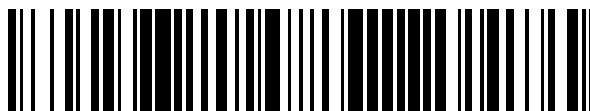


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 683 037**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.12.2012 PCT/US2012/068617**

87 Fecha y número de publicación internacional: **13.06.2013 WO13086450**

96 Fecha de presentación y número de la solicitud europea: **07.12.2012 E 12856015 (8)**

97 Fecha y número de publicación de la concesión europea: **11.07.2018 EP 2788509**

54 Título: **Diagnóstico de tumores malignos linfoides y detección de enfermedad residual mínima**

30 Prioridad:

**09.12.2011 US 201161569118 P**  
**08.05.2012 US 201261644294 P**  
**14.11.2012 US 201261726489 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**24.09.2018**

73 Titular/es:

**ADAPTIVE BIOTECHNOLOGIES CORPORATION (50.0%)**  
**1551 Eastlake Avenue East, Suite 200**  
**Seattle, Washington 98102, US y**  
**FRED HUTCHINSON CANCER RESEARCH CENTER (50.0%)**

72 Inventor/es:

**SHERWOOD, ANNA M y**  
**ROBINS, HARLAN, S**

74 Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

**ES 2 683 037 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

**Diagnóstico de tumores malignos linfoides y detección de enfermedad residual mínima**5 **Antecedentes****Campo técnico**

10 La presente descripción se refiere en general a la cuantificación altamente sensible de la representación relativa de células inmunitarias adaptativas que tienen un reordenamiento del gen que codifica el receptor de células T (TCR) o inmunoglobulina (Ig) concreto (p.ej., clonotipo), en muestras obtenidas de un sujeto antes y/o después de un tratamiento terapéutico. La determinación de las frecuencias relativas de tales clonotipos se puede utilizar para diagnosticar ciertos tumores malignos hematológicos linfoides y otros trastornos. Además, la determinación de las frecuencias relativas de tales clonotipos después del tratamiento terapéutico proporciona una detección exquisitamente sensible de enfermedad residual mínima (ERM).

**Descripción de la técnica relacionada**

20 El sistema inmunitario adaptativo protege a los organismos superiores contra infecciones y otros eventos patológicos que pueden ser atribuibles a sustancias foráneas, utilizando receptores inmunitarios adaptativos, las proteínas de reconocimiento específicas de antígenos que son expresadas por células hematopoyéticas del linaje linfóide y que son capaces de distinguir las moléculas propias de las no propias en el anfitrión. Estos linfocitos se pueden encontrar en la circulación y los tejidos de un anfitrión, y se ha descrito su recirculación entre sangre y vasos linfáticos, incluida su extravasación a través de vénulas endoteliales altas de nódulos linfáticos, así como en sitios de infección, inflamación, lesión tisular y otros insultos clínicos. (Véase, p.ej., Stein et al., 2005 Immunol. 116:1-12; DeNucci et al., 2009 Crit. Rev. Immunol. 29:87-109; Marelli-Berg et al., 2010 Immunol. 130:158; Ward et al., 2009 Biochem. J. 418:13; Gonzalez et al., 2011 Ann. Rev. Immunol. 29:215; Kehrl et al., 2009 Curr. Top. Microb. Immunol. 334:107; Steinmetz et al., 2009 Front. Biosci. (Schol. Ed.) 1:13.)

30 En consecuencia, la naturaleza dinámica del movimiento de los linfocitos a través de un organismo anfitrión se refleja en cambios en la distribución cualitativa (p.ej., especificidad por el antígeno del receptor inmunitario adaptativo expresado clonalmente (inmunoglobulina o receptor de células T), célula T frente a célula B, célula T colaboradora ( $T_H$ ) frente a célula T reguladora ( $T_{reg}$ ), célula T efectora frente a célula T de memoria, etc.) y cuantitativa de linfocitos entre tejidos, como una función de los cambios en el estado inmunitario del anfitrión.

35 El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígenos de células T y B con diversidad suficiente para reconocer el universo de posibles patógenos. Los linfocitos B maduran para expresar anticuerpos (inmunoglobulinas, Ig o Ig, también conocidas como receptores de células B, BCR) que se presentan como heterodímeros de un polipéptido de cadena pesada (H) o ligera (L), mientras que los linfocitos T expresan receptores de células T (TCR) heterodiméricos. La capacidad de las células T para reconocer el universo de antígenos asociados con diversos cánceres u organismos infecciosos es conferida por su receptor de antígeno de células T (TCR), que es un heterodímero que comprende una cadena  $\alpha$  (alfa) y una cadena  $\beta$  (beta), o una cadena  $\gamma$  (gamma) y una cadena  $\delta$  (delta). Las proteínas que componen estas cadenas están codificadas por ADN, que emplea un mecanismo único para generar la tremenda diversidad del TCR. Este receptor de reconocimiento inmunitario de múltiples subunidades se asocia con el complejo CD3 y se une a los péptidos presentados por las proteínas de clase I y II del complejo mayor de histocompatibilidad (MHC) en la superficie de las células presentadoras de antígeno (APC). La unión de TCR al péptido antigénico en la APC es un evento central en la activación de células T, que se produce en una sinapsis inmunológica en el punto de contacto entre la célula T y la APC.

50 Cada péptido de TCR contiene regiones determinantes de complementariedad (CDR) variables, así como regiones estructurales (FR) y una región constante. La diversidad de secuencias de las células T  $\alpha\beta$  está determinada en gran parte por la secuencia de aminoácidos de los bucles de la tercera región determinante de la complementariedad (CDR3) de los dominios variables de la cadena  $\alpha$  y  $\beta$ , cuya diversidad es el resultado de la recombinación entre segmentos de genes variables ( $V_\beta$ ), de diversidad ( $D_p$ ), y de empalme ( $J_p$ ) en el locus de la cadena  $\beta$ , y entre segmentos de genes  $V_\alpha$  y  $J_\alpha$  análogos en el locus de la cadena  $\alpha$ , respectivamente. La existencia de múltiples de tales segmentos de genes en los loci de la cadena  $\alpha$  y  $\beta$  de TCR permite codificar una gran cantidad de secuencias de CDR3 distintas. La diversidad de la secuencia de CDR3 aumenta adicionalmente mediante la adición y delección independiente de nucleótidos a las uniones  $V_\beta$ - $D_\beta$ ,  $D_\beta$ - $J_\beta$ , y  $V_\alpha$ - $J_\alpha$  durante el proceso de reordenamiento del gen de TCR. A este respecto, la inmunocompetencia se refleja en la diversidad de los TCR.

60 El TCR  $\gamma\delta$  es distintivo del TCR  $\alpha\beta$  porque codifica un receptor que interactúa estrechamente con el sistema inmunitario innato. TCR $\gamma\delta$ , se expresa al inicio del desarrollo, tiene una distribución anatómica especializada, tiene especificidades únicas de patógenos y moléculas pequeñas, y tiene un amplio espectro de interacciones celulares

innatas y adaptativas. En la ontogenia se establece temprano un patrón sesgado de la expresión del segmento V y J de TCR $\gamma$  ya que los subconjuntos restringidos de células TCR $\gamma\delta$  pueblan la boca, la piel, el intestino, la vagina y los pulmones prenatalmente. En consecuencia, el repertorio diverso de TCR $\gamma$  en tejidos adultos es el resultado de una amplia expansión periférica después de la estimulación por exposición ambiental a patógenos y moléculas tóxicas.

Las Ig (BCR) expresadas por las células B son proteínas que consisten en cuatro cadenas polipeptídicas, dos cadenas pesadas (cadenas H) y dos cadenas ligeras (cadenas L), que forman una estructura H<sub>2</sub>L<sub>2</sub>. Cada par de cadenas H y L contiene un dominio hipervariable, que consiste en una región V<sub>L</sub> y una V<sub>H</sub>, y un dominio constante. Las cadenas H de las Ig son de varios tipos,  $\mu$ ,  $\delta$ ,  $\gamma$ ,  $\alpha$  y  $\beta$ . La diversidad de las Ig dentro de un individuo está determinada principalmente por el dominio hipervariable. De manera similar al TCR, el dominio V de las cadenas H de Ig se crea mediante la unión combinatoria de los segmentos génicos V<sub>H</sub>, D<sub>H</sub> y J<sub>H</sub>. La diversidad de secuencias de los dominios hipervariables se incrementa adicionalmente mediante la adición y delección independientes de nucleótidos a las uniones V<sub>H</sub>-D<sub>H</sub>, D<sub>H</sub>-J<sub>H</sub>, y V<sub>H</sub>-J<sub>H</sub> durante el proceso de reordenamiento del gen Ig. A este respecto, la inmunocompetencia se refleja en la diversidad de las Ig.

La caracterización cuantitativa de las células inmunitarias adaptativas basada en la presencia en tales células de genes que codifican Ig y TCR funcionalmente reordenados que dirigen la expresión productiva de los receptores inmunitarios adaptativos se ha logrado utilizando muestras biológicas a partir de las cuales las células inmunitarias adaptativas pueden aislarse fácilmente en cantidades significativas, tales como sangre, linfa u otros fluidos biológicos. En estas muestras, las células inmunitarias adaptativas se presentan como partículas en suspensión fluida. Véase, p.ej., el documento US 2010/0330571; véase también, p.ej., Murphy, Janeway's Immunobiology (8<sup>a</sup> Ed.), 2011 Garland Science, NY, Apéndice I, pág. 717-762.

La leucemia/linfoma linfoblástico de células T agudo (T-ALL) es una neoplasia maligna, agresiva de células T inmaduras que afecta tanto a pacientes adultos como pediátricos. Aunque ha habido un progreso significativo en el tratamiento de estos pacientes con mejoras en la obtención de respuestas duraderas, sigue estando claro que un subconjunto de estos pacientes no se tratan adecuadamente y con frecuencia presentan recaída de la enfermedad, mientras que otros pueden recibir un tratamiento excesivo debido a la incapacidad para individualizar suficientemente el tratamiento clínico. Varios estudios han confirmado la importancia de evaluar la posible presencia de enfermedad residual mínima (ERM) después de un régimen de tratamiento, para ayudar a predecir los resultados clínicos de los pacientes (1-3). Por ejemplo, los pacientes que demuestran una respuesta temprana a la terapia, y que no muestran una obtención sostenida de ERM, tienen mejores resultados que los que no lo hacen (3).

De manera similar, la leucemia/linfoma linfoblástico de células  $\beta$  agudo (B-ALL) es una neoplasia maligna agresiva de células B inmaduras que afecta a pacientes adultos y pediátricos. Si bien se ha logrado un progreso significativo en el aumento del número de pacientes que logran una remisión duradera a largo plazo, un subconjunto de pacientes recae. Al igual que en T-ALL, múltiples estudios han confirmado que la presencia y la frecuencia de enfermedad residual mínima son marcadores pronósticos importantes en la B-ALL. Además, varios de estos estudios también respaldan el uso de estos datos para notificar e individualizar la terapia (p.ej., Yamaji et al., 2010 *Pediatr. Blood Cane.* 55(7):1287-95; Bhojwani et al., 2009 *Clin. Lymphoma Myeloma* 9 (Supl. 3):S222-30.

Las estrategias clínicas actuales para la evaluación de la enfermedad residual mínima incluyen citometría de flujo multiparamétrica (mpFC) y métodos basados en PCR cuantitativa que utilizan cebadores específicos del paciente (4, 5). La mpFC típicamente permite la detección de células potencialmente responsables de la enfermedad recurrente/persistente con una sensibilidad del orden de 10<sup>-4</sup> a 10<sup>-5</sup> células nucleadas. Sin embargo, la interpretación de estos datos depende del operador y del laboratorio y, en consecuencia, está limitada por una mala normalización. Además, la expresión variable de antígenos leucémicos en el contexto posterior a la terapia confunde la detección de ERM mediante mpFC (6).

En comparación, los métodos moleculares para la detección de enfermedad residual mínima pueden lograr una sensibilidad relativamente mayor, del orden de 10<sup>-5</sup> a 10<sup>-6</sup> células (7, 8). Sin embargo, las configuraciones previas de estos ensayos moleculares, principalmente ensayos basados en PCR cuantitativa en tiempo real (RT-qPCR) que utilizan cebadores específicos del paciente que se dirigen a secuencias de empalme de la región variable del receptor inmunitario adaptativo (p.ej., receptor de células T (TCR) o inmunoglobulina (Ig)), o translocaciones específicas del paciente, son complejas y difíciles de implementar en una materia uniforme (7). Por ejemplo, estos enfoques requieren la producción y el uso de sondas oligonucleotídicas individualizadas, específicas del paciente para cada paciente, lo que es laborioso, costoso y lento, e incompatible con el marco temporal en el que deben tomarse decisiones clínicas.

La secuenciación de alto rendimiento (HTS) es una tecnología emergente que puede proporcionar información sobre la complejidad de la respuesta inmunitaria adaptativa a través del análisis de la reorganización del gen del receptor linfocitario (9). Los estudios que utilizan esta tecnología han desafiado la comprensión en la técnica de la extensión de la diversidad de los linfocitos que presentan, y comparten, los individuos (9, 10), y han proporcionado una visión mecanicista de los primeros eventos genéticos moleculares críticos para la maduración del linaje de células T (11).

Recientemente, la secuenciación de alto rendimiento de los genes del receptor inmunitario adaptativo de células linfoides se ha utilizado para controlar la diversidad de linfocitos después de la inmunoterapia adoptiva con células T modificadas por receptor antigénico quimérico para el tratamiento de la leucemia linfocítica crónica refractaria a la quimioterapia (12). Por separado, la secuenciación de alto rendimiento de los genes del receptor inmunitario adaptativo de las células linfoides se ha utilizado para controlar la enfermedad en los trastornos linfoproliferativos B (13). La HTS ha demostrado la capacidad de identificar clones de células T raras (una célula T en 100.000) con alta precisión y reproducibilidad (14). Sherwood et al., 2001 describen la secuenciación profunda de los repertorios de TCR $\alpha$  y TCR $\beta$  humanos (Sci. Trans. Med. 2011, vol. 3, núm. 90-93, páginas 90-96). Weng et al., 2011 describe la secuenciación de alto rendimiento de TCR en pacientes con micosis fungoide y síndrome de Sezary (Blood (Resúmenes de reuniones de ASH) 2011, 118, Resumen 3114).

Claramente, existe una necesidad de mejorar la sensibilidad y especificidad en el diagnóstico de tumores malignos hematológicos linfoides y otras afecciones que se reflejan en la heterogeneidad y frecuencias relativas de aparición de receptores inmunitarios adaptativos únicos concretos, y en la capacidad de detectar enfermedad residual mínima (ERM). Las realizaciones descritas actualmente abordan estas necesidades y proporcionan otras ventajas relacionadas.

### Breve compendio

La invención proporciona un método para detectar enfermedad residual mínima para un tumor maligno hematológico linfoide en un sujeto después de la terapia, comprendiendo el método

(a) amplificación de ADN extraído de una primera muestra en una reacción en cadena de polimerasa múltiple (PCR) utilizando cebadores del segmentos V y J para producir una multiplicidad de moléculas de ADN reordenado amplificado que comprenden cada una una región que codifica CDR3 de TCR, en donde la primera muestra se ha obtenido del sujeto antes de la terapia y comprende una pluralidad de células hematopoyéticas linfoides,

(b) secuenciación de alto rendimiento de dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar las frecuencias relativas de regiones que codifican CDR3 de TCR reordenadas de manera única en la primera muestra, identificando así la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas,

(c) amplificación de ADN extraído de una segunda muestra en una reacción en cadena de polimerasa múltiple (PCR) utilizando los cebadores del segmento V y J para producir una multiplicidad de moléculas de ADN reordenado amplificado comprendiendo cada una una región que codifica CDR3 de TCR, en donde la segunda muestra ha sido obtenida del sujeto después de la terapia y comprende una pluralidad de células hematopoyéticas linfoides, y

(d) secuenciación de alto rendimiento de dicha multiplicidad de moléculas de ADN reordenado amplificado producidas a partir de la segunda muestra para determinar la frecuencia relativa de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas en la segunda muestra, en donde la presencia de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas a una frecuencia relativa de al menos una frecuencia de una de cada  $10^6$  a una de cada  $10^5$  regiones que codifican CDR3 de TCR indica enfermedad residual mínima para el tumor maligno hematológico linfoide;

en donde la PCR múltiple comprende adicionalmente una pluralidad de oligonucleótidos molde que tiene una pluralidad de secuencias de oligonucleótidos de fórmula general

[I] 5'-U1-B1-V-B2-R-B3-J-B4-U2-3'

en donde

(i) V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única,

(ii) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única,

(iii) U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (1) una primera secuencia de oligonucleótido adaptador universal, y (2) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a primera secuencia de oligonucleótido adaptador universal,

(iv) U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (1) una

segunda secuencia de oligonucleótido adaptador universal, y (2) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal,

(v) B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación emparejada, (1) la secuencia oligonucleotídica V única de (i) y (2) la secuencia oligonucleotídica J única de (ii), y

(vi) R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de (i)-(v).

En la presente memoria se describe un método para diagnosticar un tumor maligno hematológico linfoide en un sujeto antes de la terapia, y para detectar enfermedad residual mínima para el tumor maligno hematológico linfoide en el sujeto después de la terapia, comprendiendo el método: (I) para cada uno de (i) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que se obtienen cada una del sujeto en uno o más puntos temporales antes de la terapia, y (ii) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que se obtienen cada una del sujeto en uno o más puntos temporales después de la terapia, (a) amplificar ADN extraído de la muestra biológica en una reacción en cadena de la polimerasa múltiple (PCR) que comprende: (i) una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor de células T (TCR) humano, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica V $\gamma$  de TCR o que codifica V $\beta$  de TCR y en donde la pluralidad de cebadores del segmento V hibridan específicamente sustancialmente con todos los segmentos génicos funcionales que codifican V $\gamma$  de TCR o que codifica V $\beta$  de TCR que están presentes en la muestra, y (ii) una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor de células T (TCR) humano, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica J $\gamma$  de TCR o que codifica J $\beta$  de TCR y en donde la pluralidad de cebadores del segmento J hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican J $\gamma$  de TCR o que codifican J $\beta$  de TCR que están presentes en la muestra, en donde los cebadores del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 de TCR $\gamma$  reordenadas o sustancialmente todas las regiones que codifican CDR3 de TCR $\beta$  en la muestra para producir una multiplicidad de moléculas de ADN reordenado amplificado a partir de una población de células T en la muestra, siendo suficiente dicha multiplicidad de moléculas de ADN reordenado amplificado para cuantificar la diversidad de la región que codifica CDR3 de TCR $\gamma$  o la región que codifica CDR3 de TCR $\beta$  en la población de células T, y en donde cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud; y (b) secuenciar dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar, para cada molécula única de ADN reordenado en dicha multiplicidad de moléculas de ADN reordenado amplificado, (i) una secuencia de ADN reordenado y (ii) una frecuencia de aparición relativa de la secuencia de ADN reordenado; y (II) comparar (i) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales antes de la terapia, con (ii) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales después de la terapia, en donde (A) la presencia, en una muestra obtenida del sujeto antes de la terapia, de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15% de regiones que codifican CDR3 de TCR $\gamma$  reordenadas o regiones que codifican CDR3 de TCR $\beta$  reordenadas es diagnóstica de un tumor maligno hematológico linfoide clonal en el sujeto antes de la terapia, y en donde (B) la presencia, en una muestra obtenida del sujeto después de la terapia, de dicha al menos una secuencia única de ADN reordenado que es diagnóstica del tumor maligno hematológico linfoide clonal a una frecuencia relativa de al menos una de cada 10<sup>5</sup> regiones que codifican CDR3 de TCR indica enfermedad residual mínima para el tumor maligno hematológico linfoide.

En ciertas realizaciones adicionales, cada segmento génico funcional que codifica V de TCR comprende una secuencia señal de recombinación (RSS) del gen V y cada segmento génico que codifica J de TCR comprende una RSS del gen J, y en donde cada molécula de ADN reordenado amplificado comprende (i) al menos 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR, estando situados dichos al menos 10, 20, 30 o 40 nucleótidos contiguos 5' respecto a la RSS del gen V y (ii) al menos 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR, estando situados dichos al menos 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J.

En ciertas realizaciones la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J comprenden al menos una de (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En ciertas realizaciones cualquiera o ambos de: (i) los cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, y 6222-6273, y (ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO: 5631-5643, 5780-5798, 6212-6215, y 6274-6286.

En la presente memoria se describe un método para diagnosticar un tumor maligno hematológico linfóide en un sujeto, que comprende (I) para cada uno de (i) una o una pluralidad de muestras biológicas que se obtienen del sujeto y que comprenden cada una una pluralidad de células hematopoyéticas linfoides, (a) amplificar ADN extraído de la muestra biológica en una reacción en cadena de polimerasa múltiple (PCR) que comprende: (i) una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor de células T (TCR) humano, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica V $\gamma$  de TCR o que codifica V $\beta$  de TCR y en donde la pluralidad de cebadores del segmento V hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican V $\gamma$  de TCR o que codifican V $\beta$  de TCR que están presentes en la muestra, y (ii) una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor de células T (TCR) humano, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica J $\gamma$  de TCR o que codifica J $\beta$  de TCR y en donde la pluralidad de cebadores del segmento J hibridan específicamente sustancialmente con todos los segmentos génicos funcionales que codifican J $\gamma$  de TCR o que codifican J $\beta$  de TCR que están presentes en la muestra, en donde los cebadores del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 de TCR $\gamma$  o sustancialmente todas las regiones que codifican CDR3 de TCR $\beta$  reordenadas en la muestra para producir una multiplicidad de moléculas de ADN reordenado amplificado de una población de células T en la muestra, siendo suficiente dicha multiplicidad de moléculas de ADN reordenado amplificado para cuantificar la diversidad de la región que codifica CDR3 de TCR $\gamma$  o CDR3 de TCR $\beta$  en la población de células T, y en donde cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud; y (b) secuenciar dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar, para cada molécula única de ADN reordenado en dicha multiplicidad de moléculas de ADN reordenado amplificado, (i) una secuencia de ADN reordenado y (ii) una frecuencia de aparición relativa de la secuencia de ADN reordenado; y (II) determinar la frecuencia de aparición relativa para cada secuencia única de ADN reordenado a partir de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto, en donde (A) la presencia, en una muestra obtenida del sujeto, de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15% de las regiones que codifican CDR3 de TCR $\gamma$  reordenadas o las regiones que codifican CDR3 de TCR $\beta$  reordenadas es diagnóstica de un tumor maligno hematológico linfóide clonal en el sujeto, en donde (B) la presencia, en una muestra obtenida del sujeto, de dicha al menos una secuencia única de ADN reordenado que es diagnóstica del tumor maligno hematológico linfóide clonal a una frecuencia relativa de al menos una de cada 10<sup>5</sup> regiones que codifican CDR3 de TCR, indica enfermedad residual mínima para el tumor maligno hematológico linfóide, y en donde (C) la ausencia en la muestra obtenida del sujeto de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15% de regiones que codifican CDR3 de TCR $\beta$  reordenadas indica que el sujeto debe evaluarse para detectar el inmunofenotipo de precursor tímico temprano.

En ciertos casos, cada segmento génico funcional que codifica V de TCR o Ig comprende una secuencia señal de recombinación (RSS) del gen V y cada segmento génico funcional que codifica J de TCR o Ig comprende una RSS del gen J, y en donde cada molécula de ADN reordenado amplificado comprende (i) al menos 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR o Ig, estando situados dichos al menos 10, 20, 30 o 40 nucleótidos contiguos 5' con respecto a la RSS del gen V y (ii) al menos 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR o Ig, estando situados dichos al menos 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J. En ciertos casos, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J comprenden al menos una de (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las

secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En ciertos casos, cualquiera o ambos de (i) los cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5579-5630, 5644-5779, 5799-5816, 6124-6127, y 6222-6273, y (ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5631-5643, 5780-5798, 6212-6215, y 6274-6286.

En la presente memoria se describe un método para detectar enfermedad residual mínima para un tumor maligno hematológico linfóide en un sujeto después de la terapia, comprendiendo el método: (I) para cada una de (i) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que son obtenidas cada una del sujeto en uno o más puntos temporales antes de la terapia, y (ii) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que son obtenidas cada una del sujeto en uno o más puntos temporales después de la terapia, (a) amplificar ADN extraído de la muestra biológica en una reacción en cadena de polimerasa múltiple (PCR) que comprende: (i) una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces de hibridar específicamente por separado con al menos un polinucleótido que codifica un polipéptido de la región V del receptor de células T (TCR) humano, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica V $\alpha$  de TCR o que codifica V $\beta$  de TCR y en donde la pluralidad de cebadores del segmento V hibridan específicamente sustancialmente con todos los segmentos génicos funcionales que codifican V $\gamma$  de TCR o que codifican V $\delta$  de TCR que están presentes en la muestra, y (ii) una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor de células T (TCR) humano, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica J $\gamma$  de TCR o que codifica J $\delta$  de TCR y en donde la pluralidad de cebadores del segmento J hibridan específicamente sustancialmente con todos los segmentos génicos funcionales que codifican J $\gamma$  de TCR o que codifican J $\delta$  de TCR que están presentes en la muestra, en donde los cebadores del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 de TCR $\gamma$  reordenadas o sustancialmente todas las regiones que codifican CDR3 de TCR $\beta$  en la muestra para producir una multiplicidad de moléculas de ADN reordenado amplificado de una población de células T en la muestra, siendo suficiente dicha multiplicidad de moléculas de ADN reordenado amplificado para cuantificar la diversidad de la región que codifica CDR3 de TCR $\gamma$  o de la región que codifica CDR3 de TCR $\beta$  en la población de células T, y en donde cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud; y (b) secuenciar dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar, para cada molécula única de ADN reordenado en dicha multiplicidad de moléculas de ADN reordenado amplificado, (i) una secuencia de ADN reordenado y (ii) una frecuencia de aparición relativa de la secuencia de ADN reordenado; y (II) comparar (i) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales antes de la terapia, con (ii) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales después de la terapia, en donde (A) la presencia, en una muestra obtenida del sujeto antes de la terapia, de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15% de regiones que codifican CDR3 de TCR $\gamma$  reordenadas o regiones que codifican CDR3 de TCR $\beta$  reordenadas es diagnóstica de un tumor maligno hematológico linfóide clonal en el sujeto antes de la terapia, y en donde (B) la presencia, en una muestra obtenida del sujeto después de la terapia, de dicha al menos una secuencia única de ADN reordenado que es diagnóstica de tumor maligno hematológico linfóide clonal a una frecuencia relativa de al menos una de cada 10<sup>6</sup> o de al menos una de cada 10<sup>5</sup> regiones que codifican CDR3 de TCR indica enfermedad residual mínima para el tumor maligno hematológico linfóide.

En ciertos casos, cada segmento génico funcional que codifica V de TCR o Ig comprende una secuencia señal de recombinación (RSS) del gen V y cada segmento génico funcional que codifica J de TCR o Ig comprende una RSS del gen J, y en donde cada molécula de ADN reordenado amplificado comprende (i) al menos 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR o Ig, estando situados dichos al menos 10, 20, 30 o 40 nucleótidos contiguos 5' con respecto a la RSS del gen V y (ii) al menos 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR o Ig, estando situados dichos al menos 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J. En ciertos casos, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J comprende al menos una de (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En ciertos casos cualquiera o ambos de (i) los

cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5579-5630, 5644-5779, 5799-5816, 6124-6127, y 6222-6273, y (ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5631-5643, 5780-5798, 6212-6215, y 6274-6286.

En la presente memoria se describe un método para diagnosticar un tumor maligno hematológico linfóide en un sujeto antes de la terapia, y para detectar enfermedad residual mínima para el tumor maligno hematológico linfóide en el sujeto después de la terapia, comprendiendo el método (I) para cada uno de (i) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que se obtienen cada una del sujeto en uno o más puntos temporales antes de la terapia, y (ii) una o una pluralidad de muestras biológicas que comprenden cada una una pluralidad de células hematopoyéticas linfoides y que se obtienen cada una del sujeto en uno o más puntos temporales después de la terapia, (a) amplificar el ADN extraído de la muestra biológica en una reacción en cadena de polimerasa múltiple (PCR) que comprende: (i) una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor de células T (TCR) humano o un polipéptido de la región V de inmunoglobulina (Ig) humano, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica la región V de TCR o que codifica la región V de Ig y en donde la pluralidad de cebadores del segmento V hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican V de TCR o que codifican V de Ig que están presentes en la muestra, y (ii) una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor de células T (TCR) humano o un polipéptido de la región J de Ig humano, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica J de TCR o que codifica J de Ig y en donde la pluralidad de cebadores del segmento J hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican J de TCR o que codifica J de Ig que están presentes en la muestra, en donde los cebadores del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 de TCR reordenadas o sustancialmente todas las regiones que codifican CDR3 de Ig en la muestra para producir una multiplicidad de moléculas de ADN reordenado amplificado de una población de células linfoides en la muestra, siendo suficiente dicha multiplicidad de moléculas de ADN reordenado amplificado para cuantificar la diversidad de la región que codifica CDR3 de TCR o la región que codifica CDR3 de Ig en la población de células linfoides, y en donde cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud; y (b) secuenciar dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar, para cada molécula única de ADN reordenado en dicha multiplicidad de moléculas de ADN reordenado amplificado, (i) una secuencia de ADN reordenado y (ii) una frecuencia de aparición relativa de la secuencia de ADN reordenado; y (II) comparar (i) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales antes de la terapia, con (ii) la frecuencia de aparición relativa para cada secuencia única de ADN reordenado de dichas una o una pluralidad de muestras biológicas que se obtienen cada una del sujeto en uno o más puntos temporales después de la terapia, en donde (A) la presencia, en una muestra obtenida del sujeto antes de la terapia, de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26 %, 27%, 28%, 29% o 30% de las regiones que codifican CDR3 de TCR reordenadas o las regiones que codifican CDR3 de Ig reordenadas es diagnóstica para un tumor maligno hematológico linfóide clonal en el sujeto antes de la terapia, y en donde (B) la presencia, en una muestra obtenida del sujeto después de la terapia, de dicha al menos una secuencia única de ADN reordenado que es diagnóstica de neoplasia maligna hematológica linfóide clonal a una frecuencia relativa de al menos una de cada  $10^5$  regiones que codifican CDR3 de TCR o de Ig indica enfermedad residual mínima para el tumor maligno hematológico linfóide.

En ciertos casos, cada segmento génico funcional que codifica V de TCR o Ig comprende una secuencia señal de recombinación (RSS) del gen V y cada segmento génico funcional que codifica J de TCR o Ig comprende una RSS del gen J, y en donde cada molécula de ADN reordenado amplificado comprende (i) al menos 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR o Ig, estando situados dichos al menos 10, 20, 30 o 40 nucleótidos contiguos 5' con respecto a la RSS del gen V y (ii) al menos 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR o Ig, estando situados dichos al menos 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J. En ciertos casos, las regiones que codifican CDR3 de TCR o Ig reordenadas se seleccionan entre regiones que codifican CDR3 de TCR $\alpha$ , regiones que codifican CDR3 de TCR $\beta$ , regiones que codifican CDR3 de TCR $\gamma$ , regiones que codifican CDR3 de TCR $\delta$ , regiones que codifican CDR3 de IgH, regiones que codifican CDR3 de Igk, y regiones que codifican CDR3 de Igl reordenadas. En ciertos casos, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J comprende al menos una de (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID



NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En ciertos casos cualquiera o ambos de (i) los cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5579-5630, 5644-5779, 5799-5816, 5833-6211, y 6222-6273, y (ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO:5631-5643, 5780-5798, 5817-5832, 6212-6221, y 6274-6286.

En ciertos casos de los métodos descritos anteriormente, cualquiera o ambos de (i) los cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con uno o más de: (1) las secuencias de nucleótidos expuestas en SEQ ID NO:5579-5630, (2) las secuencias de nucleótidos expuestas en SEQ ID NO:5644-5779, (3) las secuencias de nucleótidos expuestas en SEQ ID NO:5799-5816, (4) las secuencias de nucleótidos expuestas en SEQ ID NO:5833-6123, (5) las secuencias de nucleótidos expuestas en SEQ ID NO:6124-6127, (6) las secuencias de nucleótidos expuestas en SEQ ID NO:6128-6211, (7) las secuencias de nucleótidos expuestas en SEQ ID NO:6222-6273, y (8) las secuencias de nucleótidos expuestas en SEQ ID NO:6287-6351; y (ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con uno o más de: (1) las secuencias de nucleótidos expuestas en SEQ ID NO:5631-5643, (2) las secuencias de nucleótidos expuestas en SEQ ID NO:5780-5792, (3) las secuencias de nucleótidos expuestas en SEQ ID NO:5793-5798, (4) las secuencias de nucleótidos expuestas en SEQ ID NO:5817-5832, (5) las secuencias de nucleótidos expuestas en SEQ ID NO:6212-6215, (6) las secuencias de nucleótidos expuestas en SEQ ID NO:6216-6221, y (7) las secuencias de nucleótidos expuestas en SEQ ID NO:6274-6286.

En ciertas realizaciones de los métodos descritos anteriormente, el tumor maligno hematológico linfoide se selecciona entre leucemia linfoblástica de células T aguda (T-ALL), leucemia linfoblástica de células  $\beta$  (B-ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), leucemia linfoblástica aguda (ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), linfoma de Hodgkin, linfoma no Hodgkin, linfoma cutáneo de células T, linfoma de células del manto, linfoma de células T periférico, leucemia linfocítica de células pilosas, linfoma prolinfocítico de células T, linfoma angioinmunoblástico de células T, leucemia/linfoma linfoblástico de células T, linfoma periférico de células T, leucemia/linfoma de células T adultas, micosis fungoide, síndrome de Sezary, leucemia linfoblástica de células T, neoplasia mieloproliferativa y síndrome mielodisplásico.

En ciertas realizaciones de los métodos descritos anteriormente, la PCR múltiple comprende adicionalmente una composición molde para la determinación del factor de amplificación en el que está presente un número conocido de una pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única, comprendiendo el molde una pluralidad de oligonucleótidos molde que tiene una pluralidad de secuencias de oligonucleótidos de fórmula general: 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' [I] en donde: (a) V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) del receptor inmunitario adaptativo, o el complemento de los mismos, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única; (b) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) de un receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única; (c) U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una primera secuencia de oligonucleótido adaptador universal, y (ii) una secuencia oligonucleotídica específica de la primera secuencia de secuenciación que está conectada y posicionada 5' con respecto a la primera secuencia de oligonucleótido adaptador universal; (d) U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una segunda secuencia de oligonucleótido adaptador universal, y (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal; (e) B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación pareada, (i) la secuencia oligonucleotídica de V única de (a) y (ii) la secuencia oligonucleotídica de J única (b); (f) R es o bien nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de (a) - (e), y en donde: (g) al menos uno de: (i) la pluralidad de oligonucleótidos molde comprende al menos 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, 51-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001-1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, 1701-2000, o 2001-2500 secuencias

de oligonucleótidos únicas, (ii) la pluralidad de oligonucleótidos molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento V puede hibridar específicamente y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento J puede hibridar específicamente, o (iii) la pluralidad de oligonucleótidos molde comprende al menos a o al menos b secuencias de oligonucleótidos únicas, la que sea mayor, donde a es la cantidad de segmentos génicos que codifican la región V del receptor inmunitario adaptativo único en el sujeto y b es la cantidad de segmentos génicos que codifican la región J del receptor inmunitario adaptativo único en el sujeto, y la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único; comprendiendo adicionalmente dicho método cuantificar moléculas de ADN reordenado que codifican uno o una pluralidad de receptores inmunitarios adaptativos en el ADN extraído de la muestra biológica mediante las etapas de: (1) secuenciar cuantitativamente la totalidad o una porción suficiente de cada una de dichas moléculas de ADN molde amplificado y cada una de dichas moléculas de ADN reordenado amplificado para cuantificar (a) una cantidad de producto molde de moléculas de ADN modificado que contienen al menos una secuencia código de barras de oligonucleótidos, y (b) una cantidad de producto reordenado de moléculas de ADN reordenado amplificado que carecen de una secuencia código de barras de oligonucleótidos; (2) calcular un factor de amplificación dividiendo la cantidad de producto molde de (1) (a) por la cantidad conocida de cada uno de la pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única; y (3) dividir la cantidad de producto reordenado de (1) (b) por el factor de amplificación calculado en (2) para cuantificar las moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra.

En ciertas realizaciones adicionales, en la composición molde para la determinación del factor de amplificación, todos los oligonucleótidos molde en la pluralidad de oligonucleótidos molde son de longitud sustancialmente idéntica. En determinadas realizaciones, cada cebador en el conjunto de cebadores de amplificación oligonucleotídicos es capaz de hibridar específicamente con al menos un oligonucleótido molde en la pluralidad de oligonucleótidos de la composición molde. En ciertas realizaciones, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde comprende adicionalmente un codón de parada entre V y B2. En ciertas realizaciones, en la composición molde para la determinación del factor de amplificación, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde está presente en una cantidad sustancialmente equimolar. En ciertas realizaciones, el receptor inmunitario adaptativo se selecciona entre TCRB, TCRG, TCRA, TCRD, IGH, IGK e IGL, en donde (i) el polinucleótido V de (a) codifica, respectivamente, un polipéptido de la región V del receptor TCRB, TCRG, TCRA, TCRD, IGH, IGK, o IGL, y (ii) el polinucleótido J de (b) codifica, respectivamente, un polipéptido de la región J del receptor TCRB, TCRG, TCRA, TCRD, IGH, IGK o IGL.

En determinadas realizaciones, el conjunto de cebadores de amplificación oligonucleotídicos es capaz de amplificar ADN reordenado productivamente que codifica uno o una pluralidad de receptores inmunitarios adaptativos, en donde en dicha etapa (A) de amplificación, el conjunto de cebadores de amplificación oligonucleotídicos no incluye ningún cebador oligonucleotídico que hibride específicamente con un pseudogen u orfón de la región V o con un pseudogen u orfón de la región J. En ciertas realizaciones, la pluralidad de oligonucleótidos molde tienen una pluralidad de secuencias de fórmula general (I) que se seleccionan entre: (1) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de TCRB expuestas en al menos un conjunto de 68 SEQ ID NO de V y J de TCRB, respectivamente, como se expone en la Figura 9 como Conjunto 1 de V/J de TCRB, Conjunto 2 de V/J de TCRB, Conjunto 3 de V/J de TCRB, Conjunto 4 de V/J de TCRB, Conjunto 5 de V/J de TCRB, Conjunto 6 de V/J de TCRB, Conjunto 7 de V/J de TCRB, Conjunto 8 de V/J de TCRB, Conjunto 9 de V/J de TCRB, Conjunto 10 de V/J de TCRB, Conjunto 11 de V/J de TCRB, Conjunto 12 de V/J de TCRB y Conjunto 13 de V/J de TCRB; (2) la pluralidad de secuencias oligonucleotídicas de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de TCRG expuestas en al menos un conjunto de 14 SEC ID NO de V y J de TCRG, respectivamente, como se expone en la Figura 10 como conjunto 1 de V/J de TCRG, conjunto 2 de V/J de TCRG, conjunto 3 de V/J de TCRG, conjunto 4 de V/J de TCRG y conjunto 5 de V/J de TCRG; (3) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de IGH expuestas en al menos un conjunto de 127 SEQ ID NO de V y J de IGH, respectivamente, como se expone en Figura 11 como conjunto 1 de V/J de IGH, conjunto 2 de V/J de IGH, conjunto 3 de V/J de IGH, conjunto 4 de V/J de IGH, conjunto 5 de V/J de IGH, conjunto 6 de V/J de IGH, conjunto 7 de V/J de IGH, conjunto 8 de V/J de IGH y conjunto 9 de V/J de IGH; (4) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) como se expone en SEQ ID NO: 3157-4014; (5) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) como se expone en SEQ ID NO: 4015-4084; y (6) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) como se expone en SEQ ID NO: 4085-5200.

En ciertas realizaciones de los métodos descritos anteriormente, la pluralidad de cebadores oligonucleotídicos del segmento V y del segmento J es capaz de amplificar el ADN reordenado productivamente que codifica uno o una pluralidad de receptores inmunitarios adaptativos, y en dicha etapa (a) de amplificación, los cebadores oligonucleotídicos del segmento V y del segmento J no incluyen ningún cebador oligonucleotídico que hibride específicamente con un pseudogén u orfón de la región V o con un pseudogén u orfón de la región J.

Por consiguiente, en ciertas realizaciones de algunos de los métodos descritos anteriormente, el método comprende adicionalmente cuantificar moléculas de ADN reordenado que codifican uno o una pluralidad de receptores inmunitarios adaptativos en ADN extraído de cada muestra biológica que comprende ADN de células linfoides del sujeto, comprendiendo cada receptor inmunitario adaptativo una región variable y una región de empalme, comprendiendo dicha etapa de cuantificación: (A) amplificar ADN en una reacción en cadena de polimerasa múltiple (PCR) que comprende: (1) ADN de la muestra biológica que comprende células linfoides del sujeto, (2) un conjunto de cebadores de amplificación oligonucleotídicos que es capaz de amplificar ADN reordenado que codifica uno o una pluralidad de receptores inmunitarios adaptativos en el ADN de la muestra biológica, comprendiendo el conjunto de cebadores: (a) en cantidades sustancialmente equimolares, una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor inmunitario adaptativo o con el complemento del mismo, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica la región V del receptor inmunitario adaptativo y en donde la pluralidad de cebadores del segmento V hibrida específicamente con todos los segmentos génicos funcionales que codifican la región V del receptor inmunitario adaptativo que están presentes en la composición molde, y (b) en cantidades sustancialmente equimolares, una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor inmunitario adaptativo o con el complemento del mismo, donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica la región J del receptor inmunitario adaptativo y en donde la pluralidad de cebadores del segmento J hibridan específicamente sustancialmente con todos los segmentos génicos que codifican la región J del receptor inmunitario adaptativo que están presentes en la composición molde, en donde los cebadores oligonucleotídicos del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de (i) sustancialmente todos los oligonucleótidos molde en la composición molde para producir una multiplicidad de moléculas de ADN molde amplificado, siendo dicha multiplicidad de moléculas de ADN molde amplificado suficientes para cuantificar la diversidad de los oligonucleótidos molde en la composición molde, y (ii) sustancialmente todas las moléculas de ADN reordenado que codifican receptores inmunitarios adaptativos en la muestra biológica para producir una multiplicidad de moléculas de ADN reordenado amplificado, siendo dicha multiplicidad de moléculas de ADN reordenado amplificado suficiente para cuantificar la diversidad de las moléculas de ADN reordenado en el ADN de la muestra biológica, y en donde cada molécula de ADN amplificada en la multiplicidad de moléculas de ADN molde amplificado y en la multiplicidad de moléculas de ADN reordenado amplificado es inferior a 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80 o 70 nucleótidos de longitud, y (3) una composición molde para la determinación del factor de amplificación en la que está presente un número conocido de una pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única, comprendiendo la composición molde: una pluralidad de oligonucleótidos molde que tienen una pluralidad de secuencias oligonucleotídicas de fórmula general 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' [I] en donde: (a) V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180, o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica una región variable (V) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única; (b) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única; (c) U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una primera secuencia de oligonucleótido adaptador universal, y (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una primera secuencia de oligonucleótido adaptador universal; (d) U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una segunda secuencia de oligonucleótido adaptador universal, y (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal; (e) B1, B2, B3 y B4 son cada uno independientemente, nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación emparejada, (i) la secuencia oligonucleotídica V única de (a) ) y (ii) la secuencia oligonucleotídica J única de (b); (f) R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de (a) - (e), y en donde: g) al menos uno de: (i) la pluralidad de oligonucleótidos molde comprende al menos 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, 51-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001-1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, 1701-2000 o 2001-2500 secuencias de oligonucleótidos únicas, (ii) la pluralidad de oligonucleótidos molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar específicamente cada cebador oligonucleotídico del segmento V y al menos un

oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la cual puede hibridar específicamente cada cebador oligonucleotídico del segmento J, o (iii) la pluralidad de oligonucleótidos molde comprende al menos a o al menos b secuencias de oligonucleotídicos únicas, la que sea mayor, donde a es la cantidad de segmentos génicos que codifican la región V del receptor inmunitario adaptativo único en el sujeto y b es la cantidad de segmentos génicos que codifican la región J del receptor inmunitario adaptativo único en el sujeto, y la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único; (B) secuenciar cuantitativamente toda o una porción suficiente de cada una de dichas moléculas de ADN molde amplificado y cada una de dichas moléculas de ADN reordenado amplificado para cuantificar (i) una cantidad de producto molde de moléculas de ADN molde amplificado que contienen al menos una secuencia código de barras de oligonucleótido, y (ii) una cantidad de producto reordenado de moléculas de ADN reordenado amplificado que carecen de una secuencia código de barras de oligonucleótido; (C) calcular un factor de amplificación dividiendo la cantidad de producto molde de (B)(i) por la cantidad conocida de cada uno de la pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única de (A)(2); y (D) dividir la cantidad de producto reordenado de (B)(ii) por el factor de amplificación calculado en (C) para cuantificar las moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra.

En ciertas realizaciones adicionales, en la composición molde para la determinación del factor de amplificación, todos los oligonucleótidos molde en la pluralidad de oligonucleótidos molde tienen una longitud sustancialmente idéntica. En ciertas otras realizaciones adicionales, cada cebador en el conjunto de cebadores de amplificación oligonucleotídicos es capaz de hibridar específicamente con al menos un oligonucleótido molde en la pluralidad de oligonucleótidos de la composición molde. En ciertas otras realizaciones adicionales, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde comprende adicionalmente un codón de parada entre V y B2. En ciertas otras realizaciones adicionales, en la composición molde para la determinación del factor de amplificación, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde está presente en una cantidad sustancialmente equimolar. En ciertas otras realizaciones adicionales, el receptor inmunitario adaptativo se selecciona entre TCRB, TCRG, TCRA, TCRD, IGH, IGK e IGL, en donde (i) el polinucleótido V de (a) codifica, respectivamente, un TCRB, TCRG, TCRA, TCRD, IGH, IGK o polipéptido de la región V del receptor IGL, y (ii) el polinucleótido J de (b) codifica, respectivamente, un polipéptido de la región J del receptor TCRB, TCRG, TCRA, TCRD, IGH, IGK o IGL. En ciertas otras realizaciones adicionales, el conjunto de cebadores de amplificación oligonucleotídicos es capaz de amplificar ADN reordenado productivamente que codifica uno o una pluralidad de receptores inmunitarios adaptativos, y en donde en dicha etapa (A) de amplificación, el conjunto de cebadores de amplificación oligonucleotídicos no incluye ningún cebador oligonucleotídico que hibride específicamente con un pseudogen u orfón de la región V o con un pseudogen u orfón de la región J. En ciertas otras realizaciones adicionales, la pluralidad de oligonucleótidos molde tienen una pluralidad de secuencias de fórmula general (I) que se selecciona entre: (1) la pluralidad de secuencias de oligonucleotídicos de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de TCRB expuestas en al menos un conjunto de 68 SEQ ID NO de V y J de TCRB, respectivamente, como se expone en la Figura 9 como Conjunto 1 de V/J de TCRB, Conjunto 2 de V/J de TCRB, Conjunto 3 de V/J de TCRB, Conjunto 4 de V/J de TCRB, Conjunto 5 de V/J de TCRB, Conjunto 6 de V/J de TCRB, Conjunto 7 de V/J de TCRB, Conjunto 8 de V/J de TCRB, Conjunto 9 de V/J de TCRB, Conjunto 10 de V/J de TCRB, Conjunto 11 de V/J de TCRB, Conjunto 12 de V/J de TCRB y Conjunto 13 de V/J de TCRB; (2) la pluralidad de secuencias oligonucleotídicas de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de TCRG expuestas en al menos un conjunto de 14 SEQ ID NO de V y J de TCRG, respectivamente, como se expone en la Figura 10 como conjunto 1 de V/J de TCRG, conjunto 2 de V/J de TCRG, conjunto 3 de V/J de TCRG, conjunto 4 de V/J de TCRG y conjunto 5 de V/J de TCRG; (3) la pluralidad de secuencias de oligonucleotídicos de fórmula general (I) en donde los polinucleótidos V y J tienen las secuencias V y J de IGH expuestas en al menos un conjunto de 127 SEQ ID NO de V y J de IGH, respectivamente, como se expone en Figura 11 como conjunto 1 de V/J de IGH, conjunto 2 de V/J de IGH, conjunto 3 de V/J de IGH, conjunto 4 de V/J de IGH, conjunto 5 de V/J de IGH, conjunto 6 de V/J de IGH, conjunto 7 de V/J de IGH, conjunto 8 de V/J de IGH y conjunto 9 de V/J de IGH; (4) la pluralidad de secuencias de oligonucleotídicos de fórmula general (I) como se expone en SEQ ID NO: 3157-4014; (5) la pluralidad de secuencias de oligonucleotídicos de fórmula general (I) como se expone en SEQ ID NO: 4015-4084; ; y (6) la pluralidad de secuencias de oligonucleotídicos de fórmula general (I) como se expone en SEQ ID NO: 4085-5200.

En la presente memoria se describe un método que comprende recibir un conjunto de datos que comprende datos de secuenciación para una pluralidad de amplicones obtenidos a partir de una primera muestra que comprende secuencias de ADN reordenado de una región que codifica la región determinante de la complementariedad 3 (CDR3) del receptor de células T (TCR) reordenada, obtenida la primera muestra de un sujeto; determinar a partir del conjunto de datos: la presencia de al menos una secuencia única de ADN reordenado indicativa de un tumor maligno hematológico linfóide clonal en la primera muestra, y una frecuencia de aparición relativa de la al menos una única secuencia de ADN reordenado en la primera muestra; y comparar la frecuencia de aparición relativa de la al menos una única secuencia de ADN reordenado en la primera muestra con una frecuencia de aparición relativa de la secuencia única de ADN reordenado en una segunda muestra, en donde una frecuencia de aparición relativa de la al menos una secuencia única de ADN reordenado de al menos 1 de cada  $10^5$  regiones codificantes de CDR-3 DE TCR es diagnóstica de enfermedad residual mínima para tumor maligno hematológico linfóide en el sujeto.

En ciertos casos, la primera muestra comprende células T. En ciertos casos, la segunda muestra comprende células T. En ciertos casos, la primera muestra comprende una pluralidad de células hematopoyéticas linfoides obtenidas del sujeto en un momento posterior al tratamiento del tumor maligno hematológico linfoide. En ciertos casos, la segunda muestra comprende una pluralidad de células hematopoyéticas linfoides obtenidas del sujeto en un punto temporal anterior al tratamiento. En ciertos casos, la etapa para determinar a partir del conjunto de datos la presencia de al menos una secuencia única de ADN reordenado indicativa de un tumor maligno hematológico linfoide clonal en la primera muestra comprende identificar una secuencia clonal que tiene una frecuencia relativa de al menos 15% de regiones que codifican CDR3 de TCR $\gamma$  reordenadas o regiones que codifican CDR3 de TCR $\beta$  reordenadas. En ciertos casos, el conjunto de datos se obtiene amplificando mediante una reacción en cadena de polimerasa múltiple (PCR) una pluralidad de segmentos génicos de una región que codifica CDR3 de TCR $\gamma$  reordenada o una región que codifica CDR3 de TCR $\beta$  reordenada, comprendiendo cada segmento génico una región variable (V) y una región de empalme (J), utilizando una pluralidad de cebadores oligonucleotídicos del segmento V, cada cebador oligonucleotídico de segmento V capaz de hibridar con uno o más polinucleótidos de la región V de TCR de los segmentos génicos; una pluralidad de cebadores oligonucleotídicos del segmento J, cada cebador del segmento J capaz de hibridar con uno o más polinucleótidos de la región J de TCR de los segmentos génicos; en donde la pluralidad de cebadores oligonucleotídicos del segmento V y cebadores oligonucleotídicos del segmento J promueven la amplificación de las regiones que codifican CDR3 de TCR $\gamma$  o las regiones que codifican CDR3 de TCR $\beta$  en la primera muestra para producir una pluralidad de amplicones; y secuenciar la pluralidad de amplicones para determinar una secuencia de nucleótidos para cada uno de la pluralidad de amplicones.

En una realización adicional, el polinucleótido de la región V de TCR codifica un polipéptido V $\gamma$  de TCR o un polipéptido V $\beta$  de TCR. En otra realización adicional, el polinucleótido de la región J de TCR codifica un polinucleótido J de TCR $\gamma$  o un polipéptido J de TCR $\beta$ . En otra realización adicional, cada uno de la pluralidad de cebadores oligonucleotídicos del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos. En ciertas realizaciones, la secuencia de nucleótidos es complementaria a un segmento génico que codifica V de TCR $\gamma$  funcional o un segmento génico que codifica V $\beta$  de TCR funcional. En ciertas realizaciones, el segmento génico que codifica V de TCR $\gamma$  funcional o el segmento génico que codifica V $\beta$  de TCR funcional comprenden una secuencia señal de recombinación (RSS) del gen V. En ciertas realizaciones, cada uno de la pluralidad de cebadores oligonucleotídicos del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos. En ciertas realizaciones, la secuencia de nucleótidos es complementaria a un segmento génico que codifica J de TCR $\gamma$  funcional o un segmento génico que codifica J de TCR $\beta$  funcional. En ciertas realizaciones adicionales, el segmento génico que codifica J de TCR $\gamma$  funcional o el segmento génico que codifica J de TCR $\beta$  funcional comprenden una RSS del gen J.

De acuerdo con ciertas realizaciones, la pluralidad de cebadores oligonucleotídicos del segmento V son complementarios a al menos 10, 20, 30 o 40 nucleótidos contiguos de un segmento génico que codifica TCRV. En una realización adicional, los al menos 10, 20, 30 o 40 nucleótidos contiguos están situados 5' con respecto a un gen V de RSS. En ciertas otras realizaciones, la pluralidad de cebadores oligonucleotídicos del segmento J es complementaria a al menos 10, 20 o 30 nucleótidos contiguos de un segmento génico que codifica J de TCR. En ciertas realizaciones adicionales, los al menos 10, 20 o 30 nucleótidos contiguos de un segmento génico que codifica J de TCR están situados 3' con respecto a la RSS del gen J.

En otra realización, la etapa de amplificación comprende amplificar sustancialmente todos los segmentos génicos que comprenden regiones que codifican CDR3 de TCR $\gamma$  reordenadas o regiones CDR3 de TCR $\beta$  reordenadas en la primera muestra. En otra realización, la pluralidad de cebadores oligonucleotídicos del segmento V hibrida sustancialmente con todos los segmentos génicos funcionales que codifican V $\gamma$  de TCR o segmentos génicos que codifican V $\beta$  de TCR en la primera muestra. En otra realización, la pluralidad de cebadores oligonucleotídicos del segmento J hibrida sustancialmente con todos los segmentos génicos funcionales que codifican J $\gamma$  de TCR o segmentos génicos que codifican J $\beta$  de TCR en la primera muestra. En otra realización, la pluralidad de amplicones es suficiente para cuantificar la diversidad de las regiones que codifican CDR3 de TCR $\gamma$  o la región que codifica CDR3 de TCR $\beta$  en la primera muestra. En otra realización, cada amplicón secuenciado tiene una longitud inferior a 600 nucleótidos. En otra realización, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J se seleccionan entre: (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351.

En otra realización, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, o 6222-6273. En otra realización, uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274-6286. En otra realización, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una

secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, o 6222-6273 y en donde uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274- 6286.

En la presente memoria se describe un método que comprende: recibir un conjunto de datos que comprende datos de secuenciación para una pluralidad de amplicones obtenidos a partir de una primera muestra que comprende secuencias de ADN reordenado de una región que codifica la región determinante de la complementariedad 3 (CDR3) del receptor de células T reordenada, obtenida la primera muestra de un sujeto; determinar a partir del conjunto de datos una frecuencia de aparición relativa de cada amplicón secuenciado único; y determinar si el sujeto tiene un tumor maligno hematológico linfóide clonal basándose en la presencia o ausencia de al menos una secuencia única de ADN reordenado que tiene una frecuencia de aparición relativa que excede del 15% de regiones que codifican CDR3 de TCR reordenadas en la primera muestra, en donde la frecuencia de aparición relativa de la al menos única secuencia de ADN reordenado que excede del 15% es diagnóstica de un tumor maligno hematológico linfóide clonal en el sujeto. En otro caso, la frecuencia de aparición relativa de al menos una secuencia génica reordenada única es al menos 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29% o 30% de las regiones que codifican CDR3 de TCR $\gamma$  o las regiones que codifican CDR3 de TCR $\beta$  reordenadas en la muestra. En otro caso, el método comprende adicionalmente determinar una ausencia de la muestra de al menos una secuencia única de ADN reordenado que tiene una frecuencia relativa de al menos 15% de regiones que codifican CDR3 de TCR $\gamma$  reordenadas o regiones que codifican CDR3 de TCR $\beta$  reordenadas. En ciertos casos, la etapa de determinación indica que el sujeto debe evaluarse para determinar un inmunofenotipo tímico precursor temprano.

En ciertos casos, el conjunto de datos se obtiene amplificando mediante una reacción en cadena de polimerasa múltiple (PCR) una pluralidad de segmentos génicos de una región que codifica CDR3 de TCR $\gamma$  reordenada o una región que codifica CDR3 de TCR $\beta$  reordenada, comprendiendo cada segmento génico una región variable (V) y una región de empalme (J), utilizando: una pluralidad de cebadores oligonucleotídicos del segmento V, cada cebador oligonucleotídico de segmento V capaz de hibridar con uno o más polinucleótidos de la región V de TCR de los segmentos génicos; una pluralidad de cebadores oligonucleotídicos del segmento J, cada cebador del segmento J capaz de hibridar con uno o más polinucleótidos de la región J de TCR de los segmentos génicos; en donde la pluralidad de cebadores oligonucleotídicos del segmento V y cebadores oligonucleotídicos del segmento J promueven la amplificación de las regiones que codifican CDR3 de TCR $\gamma$  reordenadas o las regiones que codifican CDR3 de TCR $\beta$  reordenadas en la muestra para producir una pluralidad de amplicones; y secuenciar la pluralidad de amplicones para determinar una secuencia de nucleótidos para cada uno de la pluralidad de amplicones. En ciertos casos adicionales, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J se seleccionan entre: (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ. ID NOs: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO : 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En algunos otros ejemplos adicionales, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124- 6127, o 6222-6273.

De acuerdo con un caso relacionado, uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274-6286. En otro caso, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, o 6222-6273 y en donde uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274- 6286.

En la presente memoria se describe un método que comprende recibir un conjunto de datos que comprende datos de secuenciación para una pluralidad de amplicones obtenidos a partir de una primera muestra que comprende secuencias de ADN reordenado de una región que codifica la región determinante de la complementariedad 3 (CDR3) del receptor de células T (TCR) reordenada o una región que codifica CDR3 de inmunoglobulina (Ig), obtenida la primera muestra de un sujeto; determinar a partir del conjunto de datos: la presencia de al menos una secuencia única de ADN reordenado indicativa de un tumor maligno hematológico linfóide clonal en la primera muestra, y una frecuencia de aparición relativa de la al menos única secuencia de ADN reordenado en la primera muestra; determinar si el sujeto tiene enfermedad residual mínima comparando la frecuencia de aparición relativa de al menos una secuencia única de ADN reordenado en la muestra con una frecuencia de aparición relativa de la secuencia única de ADN reordenado en una segunda muestra, en donde una frecuencia de aparición relativa de al

menos 1 en 10<sup>5</sup> regiones que codifican CDR-3 de TCR o CDR-3 de Ig en la primera muestra es diagnóstica de enfermedad residual mínima para un tumor maligno hematológico linfoide. En otro caso, el conjunto de datos se obtiene amplificando mediante una reacción en cadena de polimerasa múltiple (PCR) una pluralidad de segmentos génicos de regiones que codifican la región determinante de la complementariedad 3 del receptor de células T (TCR) reordenadas o regiones que codifican CDR3 de inmunoglobulina (Ig), comprendiendo cada segmento génico una región variable (V) y una región de empalme (J), utilizando: una pluralidad de cebadores oligonucleotídicos del segmento V, cada cebador oligonucleotídico del segmento V capaz de hibridar con una o más polinucleótidos de la región V de TCR o uno o más polinucleótidos de la región V de Ig; una pluralidad de cebadores oligonucleotídicos del segmento J, cada cebador del segmento J capaz de hibridar con uno o más polinucleótidos de la región J de TCR o uno o más polinucleótidos de la región J de Ig de los segmentos génicos; en donde la pluralidad de cebadores oligonucleotídicos del segmento V y cebadores oligonucleotídicos del segmento J promueven la amplificación de las regiones que codifican CDR3 de TCR o las regiones codificantes de CDR3 de Ig en la primera muestra para producir una pluralidad de amplicones; y secuenciar la pluralidad de amplicones para determinar una secuencia de nucleótidos para cada uno de la pluralidad de amplicones.

En un caso, el polinucleótido de la región V de TCR codifica un polipéptido de la región V de TCR o un polipéptido de la región V de Ig. En un caso, el polinucleótido de la región J de TCR codifica un polinucleótido de la región J de TCR o un polipéptido de la región J de Ig. En otro caso, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J se seleccionan entre: (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en la SEC ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-6351. En otro caso, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, o 6222-6273. En otro caso, uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274-6286. En otro caso, uno o más de los cebadores oligonucleotídicos del segmento V comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127, o 6222-6273 y en donde uno o más de los cebadores oligonucleotídicos del segmento J comprenden una secuencia de nucleótidos que comprende al menos 90% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5631-5643, 5780-5798, 6212-6215 o 6274-6286.

De acuerdo con ciertos casos adicionales de los métodos descritos anteriormente, el tumor maligno hematológico linfoide se selecciona entre: leucemia linfoblástica de células T aguda (T-ALL), leucemia linfoblástica de células β aguda (B-ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), leucemia linfoblástica aguda (ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), linfoma de Hodgkin, linfoma no Hodgkin, linfoma cutáneo de células T, linfoma de células del manto, linfocitos T periféricos linfoma, leucemia de células pilosas, linfoma prolinfocítico T, linfoma de células T angioinmunoblástico, leucemia/linfoma linfoblástico de células T, linfoma de células T periférico, leucemia/linfoma de células T adultas, micosis fungoide, síndrome de Sezary, leucemia linfoblástica de células T, neoplasia mieloproliferativa y síndrome mielodisplásico.

En ciertos casos de los métodos descritos anteriormente, el método comprende adicionalmente: recibir un segundo conjunto de datos que comprende secuenciar datos para normalizar las eficacias de amplificación de una pluralidad de conjuntos de cebadores, comprendiendo el segundo conjunto de datos de secuenciación para una pluralidad de amplicones molde. En ciertos casos adicionales, el segundo conjunto de datos se obtiene amplificando mediante PCR múltiple una pluralidad de oligonucleótidos molde únicos para producir una pluralidad de amplicones molde únicos, teniendo cada oligonucleótido molde único una concentración conocida de moléculas y una secuencia código de barras única, y comprendiendo la pluralidad de conjuntos de cebadores una pluralidad de cebadores oligonucleotídicos del segmento V y una pluralidad de cebadores oligonucleotídicos del segmento J; y secuenciando cuantitativamente la pluralidad de amplicones molde únicos. En un caso, el método comprende adicionalmente el ajuste de la concentración de uno o más cebadores oligonucleotídicos del segmento V y uno o más cebadores oligonucleotídicos del segmento J para reducir la amplificación sesgada de segmentos de genes que codifican TCR o Ig en la primera muestra. En ciertos casos, la amplificación por PCR múltiple de una pluralidad de oligonucleótidos molde únicos comprende el uso de concentraciones equimolares de cebadores oligonucleotídicos del segmento V y cebadores oligonucleotídicos del segmento J. En ciertos casos, la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J se seleccionan entre: (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en la SEC ID NO: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO: 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-6211 y 6216-6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286, y (8) las secuencias expuestas en SEQ ID NO: 6222-

6351.

En otro caso, el método descrito anteriormente comprende adicionalmente: cuantificar basándose en el primer conjunto de datos y el segundo conjunto de datos: (a) una cantidad de producto molde de amplicones molde que contienen al menos una secuencia código de barras de oligonucleótidos, y (b) una cantidad de producto reordenado amplicones secuenciados que carecen de una secuencia código de barras de oligonucleótidos; calcular un factor de amplificación dividiendo la cantidad de producto molde por la cantidad conocida de cada una de la pluralidad de oligonucleótidos molde únicos; y determinar la cantidad de secuencias que codifican TCR o Ig reordenadas únicas en la primera muestra dividiendo la cantidad de producto reordenado por el factor de amplificación. En un caso, el método comprende además determinar la cantidad de genomas únicos de células T o genomas de células B en la primera muestra.

De acuerdo con ciertos métodos relacionados, la pluralidad de oligonucleótidos molde únicos comprende una fórmula general: 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' (Fórmula I), en donde V comprende un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región V, o un complemento de la misma, y donde V comprende una secuencia oligonucleotídica única, en donde J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120, o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región J, o un complemento de la misma, y en donde J comprende una secuencia oligonucleotídica única, en donde U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una primera secuencia de oligonucleótido adaptador universal, y (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada en 5' con respecto a una primera secuencia de oligonucleótido adaptador universal, en donde U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una segunda secuencia de oligonucleótido adaptador universal, y (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal, en donde B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3-25 nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación emparejada, una secuencia oligonucleotídica V única y una secuencia oligonucleotídica J única, y en donde R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica no encontrada en el oligonucleótido molde.

### Breve descripción de las diversas vistas de los dibujos

La Figura 1 (Fig. 1A-1B) representa la comparación de la evaluación de la clonalidad de células T en muestras de sangre de pacientes con T-ALL, mediante secuenciación de alto rendimiento (HTS) y citometría de flujo multiparamétrica (mpFC), antes del tratamiento (Día 0) y 29 días después del tratamiento (día 29). Los clones detectados mediante HTS (color rojo) se encontraron en las 12 muestras post-tratamiento informativas que fueron positivas para ERM mediante mpFC (color azul), así como en un nivel inferior en 10 pacientes adicionales, lo que sugiere una sensibilidad superior para HTS frente a mpFC. Los pacientes 39 y 52 tenían un inmunofenotipo próximo al del precursor tímico temprano (nETP); todos los demás tenían leucemia linfoblástica aguda de células T (T-ALL) típica. La Fig. 1A muestra la clonalidad de las células T según se detectó el día 0 utilizando citometría de flujo multiparamétrica (mpFC) y secuenciación de alto rendimiento (HTS) de regiones de ADN que codifican CDR3 de TCR $\beta$  amplificadas y reordenadas. Las muestras con una población clonal fueron identificadas mediante secuenciación *TCRB*. Los datos se representan como la frecuencia de la secuencia clonal en la población total de células T. La Fig. 1B muestra la detección de enfermedad residual mínima (ERM) el día 29 en muestras de pacientes que exhibieron un reordenamiento de TCR $\beta$  clonal identificable el día 0.

La Figura 2 muestra la detección de ERM en casos de precursores tímicos tempranos (ETP) o próximos a ETP (nETP) el día 29. Todos los casos (5/5) designados como precursor tímico temprano ETP (color amarillo), y la mayoría de los casos (6/8) designados como próximos a ETP (nETP, color verde) carecían de un reordenamiento de genes de *TCRB* clonal, completo el día 0, según se evaluó mediante HTS de regiones de ADN que codifican CDR3 de TCR $\beta$  amplificadas y reordenadas. Estos casos, sin embargo, tenían ERM detectable mediante mpFC.

La Figura 3 muestra inmunofenotipos representativos de T-ALL caracterizados mediante mpFC de subtipo ETP (fila superior), subtipo próximo a ETP (fila central) y subtipo ALL típica/no ETP (fila inferior). Las poblaciones de linfoblastos T se muestran en color verde azulado. Se utilizaron los criterios referidos para el inmunofenotipo ETP (5). Cinco de 43 casos tenían un inmunofenotipo ETP. Ocho de 43 casos tenían un inmunofenotipo próximo a ETP. Los otros 30 casos fueron T-ALL típica y no fueron ni ETP ni próximos a ETP.

La Figura 4 (Fig. 4A-4B) muestra secuencias de genes de *TCRB* clonales caracterizadas en muestras del Día 0 (pretratamiento) mediante secuenciación de alto rendimiento (HTS) de regiones de ADN que codifican CDR3 de TCR $\beta$  amplificadas y reordenadas. Las secuencias CDR3 de *TCRB* (TCR $\beta$ ) se muestran para casos típicos de T-ALL. 31 de 43 (72,1%) muestras de pretratamiento tenían una secuencia de *TCRB* clonal



detectable; al menos 27 también tenían una secuencia *TCRG* (*TCR $\gamma$* ) clonal. Se muestran secuencias de CDR3 de *TCRB* con regiones de inserciones de nucleótidos (subrayado simple) y gen D (subrayado doble) destacadas.

5 La Figura 5 muestra una comparación de los repertorios de *TCRB* clonales normal y T-ALL por medio de la frecuencia de aparición relativa de los cinco clones de *TCRB* de frecuencia más alta basados en secuencias de *TCRB* identificadas mediante HTS en muestras de sangre de seis sujetos normales sanos (izquierda) y seis casos diagnosticados de T-ALL (derecha).

10 La Figura 6 muestra un diagrama esquemático de un oligonucleótido molde sintético ilustrativo para su uso en la determinación de un factor de amplificación para estimar la cantidad de secuencias codificantes del receptor inmunitario adaptativo reordenado (*TCR* o *BCR*) en una muestra. U1, U2, oligonucleótidos adaptadores universales; B1-4, oligonucleótidos código de barras; V, oligonucleótido de la región variable; J, oligonucleótido de la región de empalme; R, sitio de reconocimiento para enzimas de restricción; S, codón de parada opcional.

15 La Figura 7 muestra los resultados del cálculo de un factor de amplificación para cada par VJ en una composición molde que se añadió a una amplificación mediante PCR multiplexada de secuencias de *IGH*, y de la posterior realización del promedio del factor de amplificación en todos los moldes sintéticos para estimar la multiplicidad de cobertura de secuencia. Cada barra representa la cantidad total de moléculas sintéticas con la misma multiplicidad de cobertura de secuencia estimada. La mediana de la multiplicidad de cobertura de secuencia estimada fue de 3,25.

20 La Figura 8 muestra un gráfico del número de células B que se estimaron mediante PCR multiplexada utilizando una composición molde sintética y un factor de amplificación calculado como se describe en la presente memoria, frente al número conocido de células B utilizadas como fuente de moldes de ADN natural. Se añadieron cantidades constantes de ADN total, de las cuales una proporción variada era ADN de células B, a reacciones de PCR multiplexadas. Además, se añadieron aproximadamente 5.000 moléculas de oligonucleótido molde sintético de fórmula general (I) (aproximadamente 4-5 moléculas de cada secuencia) en cada reacción. El número de genomas de células B presentes en el ADN añadido a cada reacción de PCR, según se estima por medio de la amplificación de las moléculas de oligonucleótidos molde sintéticas, se representa en el eje Y. En el eje X se muestra la cantidad real de ADN añadido a cada reacción de PCR. La correlación entre los valores reales y esperados fue  $r^2 = 0,9988$ .

30 La Figura 9 (Figuras 9A-9L) muestra conjuntos de V/J de *TCRB* ilustrativos (68 V + 13 J) para su uso en composiciones molde que comprenden una pluralidad de secuencias de oligonucleótidos de fórmula general 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' [I], para su uso en la determinación de un factor de amplificación cuando se amplifica ADN reordenado que codifica uno o una pluralidad de polipéptidos de la cadena  $\beta$  del receptor de células T humanas (*TCRB*).

35 La Figura 10 (Figuras 10A-10B) muestra conjuntos de V/J de *TCRG* ilustrativos (14 V + 5 J) para su uso en composiciones molde que comprenden una pluralidad de secuencias de oligonucleótidos de fórmula general 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' [I], para su uso en la determinación de un factor de amplificación cuando se amplifica ADN reordenado que codifica uno o una pluralidad de polipéptidos de la cadena  $\gamma$  del receptor de células T humanas (*TCRG*).

40 La Figura 11 (Figuras 11A-11M) muestra conjuntos de V/J de *IGH* ilustrativos (127 V + 9 J) para su uso en composiciones molde que comprenden una pluralidad de secuencias de oligonucleótidos de fórmula general 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' [I], para su uso en la determinación de un factor de amplificación cuando amplifica ADN reordenado que codifica uno o una pluralidad de polipéptidos de cadena pesada de inmunoglobulina humana (*IGH*).

45 La Figura 12 es un diagrama de bloques de alto nivel que ilustra una vista funcional de un sistema informático típico para su uso en métodos de acuerdo con ciertas realizaciones de la invención.

### Breve descripción de las secuencias

50 SEQ ID NO: 1-871 son secuencias oligonucleotídicas molde de *TCRB* de una composición molde para la determinación del factor de amplificación.

SEQ ID NO: 872-1560 son secuencias oligonucleotídicas molde de *TCRB* de una composición molde para la determinación del factor de amplificación.

55 SEQ ID NO: 1561-1630 son secuencias oligonucleotídicas molde de *TCRG* de una composición molde para la determinación del factor de amplificación.

SEQ ID NO: 1631-1695 son secuencias de oligonucleótidos cebadores para la amplificación de *TCRB* (*TCRB*-J, SEQ ID NO: 1631-1643; *TCRB*-V, SEQ ID NO: 1644-1695).

SEQ ID NO: 1696-1709 son secuencias de oligonucleótidos cebadores para secuenciación de ADN.

SEQ ID NO: 1710-1731 son secuencias de oligonucleótidos adaptadores.

60 SEQ ID NO: 1732-1745 son secuencias de oligonucleótidos cebadores para la amplificación de *TCRG* (*TCRG*-V, SEQ ID NO: 1732-1741; *TCRG*-J, SEQ ID NO: 1742-1745).

SEQ ID NO: 1746-1747 son secuencias de oligonucleótidos adaptadores.

SEQ ID NO: 1748 es una secuencia código de barras de oligonucleótido ilustrativa.

SEQ ID NO: 1749 es una secuencia de polinucleótido V ilustrativa.

SEQ ID NO: 1750 es una secuencia de polinucleótido J ilustrativa.  
 SEQ ID NO: 1751-1752 son secuencias de oligonucleótidos adaptadores Illumina Nextera™.  
 SEQ ID NO: 1753-1804 son secuencias de oligonucleótidos cebadores para la amplificación de TCRB.  
 SEQ ID NO: 1805-2920 son secuencias oligonucleotídicas molde de IGH de una composición molde para la  
 5 determinación del factor de amplificación.  
 SEQ ID NO: 2921-2988 son secuencias de polinucleótidos de V de TCRB para uso en una composición  
 molde para la determinación del factor de amplificación.  
 SEQ ID NO: 2989-3001 son secuencias de polinucleótidos de J de TCRB para utilizar en una composición  
 molde para la determinación del factor de amplificación.  
 10 SEQ ID NO: 3002-3015 son secuencias de polinucleótidos de V de TCRG para uso en una composición  
 molde para la determinación del factor de amplificación.  
 Las SEQ ID NO: 3016-3020 son secuencias de polinucleótidos de J de TCRG para uso en una composición  
 molde para la determinación del factor de amplificación.  
 Las SEQ ID NO: 3021-3147 son secuencias de polinucleótidos de V de IGH para utilizar en una composición  
 15 molde para la determinación del factor de amplificación.  
 SEQ ID NO: 3148-3156 son secuencias de polinucleótidos J de IGH para uso en una composición molde para  
 la determinación del factor de amplificación.  
 SEQ ID NO: 3157-4014 son secuencias oligonucleotídicas molde de TCRB de una composición molde para la  
 determinación del factor de amplificación.  
 20 SEQ ID NO: 4015-4084 son secuencias oligonucleotídicas molde de TCRG de una composición molde para la  
 determinación del factor de amplificación.  
 SEQ ID NO: 4085-5200 son secuencias oligonucleotídicas molde de IGH de una composición molde para la  
 determinación del factor de amplificación.  
 SEQ ID NO: 5201-5286 son cebadores directos de oligonucleótidos para la amplificación de IGH.  
 25 SEQ ID NO: 5287-5293 son cebadores inversos de oligonucleótidos para la amplificación de IGH.  
 SEQ ID NO: 5294-5386 son secuencias de oligonucleótidos cebadores para la amplificación de IGH (V de  
 IGH, SEQ ID NO: 5294-5379; J de IGH, SEQ ID NO: 5380-5386).  
 SEQ ID NO: 5387-5578 son secuencias de oligonucleótidos para cebadores de cola adaptadores.  
 SEQ ID NO: 5579-6382 son secuencias de oligonucleótidos cebadores para la amplificación de secuencias  
 30 que codifican el receptor inmunitario adaptativo.  
 SEQ ID NO: 6383-6388 son ejemplos de cebadores oligonucleotídicos para amplificar secuencias de genes  
 de control ("constitutivos").

### Descripción detallada

35 La presente invención proporciona, en ciertas realizaciones y como se describe en la presente memoria, métodos  
 sorprendentemente ventajosos para detectar enfermedad residual mínima (ERM) para tumores malignos  
 hematológicos linfoides en sujetos que siguen el tratamiento para tales afecciones. Estas y otras realizaciones  
 relacionadas están dirigidas en la parte pertinente al uso de metodologías de secuenciación de alto rendimiento  
 40 (HTS) recientemente desarrolladas para la caracterización cualitativa y cuantitativa de secuencias de ADN  
 reordenado de genes que codifican receptores de células T (TCR).

45 La amplificación multiplexada y la secuenciación de alto rendimiento de las secuencias de ADN que codifican TCR y  
 BCR (IG) reordenadas son descritas, por ejemplo, por Robins et al., 2009 Blood 114, 4099; Robins et al., 2010 Sci.  
 Translat. Med. 2: 47ra64; Robins et al., 2011 J. Immunol. Meth. doi: 10.1016/j.jim.2011.09. 001; Sherwood et al.  
 2011 Sci. Translat. Med. 3: 90ra61; Publicación de Estados Unidos Núm. 2012/0058902, Publicación de Estados  
 Unidos Núm. 2010/0330571, documento WO/2010/151416, documento WO/2011/106738 y documento  
 WO2012/027503.

50 Ciertas realizaciones presentes pueden beneficiar de forma útil el tratamiento clínico en trastornos linfoides tales  
 como tumores malignos hematológicos linfoides u otros trastornos linfoproliferativos, por ejemplo, leucemia/linfoma  
 linfoblástico agudo de células T (T-ALL), facilitando el diagnóstico de la enfermedad y/o permitiendo la evaluación de  
 enfermedad residual mínima en los pacientes después del tratamiento. Como se describe en la presente memoria, la  
 55 secuenciación de alto rendimiento de los loci del gen del receptor de células T ofrece ventajas significativas sobre la  
 evaluación previa del estado de ERM mediante citometría de flujo multiparamétrica de nueve colores (mpFC), que  
 incluye una normalización más sencilla de las metodologías de ensayo de prueba.

60 Por ejemplo, los presentes métodos proporcionan ventajas en uno o más de sensibilidad, especificidad, tiempo,  
 coste, preparación de reactivos, configuración y calibración del equipo, superación de la falta de fiabilidad de los  
 métodos anteriores debido a la dependencia de manipulaciones subjetivas del operador y otros factores, en  
 comparación ya sea con mpFC o con enfoques moleculares específicos del paciente para la evaluación de ERM.  
 Como se describe en la presente memoria en los Ejemplos, los presentes métodos permitieron inesperadamente la  
 detección de ERM con una sensibilidad comparable o mayor que la mpFC, así como una sorprendente detección de  
 ERM en muestras de pacientes para las cuales la mpFC no identificó ERM. Además, las metodologías HTS actuales

identificaron casos de T-ALL que exhiben propiedades similares a las del precursor tímico temprano (ETP). Los casos de ETP/T-ALL descritos en la presente memoria carecían de un reordenamiento del *gen TCRB* clonal en el momento del diagnóstico, como se identifica utilizando los métodos descritos en la presente memoria. En vista del curso clínico agresivo de ETP/T-ALL, la gestión mejorada de la enfermedad y otras ventajas sin precedentes asociadas con la clasificación temprana de ETP/T-ALL se proporcionan de este modo por medio de ciertas realizaciones que se divulgan en la presente memoria.

De acuerdo con algunas de las presentes realizaciones, la capacidad para caracterizar cualitativamente (p.ej., identificando secuencias únicas de regiones génicas que codifican CDR3 de TCR reordenadas en una muestra) y cuantitativamente (p.ej., determinando la frecuencia de aparición relativa de cualquier secuencia de ADN que codifica CDR3 de TCR única reordenada como una proporción del número total de secuencias codificantes de TCR en la muestra) la diversidad de receptores inmunitarios adaptativos proporciona los medios para diagnosticar enfermedades de células linfoides (p.ej., un tumor maligno hematológico linfóide), y también proporciona medios para detectar la presencia de enfermedad residual mínima (ERM) en un sujeto durante y después del tratamiento.

Los ejemplos ilustrativos y no limitantes de enfermedades de células linfoides para las que las realizaciones de la presente invención divulgadas pueden ayudar útilmente en el diagnóstico y/o detección de ERM incluyen tumores malignos hematológicos linfoides tales como leucemia linfoblástica aguda (ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), otros linfomas y leucemias incluyendo linfoma Hodgkin y no Hodgkin, linfoma cutáneo de células T, linfoma de células del manto, linfoma periférico de células T, leucemia de células pilosas, linfoma prolinfocítico de células T, linfoma angioinmunoblástico de células T, leucemia linfoblástica T/linfoma, linfoma de células T periférico-no especificado de otro modo, leucemia/linfoma de células T adultas, micosis fungoide, síndrome de Sezary, leucemia linfoblástica de células T y cualquier otro cáncer que implique células T o células B; y también puede incluir otros trastornos linfoproliferativos, incluyendo neoplasmas mieloproliferativos, síndrome mielodisplásico y otros.

La detección de ERM puede jugar un papel importante no solo en el control de la respuesta de un paciente a la terapia, sino también en el diagnóstico preciso de la causa subyacente de los principales signos clínicos. La ERM típicamente se refiere a la presencia de células malignas (generalmente en referencia a células leucémicas) que no son detectables basándose en la morfología celular. Varios estudios han demostrado que la detección cuantitativa de ERM en las neoplasias linfoides predice el resultado clínico. (Szczepanski T, et al., *Lancet Oncol* 2001; 2:409-17; van Dongen JJ, et al., *Lancet* 1998; 352:1731-8; Bruggemann M, et al., *Acta Haematol* 2004; 112:111-9; Cave H, et al., *N Engl J Med* 1998; 339:591-8; Coustan-Smith E, et al., *Blood* 2000; 96:2691-6; Coustan-Smith E, et al., *Blood* 2002; 100:52-8; Wells DA, et al., *Am J Clin Pathol* 1998; 110:84-94; Radich J, et al., *Biol Blood Marrow Transplant* 1995; 1:24-31; Bahloul M, et al., *Best Pract Res Clin Haematol* 2005; 18:97-111; Hoshino A, et al., *Tohoku J Exp Med*. 2004; 203:155-64; Ciudad J, et al., *Br J Haematol* 1999; 104:695-705; Lucio P, et al., *Leukemia* 1999; 13:419-27). Ciertas realizaciones contempladas que se relacionan con el diagnóstico, por ejemplo, pueden incluir la detección de ERM en linfomas realizando en primer lugar secuenciación de alto rendimiento (HTS) como se describe en la presente memoria de ADN que codifica uno o más receptores inmunitarios adaptativos (p.ej., TCR) en una muestra que contiene células linfoides obtenidas de linfa o ganglios linfáticos, sangre periférica u otros tejidos, para identificar una o más secuencias clonales de TCR, y pueden implicar adicionalmente el seguimiento de la frecuencia de aparición de tales secuencias clonales de TCR en muestras de sangre en uno o más puntos temporales posteriores.

El control de la respuesta de un paciente con cáncer a un tratamiento terapéutico basado en la cuantificación de la carga tumoral (p.ej., mediante la detección de ERM) puede ayudar a la evaluación de un riesgo relativo de recaída y también puede utilizarse para identificar pacientes que pueden beneficiarse de la reducción de la terapia, la intensificación de la terapia, la reducción de la inmunosupresión para el efecto de injerto contra leucemia después de un trasplante de células madre o terapia de células T adoptivas. (Bradfield SM, et al., *Leukemia* 2004; 18: 1156-8). También se puede encontrar enfermedad mínima en situaciones de diagnóstico. Por ejemplo, los bajos niveles de células B monoclonales en pacientes que se presentan clínicamente con citopenia pueden levantar sospechas para un diagnóstico de síndrome mielodisplásico. (Wells et al., *Blood* 2003; 102: 394-403). La detección de enfermedad mínima también se encuentra en la estadificación del linfoma, que puede implicar la detección de niveles bajos de células tumorales sobre un fondo de células normales. La detección de enfermedad mínima como se describe en la presente memoria (p.ej., como la detección de ERM en pacientes con cáncer linfóide después del tratamiento) no necesita limitarse a la verificación de los efectos del tratamiento, sino que también puede encontrar usos en entornos de diagnóstico donde no hay disponible una población de referencia para la comparación.

De acuerdo con ciertas realizaciones, se proporciona así un método para detectar enfermedad residual mínima que comprende amplificar el ADN extraído de una primera muestra (p.ej., médula ósea, linfa o sangre, dependiendo del tipo de cáncer) obtenida del paciente en una reacción en cadena de la polimerasa múltiple (PCR) utilizando conjuntos de cebadores del segmento V y J como se describe en la presente memoria y en uno o más de Robins et al., 2009 *Blood* 114, 4099; Robins et al., 2010 *Sci. Translat. Med.* 2:47ra64; Robins et al., 2011 *J. Immunol. Meth.* doi:10.1016/j.jim.2011.09. 001; Sherwood et al. 2011 *Sci. Translat. Med.* 3:90ra61; Publicación de Estados Unidos

Núm. 2012/0058902, Publicación de Estados Unidos Núm. 2010/0330571, documento WO/2010/151416, documento WO/2011/106738 y documento WO2012/027503, en donde la primera muestra se toma en un primer punto temporal antes o durante un tratamiento terapéutico, y en donde la primera muestra comprende una población de células T, en donde los cebadores del segmento V y del segmento J permiten la amplificación de sustancialmente todas las combinaciones de segmentos V y J de un locus de TCR reordenado en la PCR múltiple para producir una multiplicidad de moléculas de ADN amplificadas que comprenden cada una una región que codifica CDR3 de TCR; medir una frecuencia relativa de las regiones que codifican CDR3 de TCR reordenadas de manera única, identificando de este modo la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas; amplificar con los cebadores del segmento V y los cebadores del segmento J el ADN extraído de una segunda muestra del paciente, en una reacción en cadena de la polimerasa múltiple (PCR), en donde la segunda muestra se toma en un punto temporal posterior a la primera muestra y en donde la segunda muestra comprende una población de células T, que mide la frecuencia relativa de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas en la segunda muestra; en donde la presencia de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas en la segunda muestra indica la presencia de enfermedad residual mínima.

En ciertas realizaciones, la presencia de la región que codifica CDR3 reordenada de manera única de los uno o más clones malignos en la segunda muestra a una frecuencia relativa mayor que  $10^{-6}$  regiones que codifican CDR3 de TCR indica la presencia de enfermedad residual mínima. En ciertas otras realizaciones, la frecuencia relativa puede ser mayor que  $1 \times 10^{-5}$ .

Los presentes métodos se pueden utilizar, adicional o alternativamente, para diagnosticar un trastorno que afecta a células linfoides tales como un tumor maligno hematológico linfóide, por ejemplo una enfermedad maligna asociada con clones de células T específicos, donde se puede identificar al menos un clonotipo maligno específico en un paciente. Se puede detectar un clonotipo por la presencia en ADN extraído de una muestra de al menos una molécula de ADN reordenado que codifica una CDR3 de TCR y que tiene una secuencia de ADN única, donde la frecuencia relativa de la secuencia única de ADN reordenado se determina cuantificando la diversidad del receptor inmunitario adaptativo de acuerdo con los métodos descritos en la presente memoria.

Preferiblemente y en ciertas realizaciones, la detección de al menos una secuencia de ADN que codifica CDR3 reordenada única que tiene una frecuencia relativa de al menos 15% de la frecuencia de aparición de todas las secuencias que codifican CDR3 reordenadas detectables para un polipéptido del receptor inmunitario adaptativo concreto (p.ej. TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$  o TCR $\delta$ ) indica la presencia de un clonotipo asociado a tumor maligno. En ciertas otras realizaciones, la detección de al menos una secuencia de ADN que codifica CDR3 reordenada única que tiene una frecuencia relativa de al menos 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29% o 30% de la frecuencia de aparición de todas las secuencias codificantes de CDR3 reordenadas detectables para un polipéptido receptor inmunitario adaptativo concreto, indica la presencia de un clonotipo asociado a tumor maligno.

En ciertas otras realizaciones, se identifica un clonotipo que es diagnóstico para un tumor maligno hematológico linfóide clonal determinando primero un valor normalizado  $R_c$  ( $\pm$  desviación típica, DT) para la frecuencia de aparición relativa de la secuencia que codifica CDR3 reordenada única más abundante detectable para un polipéptido receptor inmunitario adaptativo concreto (p.ej., TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$  o TCR $\delta$ ) en una muestra de uno o más sujetos de control sanos normales, apropiadamente emparejados. A continuación, se determina  $R_x$  ( $\pm$  DT), la frecuencia de aparición relativa de secuencias codificantes de CDR3 reordenadas detectables para el polipéptido receptor inmunitario adaptativo concreto en una muestra de un sujeto que se sabe que tiene o se sospecha que tiene un tumor maligno hematológico linfóide. La muestra puede obtenerse del sujeto antes, durante o después del tratamiento, tal como tratamiento terapéutico dirigido al tumor maligno linfóide. Se puede considerar que existe tumor maligno hematológico linfóide clonal presente, de acuerdo con ciertas realizaciones independientes contempladas en la presente memoria, cuando  $R_x$  tiene al menos 2,5, 3,0, 3,25, 3,5, 3,75, 4,0, 4,25, 4,5, 4,75 o 5,0 desviaciones típicas más altas que  $R_c$ . Para una muestra obtenida durante o después del tratamiento,  $R_x$  por lo tanto, se puede utilizar como un indicador de ERM. Por consiguiente, se puede identificar un clonotipo maligno en estas y en realizaciones relacionadas mediante la presencia, a un nivel relativo estadísticamente significativo, de un clonotipo en la población de regiones que codifican CDR3 reordenadas de manera única que se han amplificado y secuenciado utilizando los métodos de la presente memoria.

Ciertas realizaciones divulgadas actualmente permiten de forma ventajosa la determinación de el número de células inmunitarias adaptativas (p.ej., linfocitos T o B) en una muestra que contiene células linfoides y que también puede contener células no linfoides, donde el ADN se extrae de la muestra para su uso como fuente de moldes para la amplificación multiplexada de ADN de acuerdo con los presentes métodos. En estas y otras realizaciones relacionadas, que se describen con mayor detalle en otra parte de la presente memoria, se incluye (o se "incorpora") a la reacción de amplificación multiplexada una composición molde artificial para la determinación del factor de amplificación. La secuenciación cuantitativa de los productos de amplificación de la composición molde (cuyos productos son identificables en virtud de secuencias únicas código de barras de oligonucleótidos contenidas en ella) permite el cálculo de un factor de amplificación que es una característica de la reacción de amplificación

multiplexada concreta (p.ej., moldes, cebadores de amplificación, condiciones de reacción, etc.) que se ha realizado. El factor de amplificación se puede utilizar para calcular, a partir de los datos de secuenciación cuantitativa, la cantidad de moléculas únicas de ADN que codifican el receptor inmunitario adaptativo (p.ej., genomas de células linfoides que contienen el ADN del receptor inmunitario adaptativo reordenado que codifica el ADN) en la muestra.

Estas y otras realizaciones relacionadas comprenden adicionalmente cuantificar moléculas de ADN reordenado que codifican uno o una pluralidad de receptores inmunitarios adaptativos en ADN extraído de muestras biológicas que comprenden ADN de células linfoides de un sujeto como se proporciona en la presente memoria, realizando las etapas de amplificación en presencia de un cantidad conocida de la composición molde para la determinación del factor de amplificación; secuenciar cuantitativamente los productos de amplificación del ADN que codifica el receptor inmunitario adaptativo reordenado en la muestra y de la composición molde; calcular un factor de amplificación; y dividir el número de productos reordenados amplificados del ADN que codifica el receptor inmunitario adaptativo reordenado por el factor de amplificación para cuantificar las moléculas de ADN que codifican el receptor inmunitario adaptativo único que están presentes en la muestra.

### Muestras y sujetos

El sujeto o la fuente biológica, a partir de los cuales se puede obtener una muestra biológica de prueba, puede ser un animal humano o no humano, o un organismo transgénico o clonado o modificado mediante ingeniería tisular (incluso a través del uso de células madre). En ciertas realizaciones preferidas de la invención, se puede conocer que el sujeto o fuente biológica tiene, o puede sospecharse que tiene o está en riesgo de tener, un cáncer hematopoyético linfóide u otra afección maligna, o una enfermedad autoinmunitaria, o una afección inflamatoria, y en ciertas realizaciones preferidas de la invención, se puede conocer que el sujeto o fuente biológica está libre de un riesgo o presencia de dicha enfermedad.

Ciertas realizaciones preferidas contemplan un sujeto o fuente biológica que es un sujeto humano tal como un paciente que ha sido diagnosticado de o está en riesgo de desarrollar o adquirir cáncer de acuerdo con criterios de diagnóstico clínico aceptados en la técnica, tales como los del U.S. National Cancer Institute (Bethesda, MD, EE. UU.) o como describen DeVita, Hellman y Rosenberg en *Cancer: Principles and Practice of Oncology* (2008, Lippincott, Williams y Wilkins, Filadelfia/Ovid, Nueva York); Pizzo y Poplack, *Principles and Practice of Pediatric Oncology* (Cuarta edición, 2001, Lippincott, Williams y Wilkins, Filadelfia/Ovid, Nueva York); Vogelstein y Kinzler, *The Genetic Basis of Human Cancer* (Segunda edición, 2002, McGraw Hill Professional, Nueva York); Dancey et al. (2009 *Semin. Oncol.* 36 Supl.3: S46); ciertas realizaciones contemplan un sujeto humano que se sabe que está libre de riesgo de tener, desarrollar o adquirir cáncer de acuerdo con dichos criterios.

Ciertas otras realizaciones contemplan un sujeto no humano o una fuente biológica, por ejemplo, un primate no humano tal como un macaco, chimpancé, gorila, vervet, orangután, mandril u otro primate no humano, incluyendo tales sujetos no humanos que pueden ser conocidos en la técnica como modelos preclínicos, que incluyen modelos preclínicos para tumores sólidos y/u otros cánceres. Ciertas otras realizaciones contemplan un sujeto no humano que es un mamífero, por ejemplo, un ratón, rata, conejo, cerdo, oveja, caballo, bovino, cabra, jerbo, hámster, cobaya u otro mamífero; muchos de tales mamíferos pueden ser sujetos que se conocen en la técnica como modelos preclínicos para ciertas enfermedades o trastornos, incluyendo tumores malignos hematopoyéticos linfoides y/u otros cánceres (p.ej., Li et al., 2011 *Dis. Model. Mech.* 4:311; von Euler et al., 2011 *Vet. Comp. Oncol.* 9:1; Goldstein et al., 2010 *Expert Rev. Hematol.* 3:301; Diamond et al., 2009 *J. Bone Min. Res.* 24:1150; Macor et al., 2008 *Curr. Pharm. Des.* 14:2023; Talmadge et al., 2007 *Am. J. Pathol.* 170:793; Kerbel, 2003 *Cane. Biol. Therap.* 2 (4 Supl 1): S134; Man et al., 2007 *Cane. Met. Rev.* 26:737; Cespedes et al., 2006 *Clin. Transl. Oncol.* 8:318). Sin embargo, no se pretende que el intervalo de realizaciones esté tan limitado, de modo que también se contemplan otras realizaciones en las que el sujeto o la fuente biológica puede ser un vertebrado no mamífero, por ejemplo, otro vertebrado superior, o un anfibio aviar o especies de reptiles, u otro sujeto o fuente biológica.

Según se menciona también en otra parte de la presente memoria, se han establecido criterios de diagnóstico clínico aceptados en la técnica para estos y otros tipos de cáncer, tales como los promulgados por el U.S. National Cancer Institute (Bethesda, MD, USA) o como describen DeVita, Hellman, y Rosenberg en *Cancer: Principles and Practice of Oncology* (2008, Lippincott, Williams y Wilkins, Philadelphia/ Ovid, New York); Pizzo y Poplack, *Principles and Practice of Pediatric Oncology* (Fourth edition, 2001, Lippincott, Williams y Wilkins, Philadelphia/ Ovid, New York); y Vogelstein y Kinzler, *The Genetic Basis of Human Cancer* (Second edition, 2002, McGraw Hill Professional, New York). Por ejemplo Ignatiadis et al. (2008 *Pathobiol* 75:104); Kunz (2008 *Curr. Drug Discov. Technol.* 5:9); y Auman et al. (2008 *Drug Metab. Rev.* 40:303) describen otros ejemplos no limitantes de tipificación y caracterización de cánceres concretos.

Se pueden proporcionar muestras biológicas obteniendo una muestra de sangre, una muestra de biopsia, un explante de tejido, un cultivo de órgano, un fluido biológico o cualquier otro tejido o preparación celular de un sujeto o una fuente biológica. Las células B y las células T pueden así obtenerse de una muestra biológica, tal como de una variedad de muestras de tejidos y fluidos biológicos que incluyen médula ósea, timo, ganglios linfáticos, nódulos

linfáticos, tejidos periféricos y sangre, pero se accede más fácilmente a la sangre periférica. Se puede muestrear cualquier tejido periférico para detectar la presencia de células B y T y, por lo tanto, se contempla su uso en los métodos descritos en la presente memoria. Los tejidos y fluidos biológicos a partir de los cuales se pueden obtener células inmunitarias adaptativas incluyen, entre otros, piel, tejidos epiteliales, colon, bazo, secreción de la mucosa, mucosa oral, mucosa intestinal, mucosa vaginal o secreción vaginal, tejido cervical, ganglios, saliva, líquido cefalorraquídeo (LCR), médula ósea, sangre del cordón umbilical, suero, líquido seroso, plasma, linfa, orina, líquido ascítico, líquido pleural, líquido pericárdico, líquido peritoneal, líquido abdominal, medio de cultivo, medio de cultivo acondicionado o fluido de lavado. En ciertas realizaciones, células inmunitarias adaptativas (p.ej., células hematopoyéticas de linaje linfocítico tales como células T y células B) pueden aislarse a partir de una muestra de aféresis. Las muestras de sangre periférica pueden obtenerse por flebotomía de los sujetos. Las células mononucleares de sangre periférica (PBMC) se aíslan mediante mecanismos conocidos por los expertos en la técnica, p.ej., por separación de gradiente de densidad Ficoll-Hypaque®. En ciertas realizaciones, se utilizan PBMC completas para el análisis.

En ciertas realizaciones relacionadas, se pueden preparar preparaciones que comprenden predominantemente linfocitos (p.ej., células T y B) o que comprenden predominantemente células T o predominantemente células B, incluyendo preparaciones en las que están presentes células de un tumor maligno hematológico linfocítico, para su uso como una muestra biológica como se proporciona en la presente memoria, de acuerdo con lo establecido en el art. metodologías aceptadas. En otras realizaciones relacionadas, pueden aislarse subpoblaciones específicas de células T o B antes del análisis utilizando los métodos descritos en la presente memoria. Diversos métodos y kits disponibles comercialmente para aislar diferentes subpoblaciones de células T y B son conocidos en la técnica e incluyen, pero no se limitan a, selección de subconjuntos, separación de esferas inmunomagnéticas o clasificación de células mediante inmunocitometría de flujo utilizando anticuerpos específicos para uno o más de cualquiera de una variedad de marcadores de la superficie de células T y B conocidos. Los marcadores ilustrativos incluyen, pero no están limitados a, uno o una combinación de CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD25, CD28, CD45RO, CD45RA, CD54, CD62, CD62L, CDw137 (41BB), CD154, GITR, FoxP3, CD54 y CD28. Por ejemplo, y como es sabido por los expertos, se pueden utilizar los marcadores de superficie celular, tales como CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD45RA y CD45RO para determinar los linajes y subpoblaciones de T, B y monocitos. En citometría de flujo. De forma similar, los marcadores de dispersión de luz frontal, de dispersión lateral y/o de superficie celular tales como CD25, CD62L, CD54, CD137, CD154 se pueden utilizar para determinar el estado de activación y las propiedades funcionales de las células.

Las combinaciones ilustrativas útiles en algunos de los métodos descritos en la presente memoria pueden incluir CD8<sup>+</sup>CD45RO<sup>+</sup> (células T citotóxicas de memoria), CD4<sup>+</sup>CD45RO<sup>+</sup> (T colaboradoras de memoria), CD8<sup>+</sup>CD45RO<sup>-</sup> (CD8<sup>+</sup>CD62L<sup>+</sup>CD45RA<sup>+</sup> (células T citotóxicas de tipo no tratado previamente); CD4<sup>+</sup>CD25<sup>+</sup>CD62L<sup>hi</sup>GITR<sup>+</sup>FoxP3<sup>+</sup> (células T reguladoras). Los anticuerpos ilustrativos para su uso en separaciones de células inmunomagnéticas o clasificación de células por inmunocitometría de flujo incluyen anticuerpos antihumanos marcados fluorescentemente, p.ej., CD4 FITC (clon M-T466, Miltenyi Biotec), CD8 PE (clon RPA-T8, BD Biosciences), CD45RO ECD (clon UCHL-1, Beckman Coulter), y CD45RO APC (clon UCHL-1, BD Biosciences). La tinción de PBMC totales se puede realizar con la combinación apropiada de anticuerpos, seguida de lavado de las células antes del análisis. Los subconjuntos de linfocitos se pueden aislar por clasificación celular activada por fluorescencia (FACS), p.ej., mediante un sistema de clasificación de células BD FACSAria™ (BD Biosciences) y analizando los resultados con el soporte lógico FlowJo™ (Treestar Inc.), y también mediante métodos conceptualmente similares que implican anticuerpos específicos inmovilizados en superficies o esferas.

Para la extracción de ácidos nucleicos, el ADN genómico total se puede extraer de las células utilizando métodos conocidos en la técnica y/o kits disponibles comercialmente, p.ej., mediante el uso de QIAamp® DNA Blood Mini Kit (QIAGEN®). La masa aproximada de un solo genoma haploide es de 3 pg. Preferiblemente, se utilizan para el análisis de al menos 25.000 a 250.000 células, por ejemplo, de al menos 50.000 a 125.000 células, o de al menos 75.000 a 150.000 células, o de al menos 100.000 a 200.000 células, es decir., de aproximadamente 0,15 a 1,5 µg, o por ejemplo, de 0,6 a 1,2 µg de ADN de células T o B diploides. El número de células T o B presentes en una muestra puede variar considerablemente cuando la muestra se obtiene de un paciente que tiene un tumor maligno hematológico linfocítico, tal como leucemia linfoblástica de células T aguda (T-ALL). Usando PBMC de un ser humano adulto sano normal como fuente, se puede estimar que el número de células T es aproximadamente 30% del total de células; también se puede estimar que el número de células B es aproximadamente 10% del total de células en una preparación de PBMC.

### Receptores adaptativos de células inmunitarias

El TCR nativo es una proteína de superficie celular heterodimérica de la superfamilia de inmunoglobulinas que está asociada a proteínas invariables del complejo CD3 implicado en la mediación de la transducción de señales. Los TCR existen en formas αβ y γδ, que son estructuralmente similares pero tienen ubicaciones anatómicas y probablemente funciones bastante distintas. Los ligandos MHC clase I y clase II, que se unen al TCR, también son proteínas de la superfamilia de inmunoglobulinas pero están especializados para la presentación de antígenos, con

un sitio de unión de péptidos altamente polimórficos que les permite presentar una matriz diversa de fragmentos de péptidos cortos en la superficie de la célula APC.

Las porciones extracelulares de los TCR  $\alpha\beta$  y  $\gamma\delta$  heterodiméricos nativos consisten en dos polipéptidos, cada uno de los cuales tiene un dominio constante proximal a la membrana y un dominio variable distal a la membrana. Cada uno de los dominios constantes y variables incluye un enlace disulfuro intracatenario. Los dominios variables contienen los bucles altamente polimórficos análogos a las regiones determinantes de complementariedad (CDR) de los anticuerpos. Las CDR3 de los TCR  $\alpha\beta$  interactúan con el péptido presentado por MHC, y las CDR 1 y 2 de los TCR  $\alpha\beta$  interactúan con el péptido y el MHC. La diversidad de las secuencias de TCR se genera a través de la reorganización somática de los genes variable (V), de diversidad (D), de empalme (J) y constante.

Los loci del gen de Ig y TCR contienen muchos segmentos diferentes de genes variable (V), de diversidad (D) y de empalme (J), que se someten a procesos de reordenamiento durante la diferenciación linfocítica temprana. Las secuencias de segmentos génicos V, D y J de Ig y TCR, son conocidas en la técnica y están disponibles en bases de datos públicas tales como GENBANK. Los ejemplos no limitantes de secuencias de los segmentos génicos de la región V de TCRB se exponen en el listado de secuencias de SEQ ID NO: 5579-5630, 5644-5779, 6222-6273 y 6287-6351, y se exponen ejemplos de secuencias de segmento de la región J de TCRB en SEQ ID NO: 5631-5643, 5780-5792 y 6274-6286. Las secuencias del segmento génico de la región J de TCRG ilustrativas se exponen en SEQ ID NO: 5793-5798 y 6212-6215. Las secuencias del segmento génico de la región V de TCRG ilustrativas se exponen en SEQ ID NO: 5799-5816 y 6124-6127. Las secuencias del segmento génico de la región J de IgH ilustrativas se exponen en SEQ ID NO: 5817-5832 y 6216-6221; las secuencias del segmento génico de la región V de IgH se exponen en SEQ ID NO: 5833-6123 y 6128-6211.

Los reordenamientos VDJ están mediados a través de un complejo enzimático de recombinasa en el que las proteínas RAG1 y RAG2 juegan un papel clave al reconocer y cortar el ADN en las secuencias señal de recombinación (RSS), que se encuentran aguas abajo de los segmentos génicos V, a ambos lados de los segmentos génicos D, y aguas arriba de los segmentos génicos J. La RSS inapropiada reduce o incluso evita por completo el reordenamiento. La secuencia señal de recombinación (RSS) consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3' y nonámero, 5'-ACAAAAACC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Se han identificado varias posiciones de nucleótidos como importantes para la recombinación, incluyendo el dinucleótido de CA en la posición uno y dos del heptámero, y también se ha mostrado que una C en la posición tres del heptámero es muy preferible, así como un nucleótido A en las posiciones 5, 6, 7 del nonámero. (Ramsden et. al. 1994 Nucl. C.A. Res. 22:1785; Akamatsu et. al. 1994 J. Immunol. 153:4520; Hesse et. al. 1989 Genes Dev. 3:1053). Las mutaciones de otros nucleótidos tienen efectos mínimos o inconsistentes. El espaciador, aunque es más variable, también tiene un impacto sobre la recombinación, y se ha demostrado que las sustituciones de un solo nucleótido tienen un impacto significativo sobre la eficacia de la recombinación (Fanning et. al. 1996 Cell. Immunol. Immunopath. 79:1, Larijani et. al. 1999 Nucl. C.A. Res. 27:2304; Nadel et. al. 1998 J. Immunol. 161:6068; Nadel et al., 1998 J. Exp. Med. 187:1495). Se han descrito los criterios para identificar secuencias de polinucleótidos de RSS que tienen eficacias de recombinación significativamente diferentes (Ramsden et. al. 1994 Nucl. C.A. Res. 22:1785; Akamatsu et. al. 1994 J. Immunol. 153:4520; Hesse et. al. 1989 Genes Dev. 3:1053 y Lee et al., 2003 PLoS 1(1): E1).

El proceso de reordenamiento generalmente comienza con un reordenamiento de D a J seguido de un reordenamiento de V a DJ en el caso de genes de cadena pesada de Ig (IgH), TCR beta (TCRB) TCR y delta (TCRD) o afecta a los reordenamientos de V a J directos en el caso de genes Ig kappa (IgK), Ig lambda (IgL), TCR alfa (TCRA) y TCR gamma (TCRG). Las secuencias entre los segmentos génicos que se reordenan generalmente se eliminan en forma de un producto de escisión circular, también denominado círculo de escisión TCR (TREC) o círculo de escisión del receptor de células B (BREC).

Las muchas combinaciones diferentes de segmentos de genes V, D y J representan el llamado repertorio combinatorio, que se estima que es  $\sim 2 \times 10^6$  para moléculas de Ig,  $\sim 3 \times 10^6$  para TCR $\alpha\beta$  y  $\sim 5 \times 10^3$  para moléculas de TCR $\gamma\delta$ . En los sitios de unión de los segmentos génicos V, D y J, la delección y la inserción aleatoria de nucleótidos se produce durante el proceso de reordenamiento, lo que da como resultado regiones de empalme muy diversas, que contribuyen significativamente al repertorio total de moléculas de Ig y TCR, estimadas en  $> 10^{12}$ .

Los linfocitos B maduros amplían adicionalmente su repertorio de Ig tras el reconocimiento del antígeno en los centros del folículo mediante hipermutación somática, un proceso que conduce a la maduración por afinidad de las moléculas de Ig. El proceso de hipermutación somática se centra en el exón V- (D-) J de los genes de cadena ligera de IgH e Ig y se refiere a mutaciones de un único nucleótido y, a veces, también a inserciones o delecciones de nucleótidos. Los genes de Ig mutados somáticamente también se encuentran en tumores malignos de células B maduras de origen folicular o posfolicular.

En ciertas realizaciones preferidas descritas en la presente memoria, se pueden emplear cebadores del segmento V y del segmento J en una reacción de PCR para amplificar regiones de ADN que codifican CDR3 de TCR

reordenadas en una muestra biológica de prueba, en donde cada segmento génico funcional que codifica V de TCR comprende una secuencia de señal de recombinación (RSS) del gen V y cada segmento génico funcional que codifica J de TCR comprende una RSS del gen J. En estas realizaciones y en realizaciones relacionadas, cada molécula de ADN reordenado amplificado puede comprender (i) al menos aproximadamente 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR, estando situados al menos aproximadamente 10, 20, 30 o 40 nucleótidos contiguos 5' con respecto a la RSS del gen V y/o cada molécula de ADN reordenado amplificado puede comprender (ii) al menos aproximadamente 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR, estando situados al menos aproximadamente 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J. En ciertas realizaciones preferidas, cada región codificante de CDR3 de TCR amplificada está presente en una molécula de ADN reordenado amplificado que tiene una longitud inferior a 600 nucleótidos. Sin desear estar limitados por la teoría, estas características de diseño para amplificar regiones de empalme VJ que codifican CDR3 permiten la hibridación de cebadores del segmento V a sustancialmente todos los segmentos génicos funcionales que codifican V de TCR, y también permiten la hibridación de cebadores del segmento J a sustancialmente todos los segmentos funcionales que codifican J de TCR, y también permiten la amplificación de regiones que codifican CDR3 que son susceptibles de secuenciarse mediante las plataformas HTS descritas en la presente memoria, al tiempo que incluyen información de secuencia adecuada para identificar todas las posibles combinaciones VDJ y VJ.

### PCR cuantitativa múltiple

Como se describe en la presente memoria y en vista de Robins et al., 2009 Blood 114, 4099; Robins et al., 2010 Sci. Translat. Med. 2:47ra64; Robins et al., 2011 J. Immunol. Meth. doi:10.1016/j.jim.2011.09.001; Sherwood et al. 2011 Sci. Translat. Med. 3:90ra61; la Publicación de Estados Unidos Núm. 2012/0058902, la Publicación de Estados Unidos Núm. 2010/0330571, el documento WO/2010/151416, el documento WO/2011/106738 y el documento WO2012/027503, se proporciona un método para cuantificar la representación relativa del ADN de células inmunitarias adaptativas en ADN a partir de una muestra biológica de prueba, y así determinar la frecuencia de aparición relativa para cada secuencia de ADN que codifica TCR reordenada única en una muestra que contiene células linfoides. De acuerdo con ciertas realizaciones preferidas, el método implica un método de PCR múltiple que utiliza un conjunto de cebadores directos que hibridan específicamente con los segmentos V y un conjunto de cebadores inversos que hibridan específicamente con los segmentos J, donde la reacción de PCR múltiple permite la amplificación de todas las combinaciones VJ (y VDJ) posibles dentro de una población dada de células T o B.

Se puede extraer ADN o ARN de células en una muestra, tal como una muestra de sangre o linfa u otra muestra de un sujeto que se sabe que tiene o se sospecha que tiene un tumor maligno hematológico linfóide, utilizando métodos convencionales o kits comercialmente disponibles conocidos en la técnica. Se puede utilizar un sistema de PCR múltiple para amplificar los loci del receptor de células inmunitarias adaptadas reordenadas a partir de ADN genómico, preferiblemente de una región CDR3. En ciertas realizaciones, la región CDR3 se amplifica a partir de una región CDR3 de TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$  o TCR $\delta$ . Se proporcionan composiciones que comprenden una pluralidad de cebadores del segmento V y del segmento J que son capaces de promover la amplificación en una reacción en cadena de polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 del receptor inmunitario adaptativo reordenadas productivamente en la muestra para una clase dada de tales receptores (p.ej., TCR $\gamma$ , TCR $\beta$ , etc.), para producir una multiplicidad de moléculas de ADN reordenado amplificado de una población de células T (para TCR) en la muestra. Preferiblemente y en ciertas realizaciones, los cebadores se diseñan de modo que cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud, excluyendo de este modo los productos de amplificación de loci adaptativos del receptor inmunitario no reordenados.

En el genoma humano actualmente se cree que hay aproximadamente 70 segmentos de genes V $\alpha$  de TCR y aproximadamente 61 J $\alpha$ , aproximadamente 52 segmentos de genes V $\beta$  de TCR, aproximadamente 2 D $\beta$  y aproximadamente 13 J $\beta$ , aproximadamente 9 segmentos de genes V $\gamma$  de TCR y aproximadamente 5 J $\gamma$ , y aproximadamente 46 cadena pesada de inmunoglobulina (IGH) V $H$ , aproximadamente 23 segmentos de genes D $H$  y aproximadamente 6 J $H$ . Por consiguiente, cuando se conocen secuencias genómicas para estos loci de manera que se pueden producir fácilmente sondas moleculares específicas para cada uno de ellos, se cree de acuerdo con la teoría no limitante que las presentes composiciones y métodos se refieren sustancialmente a todos (p.ej., más de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99%) estos segmentos génicos que codifican la región V, D y J de receptores inmunitarios adaptativos conocidos y fácilmente detectables.

Los genes de TCR e Ig pueden generar millones de proteínas distintas a través de mutación somática. Debido a este mecanismo generador de diversidad, las regiones determinantes de complementariedad hipervariables (CDR) de estos genes pueden codificar secuencias que pueden interactuar con millones de ligandos, y estas regiones están unidas a una región constante que puede transmitir una señal a la célula que indica la unión del ligando cognado de la proteína. El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígenos de células T y B con diversidad suficiente para reconocer el universo de posibles patógenos. En las células T  $\alpha\beta$  y  $\gamma\delta$ , que reconocen principalmente antígenos peptídicos presentados por moléculas del MHC, la mayor



parte de esta diversidad de receptores está contenida dentro de la tercera región determinante de complementariedad (CDR3) de las cadenas  $\alpha$  y  $\beta$  del receptor de células T (o cadenas  $\gamma$  y  $\delta$ )

5 La tecnología de ensayo utiliza dos conjuntos de cebadores para proporcionar una reacción de PCR altamente multiplexada. El primer agrupamiento "directo" (p.ej., a modo de ilustración y no de limitación, los cebadores oligonucleotídicos del segmento V descritos en la presente memoria se pueden utilizar en ciertas realizaciones preferidas como cebadores "directos" cuando los cebadores oligonucleotídicos del segmento J se utilizan como cebadores "inversos" de acuerdo con la terminología de PCR comúnmente utilizada, pero el experto en la técnica apreciará que en ciertas otras realizaciones los cebadores del segmento J pueden considerarse cebadores "directos" cuando se utilizan con cebadores "inversos" del segmento V) incluye un cebador oligonucleotídico que es específico para (p.ej., que tiene una secuencia de nucleótidos complementaria a una región de secuencia única de) cada segmento codificante de la región V ("segmento V) en el locus del gen de TCR o Ig respectivo. En ciertas realizaciones, se utilizan cebadores que se dirigen a una región altamente conservada, para capturar simultáneamente muchos segmentos V, reduciendo así la cantidad de cebadores requeridos en la PCR múltiple. De manera similar, en ciertas realizaciones, los cebadores de grupo "inverso" hibridan con una secuencia conservada en el segmento de empalme ("J").

20 Cada cebador puede diseñarse de manera que se obtenga un segmento de ADN amplificado respectivo que incluya una porción de secuencia de longitud suficiente para identificar cada segmento J inequívocamente basándose en las diferencias de secuencia entre segmentos de genes codificantes de la región J en la base de datos del genoma humano, y también incluya una porción de secuencia con la que un cebador específico del segmento J pueda hibridar para su resecuenciación. Este diseño de cebadores específicos del segmento V y J permite la observación directa de una gran fracción de los reordenamientos somáticos presentes en el repertorio del gen del receptor inmunitario adaptativo dentro de un individuo. Esta característica a su vez permite la comparación rápida de los repertorios de TCR y/o Ig (i) en individuos que tienen una enfermedad, trastorno, afección u otra indicación de interés en particular (p.ej., cáncer, una enfermedad autoinmunitaria, un trastorno inflamatorio u otra afección) con (ii) los repertorios de TCR y/o Ig de sujetos de control que están libres de tales enfermedades, trastornos, afecciones o indicaciones.

30 El término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica tal como la totalidad o una porción de un polipéptido de TCR o Ig (p.ej., un polipéptido que contiene CDR3); incluye regiones que preceden y siguen a la región codificante "líder y trailer", así como secuencias intermedias (intrones) entre segmentos codificantes individuales (exones), y también puede incluir elementos reguladores (p.ej., promotores, potenciadores, sitios de unión represores y similares), y también puede incluir secuencias señal de recombinación (RSS) como se describe en la presente memoria.

40 Los ácidos nucleicos para su uso en las presentes realizaciones, a los que también se hace referencia en la presente memoria como polinucleótidos, y que incluyen oligonucleótidos, pueden estar en forma de ARN o en forma de ADN, cuyo ADN incluye ADNc, ADN genómico y ADN sintético. El ADN puede ser de doble hebra o de hebra sencilla, y si es de hebra sencilla puede ser la hebra codificante o la hebra no codificante (antisentido). Una secuencia codificante que codifica un TCR o una región del mismo (p.ej., una región V, un segmento D, una región J, una región C, etc.) puede ser idéntica a la secuencia codificante conocida en la técnica para cualquier región de gen de TCR o dominios de polipéptido dados (p.ej., dominios de la región V, dominios CDR3, etc.), o puede ser una secuencia codificante diferente, que, como resultado de la redundancia o degeneración del código genético, codifica la misma región o polipéptido de TCR.

50 En una realización, los métodos pueden utilizar una pluralidad de cebadores del segmento V y una pluralidad de cebadores del segmento J, en donde la pluralidad de cebadores del segmento V y la pluralidad de cebadores del segmento J amplifican sustancialmente todas las combinaciones de los segmentos V y J de un locus del receptor inmunitario reordenado. Por sustancialmente todas las combinaciones se entiende al menos 95%, 96%, 97%, 98%, 99% o más de todas las combinaciones de los segmentos V y J de un locus del receptor inmunitario reordenado. En ciertas realizaciones, la pluralidad de cebadores del segmento V y la pluralidad de cebadores del segmento J amplifican todas las combinaciones de los segmentos V y J de un locus del receptor inmunitario reordenado.

55 En general, un sistema de PCR múltiple puede utilizar al menos 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, y en ciertas realizaciones, al menos 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38 o 39, y en otras realizaciones 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 65, 70, 75, 80, 85, o más cebadores directos, en el que cada cebador directo hibrida específicamente o es complementario a una secuencia correspondiente a uno o más segmentos de la región V. Los cebadores de la región V ilustrativos para la amplificación de TCR $\beta$  se muestran en SEQ ID NO: 5579-5630. Los cebadores de la región V de TCR $\gamma$  ilustrativos se proporcionan en SEQ ID NO: 6124-6127. Los cebadores de la región V de IgH ilustrativos se proporcionan en SEQ ID NO: 6128-6211. Por lo tanto, las secuencias del segmento génico de la región V se pueden utilizar para diseñar cebadores de la región V. Las secuencias del segmento génico de la región V de TCRB ilustrativos se exponen en la lista de secuencias en SEQ ID NO: 5579-5630, 5644-5779, 6222-6273 y 6287-6351. Las secuencias

del segmento génico de la región V de TCRG ilustrativos se exponen en SEQ ID NO: 5799-5816 y 6124-6127. Las secuencias del segmento génico de la región V de IgH ilustrativas se exponen en SEQ ID NO: 5833-6123 y 6128-6211. En SEQ ID NO: 5579-5643, en los oligonucleótidos RN2, "r" representa una base ribonucleotídica en la secuencia oligonucleotídica y "/3SpC3/" representa un espaciador 3' de tres carbonos en el grupo hidroxilo, que evita la extensión y amplificación por la polimerasa. La endonucleasa de reparación del ADN escinde el oligonucleótido en el ribonucleótido después de la hibridación con una secuencia complementaria, creando un grupo hidroxilo no bloqueado que se puede extender mediante una polimerasa.

El sistema de PCR múltiple también usa al menos 3, 4, 5, 6 o 7, y en ciertas realizaciones, 8, 9, 10, 11, 12 o 13 cebadores inversos, en los que cada cebador inverso hibrida específicamente con o es complementario a una secuencia correspondiente a uno o más segmentos de la región J. Los cebadores ilustrativos del segmento J de TCRβ se proporcionan en SEQ ID NO: 5631-5643. Los cebadores del segmento TCRγ J ilustrativos se proporcionan en SEQ ID NO: 6212-6215. Los cebadores del segmento J de IgH ilustrativos se proporcionan en SEQ ID NO: 6216-6221. Por lo tanto, las secuencias del segmento génico de la región J se pueden utilizar para diseñar cebadores de la región J. Las secuencias de segmentos de regiones J de TCRB ilustrativas se exponen en SEQ ID NO: 5631-5643, 5780-5792 y 6274-6286. Las secuencias del segmento génico de la región J de TCRG ilustrativas se exponen en SEQ ID NO: 5793-5798 y 6212-6215. Las secuencias del segmento génico de la región J de IgH ilustrativas se exponen en SEQ ID NO: 5817-5832 y 6216-6221. En una realización, existe un cebador del segmento J para cada segmento J.

Los oligonucleótidos o polinucleótidos que son capaces de hibridar o reasociarse específicamente con una secuencia de ácido nucleico diana mediante la complementariedad de bases de nucleótidos pueden hacerlo bajo condiciones de rigurosidad de moderada a alta. Con fines ilustrativos, las condiciones de rigurosidad moderada a alta adecuadas para la amplificación por PCR específica de una secuencia de ácido nucleico diana serían de entre 25 y 80 ciclos de PCR, consistiendo cada ciclo en una etapa de desnaturalización (p.ej., aproximadamente 10-30 segundos a más de aproximadamente 95°C), una etapa de reasociación (p.ej., aproximadamente 10-30 s aproximadamente a 60-68°C), y una etapa de extensión (p.ej., aproximadamente 10-60 s a aproximadamente 60-72°C), opcionalmente de acuerdo con ciertas realizaciones, combinando las etapas de recocido y extensión para proporcionar una PCR en dos etapas. Como reconocería el experto en la técnica, se pueden añadir o cambiar otros reactivos de PCR en la reacción de PCR para aumentar la especificidad del emparejamiento del cebador y la amplificación, tal como alterando la concentración de magnesio, opcionalmente añadiendo DMSO, y/o utilizando cebadores bloqueados, nucleótidos modificados, ácidos péptidonucleicos y similares.

En ciertas realizaciones, se pueden utilizar técnicas de hibridación de ácidos nucleicos para evaluar la especificidad de hibridación de los cebadores descritos en la presente memoria. Las técnicas de hibridación son bien conocidas en la técnica de la biología molecular. Con fines de ilustración, las condiciones moderadamente restrictivas adecuadas para someter a ensayo la hibridación de un polinucleótido como se proporciona en la presente memoria con otros polinucleótidos incluyen prelavado en una solución de 5 X SSC, SDS al 0,5%, EDTA 1,0 mM (pH 8,0); hibridación a 50°C-60°C, 5 X SSC, durante la noche; seguido de un lavado dos veces a 65°C durante 20 minutos con cada uno de 2X, 0,5X y 0,2X SSC que contiene SDS al 0,1%. Un experto en la técnica entenderá que la rigurosidad de la hibridación puede manipularse fácilmente, por ejemplo alterando el contenido de sal de la solución de hibridación y/o la temperatura a la que se realiza la hibridación. Por ejemplo, en otra realización, las condiciones de hibridación altamente rigurosas adecuadas incluyen las descritas anteriormente, con la excepción de que se aumenta la temperatura de hibridación, p.ej., a 60°C-65°C o 65°C-70°C.

CEBADORES. De acuerdo con la presente descripción, se proporcionan cebadores oligonucleotídicos en un conjunto de cebadores oligonucleotídicos que comprende una pluralidad de cebadores del segmento V y una pluralidad de cebadores del segmento J, donde el conjunto de cebadores es capaz de amplificar el ADN reordenado que codifica receptores inmunitarios adaptativos en una muestra biológica que comprende ADN de células linfoides. En la técnica se conocen conjuntos de cebadores adecuados y se describen en esta memoria, por ejemplo, los conjuntos de cebadores de los documentos U.S.A.N. 13/217,126; U.S.A.N. 12/794,507; PCT/US2011/026373; o PCT/US2011/049012; o similares; o aquellos mostrados en la Tabla 1. En ciertas realizaciones, el conjunto de cebadores está diseñado para incluir una pluralidad de cebadores específicos de secuencia V que incluye, para cada gen de la región V único (incluyendo pseudogenes) en una muestra, al menos un cebador que se puede reasociar específicamente con una secuencia de la región V única; y para cada gen único de la región J en la muestra, al menos un cebador que pueda reasociarse específicamente con una secuencia de la región J única.

El diseño del cebador puede lograrse mediante metodologías rutinarias a la vista de las secuencias genómicas conocidas de TCR y BCR. Por consiguiente, el conjunto de cebadores es preferiblemente capaz de amplificar todas las posibles combinaciones V-J que pueden resultar de reordenamientos de ADN en el locus de TCR o BCR. Como también se describe a continuación, ciertas realizaciones contemplan conjuntos de cebadores en los cuales uno o más cebadores para V pueden ser capaces de reasociarse específicamente con una secuencia "única" que puede ser compartida por dos o más regiones V pero que no es común a todas las regiones V, y/o en donde uno o más cebadores J pueden ser capaces de reasociarse específicamente con una secuencia "única" que puede ser

compartida por dos o más regiones J pero que no es común a todas las regiones J.

En realizaciones concretas, los cebadores oligonucleotídicos para su uso en las composiciones y métodos descritos en la presente memoria pueden comprender o consistir en un ácido nucleico de al menos aproximadamente 15 nucleótidos de longitud que tiene la misma secuencia que, o es complementario a, una secuencia contigua larga de 15 nucleótidos del segmento V o J (es decir, porción de polinucleótido genómico que codifica un polipéptido de la región V o de la región J). También serán útiles en ciertas realizaciones cebadores más largos, p.ej., los de aproximadamente 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, 45 o 50, nucleótidos de longitud que tienen la misma secuencia, o una secuencia complementaria a, una secuencia contigua del segmento de polinucleótido que codifica la región V o J. Todas las longitudes intermedias de los cebadores oligonucleotídicos descritos en la presente memoria se contemplan para su uso en la presente memoria. Como reconocería el experto en la técnica, los cebadores pueden tener una secuencia adicional añadida (p.ej., nucleótidos que pueden no ser iguales o complementarios al segmento polinucleotídico que codifica la región J o la región diana), tales como sitios de reconocimiento para enzimas de restricción, secuencias adaptadoras para la secuenciación, secuencias código de barras, y similares (véase p.ej., secuencias de cebadores proporcionadas en las Tablas y en la lista de secuencias de la presente memoria). Por lo tanto, la longitud de los cebadores puede ser más larga, tal como de aproximadamente 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 80, 85, 90, 95, 100 o más nucleótidos de longitud o más, dependiendo del uso específico o la necesidad.

También se contemplan para su uso en ciertas realizaciones variantes del cebador oligonucleotídico del segmento V o del segmento J del receptor inmunitario adaptativo que pueden compartir un alto grado de identidad de secuencia con los cebadores oligonucleotídicos para los que se presentan en la presente memoria secuencias de nucleótidos, incluyendo los establecidos en la Lista de Secuencias. Por lo tanto, en estas realizaciones y en realizaciones relacionadas, las variantes del cebador oligonucleotídico adaptativo del segmento V o del segmento J pueden tener una identidad sustancial con las secuencias de cebadores oligonucleotídicos del segmento V o del segmento J del receptor inmunitario adaptativo descritas en la presente memoria, por ejemplo, tales variantes del cebador oligonucleotídico pueden comprender al menos 70% de identidad de secuencia, preferiblemente al menos 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99 % o mayor identidad de secuencia en comparación con una secuencia de polinucleótidos de referencia tal como las secuencias de oligonucleótidos cebadores descritas en la presente memoria, utilizando los métodos descritos en la presente memoria (p.ej., análisis BLAST utilizando parámetros normalizados). Un experto en esta técnica reconocerá que estos valores pueden ajustarse apropiadamente para determinar la capacidad correspondiente de una variante de cebador oligonucleotídico para reasociarse con un polinucleótido que codifica un segmento de receptor inmunitario adaptativo teniendo en cuenta la degeneración de codones, el posicionamiento de los marcos de lectura y similares.

Típicamente, las variantes de cebador oligonucleotídico contendrán una o más sustituciones, adiciones, deleciones y/o inserciones, preferiblemente de manera que la capacidad de emparejamiento del oligonucleótido variante no disminuya sustancialmente con respecto a la de una secuencia de cebador oligonucleotídico del segmento V o segmento J del receptor inmunitario adaptativo que se muestra específicamente en la presente memoria.

La Tabla 1 presenta como ejemplo no limitante un conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado productivamente que codifica cadenas  $\beta$  de TCR (TCRB) en una muestra biológica que comprende ADN de células linfoides de un sujeto. En este conjunto de cebadores, los cebadores del segmento J comparten una homología de secuencia sustancial, y por lo tanto pueden cebar de manera cruzada más de una secuencia polinucleotídica J diana, pero los cebadores del segmento V están diseñados para reasociarse específicamente con secuencias diana dentro de la región CDR2 de V y son por lo tanto únicos para cada segmento V. Sin embargo, una excepción está presente en el caso de varios cebadores V en los que las secuencias dentro de la familia de los genes diana estrechamente relacionados son idénticas (p.ej., V6-2 y V6-3 son idénticos a nivel de nucleótidos en toda la secuencia codificante del segmento V, y por lo tanto pueden tener un único cebador, TRB2V6-2/3).

Por lo tanto, se apreciará que en ciertas realizaciones la cantidad de oligonucleótidos molde diferentes en la composición molde, y/o la cantidad de cebadores oligonucleotídicos diferentes en el conjunto de cebadores, se puede reducir ventajosamente diseñando molde y/o cebadores para explotar ciertas similitudes conocidas en secuencias V y/o J. Por lo tanto, en estas realizaciones y realizaciones relacionadas, las secuencias de oligonucleótidos "únicas" como se describe en la presente memoria pueden incluir secuencias de polinucleótidos V específicas que son compartidas por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 oligonucleótidos molde distintos y/o secuencias de polinucleótidos J específicas compartidas por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 oligonucleótidos molde distintos, donde las secuencia de tales moldes difieren entre sí en otras secuencias distintas de las secuencias V y/o J compartidas.

De acuerdo con ciertas realizaciones contempladas actualmente, puede ser útil disminuir (p.ej., reducir de una manera estadísticamente significativa) el sesgo de amplificación del molde tal como el potencial de amplificación de ácido nucleico no uniforme entre los miembros de un conjunto de cebadores de amplificación que pueden resultar de

eficacias de cebador desiguales (p.ej., utilización de cebadores desiguales) solo para un subconjunto limitado de todos los genes V y J de origen natural. Por ejemplo, en análisis del repertorio inmunitario de TCR o BCR implicado en una respuesta inmunitaria, ya sea a un antígeno específico, como en una vacuna, o a un tejido, como en una enfermedad autoinmunitaria, solo pueden interesar los reordenamientos productivos de TCR o IG. En tales circunstancias, puede ser económicamente ventajoso identificar y corregir el potencial de amplificación de ácidos nucleicos no uniforme solo para aquellos cebadores del segmento V y J que contribuyen a reordenamientos productivos de ADN que codifica TCR o BCR, y excluir los esfuerzos para corregir la amplificación no uniforme de pseudogenes y orfones (es decir, segmentos codificantes de la región V de TCR o BCR que se han duplicado en otros cromosomas).

En el locus de IGH humano, por ejemplo, la base de datos ImmunoGeneTics (IMGT) (M.-P. LeFranc, Université Montpellier, Montpellier, Francia; [www.imgt.org](http://www.imgt.org)) anota 165 genes del segmento V, de los cuales 26 son orfones en otros cromosomas y 139 están en el locus IGH en el cromosoma 14. Entre los 139 segmentos V dentro del locus de IGH, 51 tienen al menos un alelo funcional, mientras que 6 son ORF (marcos de lectura abiertos) a los que les falta al menos un residuo de aminoácido altamente conservado, y 81 son pseudogenes. Los pseudogenes pueden incluir segmentos V que contienen un codón de terminación dentro del marco dentro de la secuencia codificante del segmento V, un desplazamiento del marco entre el codón de inicio y la secuencia codificante CDR3, una o más inserciones de elementos repetidos y deleciones de regiones críticas, tales como primer exón o la RSS. Para caracterizar reordenamientos de IGH funcionales en una muestra mientras se evita el tiempo y el gasto de caracterizar pseudogenes y/u orfones, se contempla por lo tanto utilizar un subconjunto de los oligonucleótidos molde sintéticos descritos en la presente memoria que está diseñado para incluir solo aquellos segmentos V que participan en un reordenamiento funcional para codificar un TCR o BCR, sin tener que sintetizar ni calibrar cebadores de amplificación y oligonucleótidos molde específicos de las secuencias de pseudogen. De este modo se obtienen eficacias ventajosas con respecto, entre otros, a tiempo y gastos.

Tabla 1. Conjunto de cebadores oligonucleotídicos ilustrativos (cebadores de PCR hsTCRB)

Nombre	Secuencia	SEQ ID NO:
TRBJ1-1	TTACCTACAACACTGTGAGTCTGGTGCCTTGCCAAA	1631
TRBJ1-2	ACCTACAACGGTTAACCTGGTCCCCGAACCGAA	1632
TRBJ1-3	ACCTACAACAGTGAGCCAACCTCCCTCTCCAAA	1633
TRBJ1-4	CCAAGACAGAGAGCTGGGTTCCACTGCCAAA	1634
TRBJ1-5	ACCTAGGATGGAGAGTCGAGTCCCATCACAAA	1635
TRBJ1-6	CTGTCACAGTGAGCCTGGTCCC GTTCCAAA	1636
TRBJ2-1	CGGTGAGCCGTGTCCCTGGCCCGAA	1637
TRBJ2-2	CCAGTACGGTCAGCCTAGAGCCTTCTCCAAA	1638
TRBJ2-3	ACTGTCAGCCGGTGCCTGGGCCAAA	1639
TRBJ2-4	AGAGCCGGGTCCCGGCGCCGAA	1640
TRBJ2-5	GGAGCCGCGTGCCTGGCCCGAA	1641
TRBJ2-6	GTCAGCCTGCTGCCGGCCCCGAA	1642
TRBJ2-7	GTGAGCCTGGTGCCCGGCCGAA	1643
TRB2V10-1	AACAAAGGAGAAGTCTCAGATGGCTACAG	1644
TRB2V10-2	GATAAAGGAGAAGTCCCCGATGGCTATGT	1645
TRB2V10-3	GACAAAGGAGAAGTCTCAGATGGCTATAG	1646
TRB2V6-2/3	GCCAAAGGAGAGGTCCCTGATGGCTACAA	1647

ES 2 683 037 T3

Nombre	Secuencia	SEQ ID NO:
TRB2V6-8	CTCTAGATTAAACACAGAGGATTTCCCAC	1648
TRB2V6-9	AAGGAGAAGTCCCCGATGGCTACAATGTA	1649
TRB2V6-5	AAGGAGAAGTCCCCAATGGCTACAATGTC	1650
TRB2V6-6	GACAAAGGAGAAGTCCCGAATGGCTACAAC	1651
TRB2V6-7	GTTCCCAATGGCTACAATGTCTCCAGATC	1652
TRB2V6-1	GTCCCAATGGCTACAATGTCTCCAGATT	1653
TRB2V6-4	GTCCCTGATGGTTATAGTGTCTCCAGAGC	1654
TRB2V24-1	ATCTCTGATGGATACAGTGTCTCTCGACA	1655
TRB2V25-1	TTTCCTCTGAGTCAACAGTCTCCAGAATA	1656
TRB2V27	TCCTGAAGGGTACAAAGTCTCTCGAAAAG	1657
TRB2V26	CTCTGAGAGGTATCATGTTTCTTGAATA	1658
TRB2V28	TCCTGAGGGGTACAGTGTCTCTAGAGAGA	1659
TRB2V19	TATAGCTGAAGGGTACAGCGTCTCTCGGG	1660
TRB2V4-1	CTGAATGCCCCAACAGCTCTCTTTAAAC	1661
TRB2V4-2/3	CTGAATGCCCCAACAGCTCTCACTTATTC	1662
TRB2V2P	CCTGAATGCCCTGACAGCTCTCGCTTATA	1663
TRB2V3-1	CCTAAATCTCCAGACAAAGCTCACTTAAA	1664
TRB2V3-2	CTCACCTGACTCTCCAGACAAAGCTCAT	1665
TRB2V16	TTCAGCTAAGTGCCTCCCAAATTCACCCT	1666
TRB2V23-1	GATTCTCATCTCAATGCCCAAGAACGC	1667
TRB2V18	ATTTTCTGCTGAATTTCCCAAAGAGGGCC	1668
TRB2V17	ATTCACAGCTGAAAGACCTAACGGAACGT	1669
TRB2V14	TCTTAGCTGAAAGGACTGGAGGGACGTAT	1670
TRB2V2	TTCGATGATCAATTCTCAGTTGAAAGGCC	1671
TRB2V12-1	TTGATTCTCAGCACAGATGCCTGATGT	1672
TRB2V12-2	GCGATTCTCAGCTGAGAGGCCTGATGG	1673
TRB2V12-3/4	TCGATTCTCAGCTAAGATGCCTAATGC	1674
TRB2V12-5	TTCTCAGCAGAGATGCCTGATGCAACTTTA	1675
TRB2V7-9	GGTTCTCTGCAGAGAGGCCTAAGGGATCT	1676

Nombre	Secuencia	SEQ ID NO:
TRB2V7-8	GCTGCCCAGTGATCGCTTCTTTGCAGAAA	1677
TRB2V7-4	GGCGGCCCCAGTGGTCGGTTCTCTGCAGAG	1678
TRB2V7-6/7	ATGATCGGTTCTCTGCAGAGAGGCCTGAGG	1679
TRB2V7-2	AGTGATCGCTTCTCTGCAGAGAGGACTGG	1680
TRB2V7-3	GGCTGCCCAACGATCGGTTCTTTGCAGT	1681
TRB2V7-1	TCCCCGTGATCGGTTCTCTGCACAGAGGT	1682
TRB2V11-123	CTAAGGATCGATTTTCTGCAGAGAGGCTC	1683
TRB2V13	CTGATCGATTCTCAGCTCAACAGTTCAGT	1684
TRB2V5-1	TGGTCGATTCTCAGGGCGCCAGTTCTCTA	1685
TRB2V5-3	TAATCGATTCTCAGGGCGCCAGTTCATG	1686
TRB2V5-4	TCCTAGATTCTCAGGTCTCCAGTTCCTA	1687
TRB2V5-8	GGAAACTCCCTCCTAGATTTTCAGGTGCG	1688
TRB2V5-5	AAGAGGAAACTCCCTGATCGATTCTCAGC	1689
TRB2V5-6	GGCAACTCCCTGA TCGA TTCTCAGGTCA	1690
TRB2V9	GTTCCCTGACTTGCACTCTGAACTAAAC	1691
TRB2V15	GCCGAACACTTCTTTCTGCTTTCTTGAC	1692
TRB2V30	GACCCAGGACCGGCAGTTCATCCTGAGT	1693
TRB2V20-1	ATGCAAGCCTGACCTTGCCACTCTGACA	1694
TRB2V29-1	CATCAGCCGCCCAAACCTAACATTCTCAA	1695

En ciertas realizaciones preferidas, los cebadores oligonucleotídicos del segmento V y del segmento J tal como se describe en la presente memoria están diseñados para incluir secuencias de nucleótidos de modo que la información adecuada esté presente dentro de la secuencia de un producto de amplificación de un gen del receptor inmunitario adaptativo (TCR) reordenado para identificar únicamente los genes tanto específicos de V como específicos de J que dan lugar al producto de amplificación en el locus del receptor inmunitario adaptativo reordenado (p.ej., al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 pares de bases aguas arriba de la secuencia señal de recombinación (RSS) del gen V, preferiblemente al menos aproximadamente 22, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 39 o 40 pares de bases aguas arriba de la secuencia señal de recombinación (RSS) del gen V, y en ciertas realizaciones preferidas más de 40 pares de bases de secuencia aguas arriba de la secuencia de señal de recombinación (RSS) del gen V, y al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 pares de bases aguas abajo de la RSS del gen J, preferiblemente al menos aproximadamente 22, 24, 26, 28 o 30 pares de bases aguas abajo de la RSS del gen J, y en ciertas realizaciones preferidas, más de 30 pares de bases aguas abajo de la RSS del gen J).

Esta característica contrasta con los cebadores oligonucleotídicos descritos en la técnica para la amplificación de secuencias génicas que codifican TCR o que codifican Ig, que se basan principalmente en la reacción de amplificación simplemente para detectar la presencia o ausencia de productos de tamaños apropiados para segmentos V y J (p.ej., la presencia en productos de reacción de PCR de un amplicón de un tamaño concreto indica la presencia de un segmento V o J pero no proporciona la secuencia del producto de PCR amplificado y por lo tanto no confirma su identidad, tal como la práctica común de espectratipificación). En ciertas realizaciones, los cebadores están diseñados para no cruzar un límite intrón/exón. Los cebadores directos

- en ciertas realizaciones se reasocian con los segmentos V en una región de conservación de secuencia relativamente fuerte entre segmentos V con el fin de maximizar la conservación de la secuencia entre estos cebadores. Por consiguiente, esto minimiza el potencial de propiedades de reasociación diferencial de cada cebador, y de modo que la región amplificada entre los cebadores V y J contiene suficiente información de secuencia V de TCR o Ig para identificar el segmento específico del gen V utilizado. En una realización, los cebadores del segmento J hibridan con un elemento conservado del segmento J y tienen una resistencia a la reasociación similar. En una realización concreta, los cebadores del segmento J se reasocian con el mismo motivo de la región marco conservada.
- 5
- 10 Se pueden preparar oligonucleótidos (p.ej., cebadores) mediante cualquier método adecuado, incluyendo la síntesis química directa mediante un método tal como el método del fosfotriéster de Narang et al., 1979, Meth. Enzymol. 68:90-99; el método del fosfodiéster de Brown et al., 1979, Meth. Enzymol. 68:109-151; el método de dietilfosforamida de Beaucage et al., 1981, Tetrahedron Lett. 22:1859-1862; y el método de soporte sólido de la Patente de Estados Unidos Núm. 4.458.066, cada una incorporada a la presente memoria como referencia. Una revisión de los métodos de síntesis de productos conjugados de oligonucleótidos y nucleótidos modificados es proporcionada por Goodchild, 1990, en Bioconjugate Chemistry 1(3):165-187, incorporada a la presente memoria como referencia.
- 15
- 20 El término "cebador", según se usa en la presente memoria, se refiere a un oligonucleótido capaz de actuar como un punto de iniciación de la síntesis de ADN en condiciones adecuadas. Tales condiciones incluyen aquellas en las que se induce la síntesis de un producto de extensión de cebador complementario a una hebra de ácido nucleico en presencia de cuatro nucleósidos trifosfato diferentes y un agente de extensión (p.ej., una ADN polimerasa o transcriptasa inversa) en un tampón apropiado y a una temperatura adecuada.
- 25
- 30 Un cebador es preferiblemente un ADN de hebra sencilla, pero en algunas realizaciones también puede incluir ARN. La longitud apropiada de un cebador depende del uso pretendido del cebador, pero típicamente varía de 6 a 50 nucleótidos, o en ciertas realizaciones, de 15-35 nucleótidos. Las moléculas de cebador cortas generalmente requieren temperaturas más frías para formar complejos híbridos suficientemente estables con el molde. Un cebador no necesita reflejar la secuencia exacta del ácido nucleico molde, sino que debe ser suficientemente complementario para hibridar con el molde. El diseño de cebadores adecuados para la amplificación de una secuencia diana dada es bien conocido en la técnica y se describe en la bibliografía citada en la presente memoria.
- 35
- Como se describe en la presente memoria, los cebadores pueden incorporar características adicionales que permiten la detección o inmovilización del cebador, pero no alteran la propiedad básica del cebador, que actúa como un punto de iniciación de la síntesis de ADN. Por ejemplo, los cebadores pueden contener una secuencia de ácido nucleico adicional en el extremo 5' que no se hibrida con el ácido nucleico diana, pero que facilita la clonación, detección o secuenciación del producto amplificado. La región del cebador que es suficientemente complementaria al molde para hibridar se denomina en la presente memoria región de hibridación.
- 40
- 45 Según se utiliza en la presente memoria, un cebador es "específico" para una secuencia diana si, cuando se usa en una reacción de amplificación en condiciones suficientemente rigurosas, el cebador hibrida principalmente con el ácido nucleico diana. Típicamente, un cebador es específico para una secuencia diana si la estabilidad del dúplex cebador-diana es mayor que la estabilidad de un dúplex formado entre el cebador y cualquier otra secuencia encontrada en la muestra. Un experto en la técnica reconocerá que diversos factores, tales como las condiciones de sal, así como la composición de bases del cebador y la ubicación de los emparejamientos erróneos, afectarán a la especificidad del cebador, y que será necesaria en muchos casos la confirmación experimental rutinaria de la especificidad del cebador. Se pueden elegir condiciones de hibridación bajo las cuales el cebador puede formar dúplex estables solo con una secuencia diana. De este modo, el uso de cebadores específicos de la diana en condiciones de amplificación adecuadamente rigurosas permite la amplificación selectiva de aquellas secuencias diana que contienen los sitios de unión del cebador diana.
- 50
- 55 En realizaciones concretas, los cebadores para su uso en los métodos descritos en la presente memoria comprenden o consisten en un ácido nucleico de al menos aproximadamente 15 nucleótidos de longitud que tiene la misma secuencia que, o es complementaria a, una secuencia larga de 15 nucleótidos contiguos del segmento V o J diana. También serán útiles en ciertas realizaciones cebadores más largos, p.ej., los de aproximadamente 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, 45 o 50, nucleótidos de longitud que tienen la misma secuencia que, o una secuencia complementaria a, una secuencia contigua del segmento V o J diana. Todas las longitudes intermedias de los cebadores mencionados anteriormente se contemplan para su uso en la presente memoria. Como reconocería el experto en la técnica, los cebadores pueden tener una secuencia adicional añadida (p.ej., nucleótidos que pueden no ser iguales o complementarios al segmento V o J diana), tales como sitios de reconocimiento de enzimas de restricción, secuencias adaptadoras para la secuenciación, secuencias código de barras, y similares (véanse p.ej., secuencias de cebadores proporcionadas en la presente memoria y en la lista de secuencias). Por lo tanto, la longitud de los cebadores puede ser más larga, tal como 55, 56, 57, 58, 59, 60, 65, 70, 75, nucleótidos de longitud o más, dependiendo del uso específico o la
- 60

necesidad. Por ejemplo, en una realización, los cebadores directo e inverso se modifican ambos en el extremo 5' con la secuencia de cebador directo universal compatible con un secuenciador de ADN.

5 También se contemplan para su uso en ciertas realizaciones variantes del cebador oligonucleotídico del segmento V o del segmento J del receptor inmunitario adaptativo que pueden compartir un alto grado de identidad de secuencia con los cebadores oligonucleotídicos para los que se presentan secuencias de nucleótidos en la presente memoria, incluyendo las expuestas en la Lista de Secuencias. Por lo tanto, en estas realizaciones y realizaciones relacionadas, las variantes del cebador oligonucleotídico del segmento V o del segmento J del receptor inmunitario adaptativo pueden tener una identidad sustancial con las secuencias de cebadores oligonucleotídicos del segmento V o del segmento J del receptor inmunitario adaptativo descritas en la presente memoria, por ejemplo, tales variantes del cebador oligonucleotídico pueden comprender al menos 70% de identidad de secuencia, preferiblemente al menos 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% o mayor identidad de secuencia en comparación con una secuencia de polinucleótidos de referencia tal como las secuencias de cebadores oligonucleotídicos descritas en la presente memoria, utilizando los métodos descritos en la presente memoria (p.ej., análisis BLAST utilizando parámetros convencionales). Un experto en esta técnica reconocerá que estos valores se pueden ajustar apropiadamente para determinar la capacidad correspondiente de una variante de cebador oligonucleotídico para reasociarse con un polinucleótido que codifica un segmento del receptor inmunitario adaptativo teniendo en cuenta la degeneración de codones, el posicionamiento de los marcos de lectura y similares.

10 Típicamente, las variantes de cebador oligonucleotídico contendrán una o más sustituciones, adiciones, deleciones y/o inserciones, preferiblemente de manera que la capacidad de reasociación del oligonucleótido variante no disminuya sustancialmente con respecto a la de una secuencia del cebador oligonucleotídico del segmento V o del segmento J del receptor inmunitario adaptativo que se expone específicamente en la presente memoria. Como también se observa en otra parte de la presente memoria, en realizaciones preferidas, los cebadores oligonucleotídicos del segmento V y del segmento J del receptor inmunitario adaptativo se diseñan para que sean capaces de amplificar una secuencia de TCR reordenada que incluye la región codificante para CDR3.

15 De acuerdo con ciertas realizaciones contempladas en la presente memoria, los cebadores para su uso en los métodos de PCR múltiple de la presente descripción pueden bloquearse funcionalmente para prevenir el cebado no específico de secuencias de células distintas de T o B. Por ejemplo, los cebadores pueden bloquearse con modificaciones químicas como se describe en la publicación de la Solicitud de Patente de Estados Unidos US2010/0167353. De acuerdo con ciertas realizaciones descritas en la presente memoria, el uso de tales cebadores bloqueados en las presentes reacciones de PCR múltiple implica cebadores que pueden tener una configuración inactiva en donde la replicación del ADN (es decir, la extensión del cebador) está bloqueada, y una configuración activada en donde prosigue la replicación del ADN. La configuración inactiva del cebador está presente cuando el cebador es de hebra sencilla o cuando el cebador híbrida específicamente con la secuencia de ADN diana de interés, pero la extensión del cebador permanece bloqueada por un radical químico que está unido al extremo 3' del cebador o cerca del mismo.

20 La configuración activada del cebador está presente cuando el cebador híbrida con la secuencia de ácido nucleico diana de interés y posteriormente se actúa con ARNasa H u otro agente de escisión para eliminar el grupo 3' bloqueante, lo que permite que una enzima (p.ej., una ADN polimerasa) catalice la extensión del cebador en una reacción de amplificación. Sin desear estar limitados por la teoría, se cree que la cinética de la hibridación de tales cebadores es similar a una reacción de segundo orden y, por lo tanto, es una función de la concentración de la secuencia del gen de células T o células B en la mezcla. Los cebadores bloqueados minimizan las reacciones no específicas requiriendo la hibridación con la diana seguido de escisión antes de que pueda proseguir la extensión del cebador. Si un cebador híbrida incorrectamente con una secuencia que está relacionada con la secuencia diana deseada pero que difiere por tener uno o más nucleótidos no complementarios que dan como resultado emparejamientos erróneos de emparejamiento de bases, se inhibe la escisión del cebador, especialmente cuando hay un emparejamiento erróneo que se encuentra en o cerca del sitio de escisión. Esta estrategia para mejorar la fidelidad de la amplificación reduce la frecuencia del falso cebado en tales ubicaciones, y por lo tanto aumenta la especificidad de la reacción. Como reconocerá el experto en la técnica, las condiciones de reacción, particularmente la concentración de ARNasa H y el tiempo permitido para hibridación y extensión en cada ciclo, pueden optimizarse para maximizar la diferencia en las eficacias de escisión entre la escisión altamente eficaz del cebador cuando éste híbrida correctamente con su secuencia diana verdadera, y con una pobre escisión del cebador cuando existe un emparejamiento erróneo entre el cebador y la secuencia molde con la que éste se puede reasociar de forma incompleta.

25 Como se describe en el documento US2010/0167353, se conocen varios grupos de bloqueo en la técnica que pueden colocarse en, o cerca del extremo 3' del oligonucleótido (p.ej., un cebador) para prevenir la extensión. Se pueden modificar un cebador u otro oligonucleótido en el nucleótido 3'-terminal para prevenir o inhibir la iniciación de la síntesis de ADN mediante, por ejemplo, la adición de un residuo desoxirribonucleótido 3' (p.ej., cordicepina), un residuo 2',3'-dideoxirribonucleótido, enlaces no nucleotídicos o modificaciones alcano-diol (Patente de Estados Unidos Núm. 5.554.516). Las modificaciones alcano-diol que se pueden utilizar para inhibir o bloquear la extensión



del cebador también han sido descritas por Wilk et al., (1990 *Nucleic Acids Res.* 18(8):2065), y por Arnold et al. (Patente de Estados Unidos Núm. 6.031.091). Los ejemplos adicionales de grupos de bloqueo adecuados incluyen sustituciones hidroxilo 3' (p.ej., 3'-fosfato, 3'-trifosfato o diésteres 3'-fosfato con alcoholes tales como 3-hidroxiopropilo), 2'3'-fosfato cíclico, sustituciones hidroxilo 2' de una base de ARN terminal (p.ej., grupos fosfato o estéricamente voluminosos tales como triisopropil sililo (TIPS) o terc-butil dimetil sililo (TBDMS)). Los grupos 2'-alquilsililo tales como TIPS y TBDMS sustituidos en el extremo 3' de un oligonucleótido son descritos por Laikhter et al., Solicitud de Patente de Estados Unidos Núm. Ser. 11/686.894, que se incorpora a la presente memoria como referencia. Los sustituyentes voluminosos también pueden incorporarse en la base del residuo 3'-terminal del oligonucleótido para bloquear la extensión del cebador.

En ciertas realizaciones, el oligonucleótido puede comprender un dominio de escisión que se localiza aguas arriba (p.ej., 5' con respecto a) del grupo de bloqueo utilizado para inhibir la extensión del cebador. Como ejemplos, el dominio de escisión puede ser un dominio de escisión de ARNasa H, o el dominio de escisión puede ser un dominio de escisión de ARNasa H2 que comprende un único residuo de ARN o el oligonucleótido puede comprender la sustitución de la base de ARN por uno o más nucleósidos alternativos. Se describen dominios ilustrativos de escisión adicionales en el documento US2010/0167353.

Por lo tanto, un sistema de PCR múltiple puede utilizar 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, 45, 50, 55, 60, 65, 70, 75, 80, 85, o más cebadores directos, en donde cada cebador directo es complementario a un solo segmento funcional V de TCR o Ig o una pequeña familia de segmentos funcionales V de TCR o Ig, p.ej., un segmento V $\beta$  de TCR (véanse p.ej., los cebadores TCRBV como se expone en la Tabla 1 y en la Lista de Secuencias, SEQ ID NO: 1644-1695) y, por ejemplo, 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, 45, 50, 55, 60, 65, 70, 75, 80, 85, o más cebadores inversos, cada uno específico para un segmento J de TCR o Ig, tal como el segmento J $\beta$  de TCR (véanse p.ej., los cebadores TCRBJ en la Tabla 1 y en la Lista de Secuencias, SEQ ID NO: 1631-1643). En otra realización, una reacción de PCR múltiple puede utilizar cuatro cebadores directos cada uno específico para uno o más segmentos funcionales V de TCR $\gamma$  y cuatro cebadores inversos, cada uno específico para uno o más segmentos J de TCR $\gamma$ . En otra realización, una reacción de PCR múltiple puede utilizar 84 cebadores directos cada uno específico para uno o más segmentos V funcionales y seis cebadores inversos, cada uno específico para uno o más segmentos J.

Las condiciones de termociclación pueden seguir los métodos de los expertos en la técnica. Por ejemplo, utilizando un termociclador PCR Express™ (Hybaid, Ashford, Reino Unido), se pueden utilizar las siguientes condiciones de ciclación: 1 ciclo a 95°C durante 15 minutos, 25 a 40 ciclos a 94°C durante 30 segundos, 59°C durante 30 segundos y 72°C durante 1 minuto, seguido de un ciclo a 72°C durante 10 minutos. Como reconocerá el experto en la técnica, las condiciones de termociclación pueden optimizarse, por ejemplo, modificando las temperaturas de reasociación, los tiempos de reasociación, la cantidad de ciclos y los tiempos de extensión. Como reconocerá el experto en la técnica, la cantidad de cebador y otros reactivos de PCR utilizados, así como los parámetros de PCR (p.ej., temperatura de reasociación, tiempos de extensión y números de ciclo), pueden optimizarse para lograr la eficacia de amplificación de PCR deseada.

Alternativamente, en ciertas realizaciones relacionadas también contempladas en la presente memoria, se pueden utilizar métodos de "PCR digital" para cuantificar la cantidad de genomas diana en una muestra, sin la necesidad de una curva patrón. En la PCR digital, la reacción de PCR para una sola muestra se realiza en una multitud de más de 100 microceldas o gotitas, de modo que cada gotita se amplifica (p.ej., la generación de un producto de amplificación proporciona evidencia de la presencia de al menos una molécula molde en la microcelda o gotita) o no se amplifica (evidencia de que el molde no estaba presente en una microcelda o gotita dadas). Simplemente contando la cantidad de microceldas positivas, es posible contar directamente la cantidad de genomas diana que están presentes en una muestra de entrada.

Los métodos de PCR digital generalmente utilizan una lectura de punto final, en lugar de una señal de PCR cuantitativa convencional que se mide después de cada ciclo en la reacción de termociclación (véase, p.ej., Pekin et al., 2011 *Lab. Chip* 11 (13):2156; Zhong et al., 2011 *Lab. Chip* 11 (13):2167; Tewhey et al., 2009 *Nature Biotechnol.* 27:1025; 2010 *Nature Biotechnol.* 28:178; Vogelstein y Kinzler, 1999 *Proc. Natl. Acad. Sci. Estados Unidos* 96: 9236-41; Pohl y Shih, 2004 *Expert Rev. Mol. Diagn.* 4(1):41-7, 2004). En comparación con la PCR tradicional, la dPCR tiene las siguientes ventajas: (1) no hay necesidad de confiar en referencias o patrones, (2) la precisión deseada se puede lograr aumentando el número total de réplicas de PCR, (3) es altamente tolerante a inhibidores, (4) es capaz de analizar mezclas complejas, y (5) proporciona una respuesta lineal al número de copias presentes en una muestra para permitir que se detecte un pequeño cambio en el número de copias. Por consiguiente, cualquiera de las composiciones descritas en la presente memoria (p.ej., composiciones molde y conjuntos de cebadores oligonucleotídicos específicos de genes receptores inmunitarios adaptativos) y los métodos pueden adaptarse para su uso en dicha metodología de PCR digital, por ejemplo, ABI QuantStudio™ 12K Flex System (Life Technologies, Carlsbad, CA), el sistema de PCR QX100™ Droplet Digital™ (BioRad, Hercules, CA), el sistema de PCR digital QuantaLife™ (BioRad, Hercules, CA) o el sistema de PCR digital de microgotita RainDance™ (RainDance

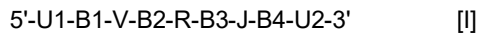
Technologies, Lexington, MA).

**Cuantificación de moléculas de ADN reordenado que codifican TCR e Ig en una muestra**

5 De acuerdo con ciertas realizaciones descritas en la presente memoria, se puede determinar la proporción de células linfoides en una muestra que son cancerosas (p.ej., células malignas). La presente descripción, sin embargo, también contempla realizaciones en las que se puede calcular el número de células cancerosas linfoides en una muestra, del número total de células de entrada de la muestra. La presente descripción también contempla realizaciones en las que se puede calcular la proporción de células cancerosas linfoides en una muestra, del número total de células de entrada de la muestra.

Brevemente y a modo de ejemplo no limitante, para calcular el número de células cancerosas linfoides del número total de células de entrada de una muestra, se puede añadir ("incorporar"), a cada reacción de amplificación de los métodos descritos actualmente un número conocido de moléculas de oligonucleótido molde sintéticas de la composición molde descrita en la presente memoria para la determinación del factor de amplificación, que tienen las secuencias que corresponden a los sitios de cebado V y J. Las moléculas de oligonucleótido molde sintéticas incluyen un conjunto de oligonucleótidos que hibridan con un cebador V y un cebador J, en la orientación correcta, y un conjunto de oligonucleótidos código de barras como se describe en la presente memoria que identifican de manera única cada una de las distintas moléculas molde sintéticas. Como se describe con mayor detalle en otra parte de la presente memoria, estas moléculas de oligonucleótido molde sintéticas también pueden incluir códigos de barras adicionales, conjuntos de identificación únicos adicionales de secuencias de oligonucleótidos y secuencias de cebadores universales.

De acuerdo con una realización, se proporciona una composición molde para la determinación del factor de amplificación en donde está presente un número conocido de una pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única, comprendiendo la composición molde: una pluralidad de oligonucleótidos molde que tienen una pluralidad de secuencias de oligonucleótidos de fórmula general



en donde:

- (a) V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única;
- (b) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única;
- (c) U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una primera secuencia de oligonucleótido adaptador universal, y (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una primera secuencia de oligonucleótido adaptador universal;
- (d) U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (i) una segunda secuencia de oligonucleótido adaptador universal, y (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal;
- (e) B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación pareada, (i) la secuencia oligonucleotídica V única de (a) y (ii) la secuencia oligonucleotídica J única de (b);
- (f) R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de (a) - (e), y en donde:
- (g) al menos uno de: (i) la pluralidad de oligonucleótidos molde comprende al menos 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, 51-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001-1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, 1701-2000 o 2001-2500 secuencias de oligonucleótidos únicas, (ii) la pluralidad de oligonucleótidos molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento V puede hibridar específicamente y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que

5 cada cebador oligonucleotídico del segmento J puede hibridar específicamente, o (iii) la pluralidad de oligonucleótidos molde comprende al menos  $a$  o al menos  $b$  secuencias de oligonucleótidos únicas, la que sea mayor, donde  $a$  es la cantidad de segmentos génicos que codifican la región V del receptor inmunitario adaptativo único en el sujeto y  $b$  es la cantidad de segmentos génicos que codifican la región J del receptor inmunitario adaptativo único en el sujeto, y la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único (Figura 6).

10 Debido a que las moléculas de oligonucleótidos molde sintéticos incluyen cada una un sitio de cebado para al menos uno de los cebadores de V y uno para los cebadores de J, los cebadores de amplificación múltiple de TCR o IG descritos en la presente memoria también amplifican estas moléculas de oligonucleótido molde sintéticas cuando los moldes sintéticos se "incorporan" a una reacción de amplificación multiplexada. El número de moléculas molde sintéticas únicas para la determinación del factor de amplificación que se pueden añadir ("incorporar") a la reacción de amplificación debe ser de al menos dos, pero puede comprender un número mucho mayor de secuencias únicas, por ejemplo, suficientes especies únicas para que la composición del molde para la determinación del factor de amplificación incluya moldes sintéticos a los que se pueda reasociar cada posible par de cebadores de V y J.

15 Ciertas realizaciones contemplan así una composición molde para la determinación del factor de amplificación que comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento V puede hibridar específicamente y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento J puede hibridar específicamente.

20 Ciertas otras realizaciones contemplan así una composición molde para la determinación del factor de amplificación que comprende al menos 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, 51-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001-1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, 1701-2000, o 2001-2500 secuencias de oligonucleótidos únicas.

25 Ciertas otras realizaciones contemplan así una composición molde para la determinación del factor de amplificación que comprende al menos  $a$  o al menos  $b$  secuencias de oligonucleótidos únicas, la que sea mayor, donde  $a$  es la cantidad de segmentos génicos que codifican la región V del receptor inmunitario adaptativo único en el sujeto y  $b$  es la cantidad de segmentos génicos que codifican la región J del receptor inmunitario adaptativo único en el sujeto, y la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único.

30 En la práctica, por ejemplo, si se añade una cantidad conocida de moléculas de oligonucleótido molde sintéticas, cada una con una secuencia de seguimiento código de barras única (B), a una reacción de amplificación multiplexada para amplificar secuencias de ADN reordenado que codifican un receptor inmunitario adaptativo, cuyos productos de amplificación se secuencian después cuantitativamente, se puede utilizar la cantidad de moléculas secuenciadas para calcular la cantidad de veces que se amplificó cada molécula de entrada. Este parámetro, descrito con mayor detalle a continuación, puede denominarse factor de amplificación.

35 En una aplicación simplificada no limitante e ilustrativa, el factor de amplificación se puede calcular como: Factor de Amplificación = Cantidad de Moléculas de Oligonucleótidos Molde Sintéticas Secuenciadas/Cantidad de Moléculas de Oligonucleótido Molde Sintéticas de Entrada. En otras realizaciones, el factor de amplificación se puede calcular para cada especie única de secuencia de molécula de oligonucleótido molde sintética que se incluye en una reacción de amplificación multiplexada, para obtener una distribución de factores de amplificación para la reacción. La mediana de esta distribución se puede utilizar como factor de amplificación final para el cálculo de la cantidad de moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra. Si el factor de amplificación se calcula de forma única para cada tipo de molécula de oligonucleótido molde sintético de la composición molde para la determinación del factor de amplificación, se puede obtener y utilizar información sobre el sesgo de la amplificación del cebador (p.ej., diferencias estadísticamente significativas detectables en las eficacias con las que diferentes cebadores de amplificación amplifican sus secuencias molde afines) para corregir dicho sesgo que se puede introducir durante la reacción de amplificación multiplexada.

40 Como otro ejemplo no limitante, utilizando (i) el valor del factor de amplificación que se determina cuantificando los productos de amplificación de las moléculas de oligonucleótido molde sintéticas de la composición molde para la determinación del factor de amplificación, y (ii) la cantidad de moléculas de ADN reordenado que codifican los receptores inmunitarios adaptativos que se detectan mediante secuenciación cuantitativa de productos de amplificación multiplexados, se puede calcular el número inicial de células inmunitarias adaptativas en la muestra. Este cálculo es: Cantidad de Células de Entrada que Tienen ADN Reordenado que Codifica Receptores Inmunitarios Adaptativos = Cantidad de Secuencias del Receptor Inmunitario Adaptativo/Factor de Amplificación. A partir de la determinación de esta manera de el número de células de entrada en la muestra que tienen el ADN reordenado que

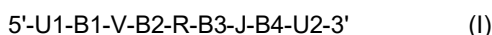
codifica receptores inmunitarios adaptativos, se puede calcular un recuento probable de el número de células cancerosas linfoides en la muestra.

5 En ciertas realizaciones, alternativamente o adicionalmente (p.ej., en paralelo), se puede calcular la proporción de células totales en la muestra que llevan una secuencia de ADN que codifica un receptor inmunitario adaptativo reordenado que está asociado con células cancerosas linfoides. En estas y otras realizaciones relacionadas, la reacción de amplificación incluye los cebadores múltiples V y J y también incluye un conjunto de cebadores de control que amplifican una región genómica con un número de copias conocido y consistente. Los genes diana adecuados para estos cebadores de control incluyen genes constitutivos, por ejemplo, HLA-DRB1, ABO, ZFX Y u otros genes expresados amplia y constitutivamente. Los ejemplos no limitantes de tales cebadores se muestran en la Tabla 2. Los cebadores de control amplifican una región establecida del genoma. Para utilizar eficazmente el espacio de la celda de flujo del secuenciador maximizando la porción usada para determinar las secuencias codificantes del receptor inmunitario adaptativo y minimizando eso para determinar las secuencias de control genómico, los cebadores pueden modificarse para reducir la amplificación por PCR eficaz. La eficacia de amplificación de los cebadores de control genómico se puede limitar, por ejemplo, alterando las secuencias de los cebadores para introducir emparejamientos erróneos, y/o estos cebadores se pueden incluir en la reacción de amplificación a baja concentración. Si bien estas modificaciones no son necesarias, pueden aumentar el rendimiento de la secuenciación. Para estimar la proporción de células totales que son células linfoides cancerosas, se puede contar a continuación el número final de secuencias del gen diana.

Tabla 2: Cebadores de control ilustrativos (Gen constitutivo)

Gen Diana	Cebador Directo	SEQ ID NO:	Cebador Inverso	SEQ ID NO:
<b>HLA-DRB 1</b>	GGATGATAAATAGGT GCCAAGTG	6383	ACTTCAATGCTGC CTGGAT	6384
<b>ABO</b>	AAGGACGAGGGCGAT TTCTA	6385	TATCTGCGATTGC GTGTCTG	6386
<b>ZFX Y</b>	TTGCATGATTTGTGG GAAGA	6387	TTCTGAAGGCCTG TGAAAGC	6388

25 A continuación se presentan detalles adicionales con respecto a la determinación de los valores del factor de amplificación utilizando la composición molde descrita en la presente memoria para la determinación del factor de amplificación. La composición molde para la determinación del factor de amplificación comprende una pluralidad de diversos oligonucleótidos molde de fórmula general (I) como se describe con mayor detalle en la presente memoria:



30 Los oligonucleótidos molde constituyentes, de los que está compuesta la composición molde, son diversos con respecto a las secuencias de nucleótidos de los oligonucleótidos molde individuales. Las secuencias de nucleótidos de los oligonucleótidos molde individuales pueden variar así considerablemente entre sí en función de la variabilidad de secuencia significativa entre el gran número de polinucleótidos de la región variable (V) y de empalme (J) de TCR o BCR posibles. Las secuencias de especies de oligonucleótidos molde individuales también pueden variar entre sí en función de las diferencias de secuencia en los oligonucleótidos U1, U2, B (B1, B2, B3 y B4) y R que se incluyen en un molde concreto dentro de la diversa pluralidad de moldes.

40 En ciertas realizaciones, los oligonucleótidos código de barras B (B1, B2, B3 y B4) pueden comprender independientemente y opcionalmente una secuencia código de barras de oligonucleótidos, en donde la secuencia código de barras se selecciona para identificar únicamente una combinación emparejada particular de una secuencia oligonucleotídica V particular única y una secuencia oligonucleotídica J particular única. El posicionamiento relativo de los oligonucleótidos código de barras B1 y B4 y de los adaptadores universales permite de forma ventajosa una rápida identificación y cuantificación de los productos de amplificación de un oligonucleótido molde único dado por lecturas de secuencia cortas y secuenciación de extremos emparejados en secuenciadores de ADN automatizados (p.ej., Illumina HiSeq™ o Illumina MiSEQ®, o GeneAnalyzer™ -2, Illumina Corp., San Diego, CA). En particular, estas realizaciones y realizaciones relacionadas permiten una determinación rápida de alto rendimiento de combinaciones específicas de una secuencia V y una J que están presentes en un producto de amplificación, para caracterizar de ese modo la eficacia de amplificación relativa de cada cebador específico de V y

cada cebador específico de J que puede estar presente en un conjunto de cebadores que es capaz de amplificar el ADN que codifica TCR o BCR reordenado en una muestra. La verificación de las identidades y/o cantidades de los productos de amplificación se puede realizar mediante lecturas de secuencia más largas, que incluyen opcionalmente lecturas de secuencia que se extienden hasta B2.

Al utilizarlo, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde está presente en una cantidad sustancialmente equimolar, que en ciertas realizaciones preferidas incluye preparaciones en las que las concentraciones molares de todos los oligonucleótidos están dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 por ciento de cada uno. En ciertas otras realizaciones preferidas proporcionadas en la presente memoria, se considera que los oligonucleótidos molde están presentes en una cantidad sustancialmente equimolar cuando las concentraciones molares de todos los oligonucleótidos están dentro de un orden de magnitud entre sí, incluyendo preparaciones en las que la mayor concentración molar que cualquier especie de oligonucleótido molde única puede tener no es mayor que 1000, 900, 800, 700, 600, 500, 440, 400, 350, 300, 250, 200, 150, 100, 90, 80, 70, 60, 50, 40 o 30 por ciento mayor que la concentración molar a la que está presente la única especie de oligonucleótido molde que tiene la concentración más baja en la composición.

De manera similar, ciertas realizaciones descritas en la presente memoria contemplan conjuntos de cebadores oligonucleotídicos para la amplificación, en los que los conjuntos de cebadores componente pueden proporcionarse en cantidades sustancialmente equimolares. Como también se describe en la presente memoria, de acuerdo con ciertas otras realizaciones, la concentración de uno o más cebadores en un conjunto de cebadores puede ajustarse deliberadamente de modo que ciertos cebadores no estén presentes en cantidades equimolares o en cantidades sustancialmente equimolares.

Preferiblemente, todos los moldes en la composición molde para la determinación del factor de amplificación, que se describe en la presente memoria y que comprende una pluralidad de oligonucleótidos molde que tienen diversas secuencias y la estructura general de fórmula general (I), son oligonucleótidos de longitud sustancialmente idéntica. Sin desear estar limitados por la teoría, generalmente se cree que en una reacción de amplificación de ácido nucleico tal como una reacción en cadena de la polimerasa (PCR), la longitud del ADN molde puede influir en la eficacia de amplificación de cebadores oligonucleotídicos al afectar la cinética de interacciones entre cebadores y moléculas de ADN molde con las que se reasocian los cebadores mediante hibridación específica dirigida a la secuencia de nucleótidos a través de la complementariedad de las bases de nucleótidos. En general, se considera que los moldes más largos funcionan de manera menos eficaz que los moldes relativamente más cortos. En ciertas realizaciones, la composición molde descrita actualmente para la determinación del factor de amplificación comprende una pluralidad de oligonucleótidos molde de fórmula general (I) como se proporciona en la presente memoria, en donde los oligonucleótidos molde tienen una longitud idéntica o una longitud sustancialmente idéntica que no es mayor que 1000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150 o 100 nucleótidos de longitud, incluyendo todos los valores enteros entre ellos.

Por consiguiente, para reducir, eliminar o minimizar la contribución potencial a sesgos indeseables en la utilización de cebadores oligonucleotídicos durante la amplificación multiplexada, las realizaciones preferidas descritas en la presente memoria pueden emplear una pluralidad de oligonucleótidos molde en donde todos los oligonucleótidos molde en la pluralidad diversa de secuencia oligonucleotídica molde tienen una longitud sustancialmente idéntica. Una pluralidad de oligonucleótidos molde pueden tener una longitud sustancialmente idéntica cuando todos (p.ej., 100%) o la mayoría (p.ej., más de 50%) de tales oligonucleótidos en una composición molde son oligonucleótidos que tienen cada uno el mismo número exacto de nucleótidos, o donde la longitud de uno o más oligonucleótidos molde en la composición molde puede variar de uno a otro en no más de 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, 60, 70, 80, 90 o 100 nucleótidos de longitud. Se apreciará a partir de la presente descripción que incluso en situaciones en las que no todos los oligonucleótidos molde tienen exactamente la misma longitud, las composiciones y métodos descritos en la presente memoria pueden emplearse todavía para determinar un factor de amplificación como se describe en la presente memoria.

De acuerdo con ciertas realizaciones descritas actualmente, (i) cada oligonucleótido molde de la composición molde actualmente descrita se proporciona en una cantidad sustancialmente equimolar, (ii) el conjunto de cebadores oligonucleotídicos que es capaz de amplificar ADN reordenado que codifica una pluralidad de receptores inmunitarios adaptativos comprende una pluralidad de cebadores oligonucleotídicos del segmento V que se proporcionan en cantidades sustancialmente equimolares, (iii) el conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica una pluralidad de receptores inmunitarios adaptativos comprende una pluralidad de cebadores oligonucleotídicos del segmento J que se proporcionan en cantidades sustancialmente equimolares, y (iv) la amplificación aumenta a escala linealmente con la cantidad de moldes iniciales de una secuencia dada.

Por lo tanto, se puede calcular un rendimiento esperado para el producto de amplificación de cada molde y asignar arbitrariamente un valor de nivel de amplificación uniforme teórico de 100%. Después de permitir que los conjuntos

de cebadores amplifiquen las secuencias de los oligonucleótidos molde en una reacción de amplificación, cualquier desviación estadísticamente significativa de la equivalencia sustancial que se observa entre las proporciones relativas de distintos productos de amplificación indica que ha habido sesgo (es decir, eficacia desigual) en la utilización del cebador durante la amplificación. En otras palabras, las diferencias cuantitativas en las cantidades

5 relativas de diferentes productos de amplificación que se obtienen indican que no todos los cebadores en el conjunto de cebadores han amplificado sus respectivas moldes con eficacias comparables. Ciertas realizaciones contemplan la asignación de un intervalo de tolerancias por encima y por debajo de un rendimiento teórico de 100%, de modo que cualquier valor de nivel de amplificación dentro del intervalo de tolerancias se puede considerar como equivalencia sustancial.

10 En ciertas de tales realizaciones, el intervalo de los rendimientos del producto de amplificación puede considerarse sustancialmente equivalente cuando los rendimientos del producto están todos dentro del mismo orden de magnitud (p.ej., difieren en menos de un factor de diez). En ciertas otras de tales realizaciones, el intervalo de rendimientos del producto de amplificación puede considerarse sustancialmente equivalente cuando los rendimientos del producto

15 difieren entre sí en no más de nueve veces, ocho veces, siete veces, seis veces, cinco veces, cuatro o tres veces. En ciertas otras realizaciones, los rendimientos del producto que pueden considerarse dentro de un intervalo de tolerancia aceptable pueden ser mayores o menores que un rendimiento calculado de 100% en hasta 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 100 o 200%.

20 Debido a que el método implica determinar la secuencia de nucleótidos de cada producto de amplificación utilizando técnicas conocidas como parte del procedimiento de cuantificación, se pueden identificar los cebadores responsables de la amplificación de cada producto único (según la definición de la secuencia) y su cantidad o cantidades relativas en el conjunto de cebadores se pueden ajustar (p.ej., aumentar o disminuir de una manera estadísticamente significativa) en consecuencia. Las concentraciones de cebadores excesivamente eficaces en el

25 conjunto de cebadores pueden reducirse con relación a las concentraciones de otros cebadores, de modo que el nivel de amplificación específica por dichos cebadores de moldes en la composición molde descrita en la presente memoria es sustancialmente equivalente al nivel de amplificación proporcionado por la mayoría de los cebadores que entregan el nivel de amplificación uniforme teórico, o que proporcionan un nivel que está dentro del intervalo de tolerancia aceptable. Las concentraciones de cebadores poco eficaces en el conjunto de cebadores pueden

30 aumentarse con respecto a las concentraciones de otros cebadores, de modo que el nivel de amplificación específica por dichos cebadores de los moldes en la composición molde descrita en la presente memoria es sustancialmente equivalente al nivel de amplificación proporcionado por la mayoría de los cebadores que proporcionan el nivel de amplificación uniforme teórico, o que proporcionan un nivel dentro del intervalo de tolerancia aceptable.

35 Proporcionando la composición molde descrita en la presente memoria como un patrón con el que pueden calibrarse conjuntos de cebadores oligonucleotídicos, y en realizaciones concretas, cuando cada oligonucleótido molde está presente en una cantidad sustancialmente equimolar de manera que las concentraciones de cebador individuales se pueden ajustar para producir amplificación sustancialmente uniforme de un conjunto estructuralmente diverso de

40 productos de amplificación, la presente descripción supera de manera ventajosa los problemas descritos anteriormente asociados con los sesgos en la eficacia del cebador individual.

Utilizando las composiciones y métodos proporcionados en la presente memoria, los cebadores individuales pueden identificarse por tener un potencial de amplificación no uniforme en virtud de su promoción de la amplificación no

45 uniforme como se evidencia por el aumento (p.ej., mayor de una manera estadísticamente significativa) o la disminución (p.ej., inferior de una manera estadísticamente significativa) de la amplificación de oligonucleótidos molde específicos con respecto al nivel de amplificación uniforme, a pesar de la presencia en una reacción de amplificación (i) de todos los oligonucleótidos molde en cantidades sustancialmente equimolares entre sí, (ii) de todos los cebadores del segmento V en cantidades sustancialmente equimolares entre sí, y (iii) de todos los

50 cebadores del segmento J en cantidades sustancialmente equimolares entre sí.

Las concentraciones relativas de dichos cebadores se pueden reducir o aumentar a continuación para obtener un conjunto completo modificado de cebadores en donde no están presentes todos los cebadores en cantidades sustancialmente equimolares entre sí, para compensar, respectivamente, el aumento o disminución del nivel de

55 amplificación relativo al nivel de amplificación uniforme. El conjunto de cebadores se puede volver a someter a ensayo a continuación para determinar su capacidad de amplificar todas las secuencias en la composición molde descrita en la presente memoria al nivel de amplificación uniforme, o dentro de un intervalo de tolerancia aceptable.

El procedimiento de someter a ensayo conjuntos de cebadores modificados por su capacidad de amplificar la composición molde descrita en la presente memoria, en donde todos los oligonucleótidos molde se proporcionan en

60 cantidades sustancialmente equimolares entre sí, puede repetirse iterativamente hasta que todos los productos se amplifiquen al nivel de amplificación uniforme, o dentro de un intervalo de tolerancia aceptable. Mediante un procedimiento de este tipo utilizando la composición molde descrita en la presente memoria, se puede normalizar la eficacia de amplificación de un conjunto de cebadores oligonucleotídicos, donde el conjunto de cebadores es capaz

de amplificar ADN reordenado productivamente que codifica uno o una pluralidad de receptores inmunitarios adaptativos en una muestra biológica que comprende ADN de células linfoides de un sujeto.

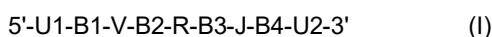
Adicional o alternativamente, de acuerdo con la presente descripción, se puede determinar si un par concreto de cebadores de amplificación oligonucleotídicos muestra un potencial de amplificación no uniforme, tal como un aumento o disminución de amplificación de la composición molde con respecto a un nivel de amplificación uniforme exhibido por la mayoría de los cebadores de amplificación oligonucleotídicos y se pueden utilizar un factor de ajuste de normalización para calcular, respectivamente, una frecuencia relativa de aparición o disminución de los productos de amplificación promovida por cada uno de dichos pares de cebadores de amplificación. Las presentes composiciones molde de esta manera, en ciertas realizaciones, proporcionan un método para corregir el potencial de amplificación de ácidos nucleicos no uniforme entre los miembros de un conjunto de cebadores de amplificación oligonucleotídicos.

Ciertas realizaciones de este tipo pueden permitir ventajosamente la corrección, calibración, homogeneización, normalización o similar de datos que se obtienen como consecuencia de eventos de amplificación no uniformes. Por lo tanto, las presentes realizaciones permiten la corrección de imprecisiones de datos, tales como las pueden resultar de la utilización de cebadores oligonucleotídicos sesgados, sin la necesidad de ajustar iterativamente las concentraciones de uno o más cebadores de amplificación y repetir las etapas de amplificar las composiciones molde descritas en la presente memoria. De este modo, se pueden obtener eficacias ventajosas cuando se puede evitar la repetición de las etapas de secuenciación cuantitativa de los productos de amplificación. Sin embargo, algunas otras realizaciones contempladas pueden emplear tal enfoque iterativo.

Por consiguiente, y como se describe en la presente memoria, en la actualidad se proporciona una composición molde para la determinación del factor de amplificación, junto con métodos para utilizar dicha composición molde. También se describen en la presente memoria métodos para corregir tales potenciales de amplificación de ácidos nucleicos no uniformes (p.ej., sesgos) entre los miembros del conjunto de cebadores oligonucleotídicos. Estas realizaciones y realizaciones relacionadas explotan beneficios previamente no reconocidos que se obtienen calibrando conjuntos de cebadores oligonucleotídicos complejos para compensar los sesgos de amplificación no deseables utilizando la composición molde para la determinación del factor de amplificación que tiene las características descritas en la presente memoria, y encontrarán usos para mejorar la precisión con la cual el TCR clonotípico específico y/o las secuencias de ADN que codifican Ig pueden cuantificarse con respecto a las metodologías descritas previamente.

Como también se indicó anteriormente y se describe en otra parte de la presente memoria, antes de la presente descripción existían, en ciertas situaciones, discrepancias insatisfactorias y difíciles de discernir entre (i) la distribución cuantitativa real de moldes de ADN que codifican receptores inmunitarios adaptativos reordenados que tienen secuencias únicas en una muestra biológica que comprende ADN de células linfoides de un sujeto, y (ii) la representación relativa observada de productos de amplificación de ácidos nucleicos de tales moldes que se detectaron, después de la amplificación multiplexada utilizando un conjunto complejo de cebadores de amplificación oligonucleotídicos diseñados para amplificar sustancialmente todos los genes del receptor inmunitario adaptativo reordenados productivamente en la muestra.

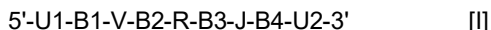
MOLDES. En la presente memoria se describe una composición molde para la determinación del factor de amplificación, para uso en una reacción de amplificación de ácidos nucleicos multiplexada con un conjunto de cebadores oligonucleotídicos que es capaz de amplificar ADN reordenado (que en ciertas realizaciones puede referirse a ADN reordenado productivamente pero que en ciertas otras realizaciones no se necesita que esté tan limitado) que codifica uno o una pluralidad de receptores inmunitarios adaptativos en una muestra biológica que comprende ADN de células linfoides de un sujeto, comprendiendo la composición molde una pluralidad de oligonucleótidos molde de fórmula general (I):



como se proporciona en la presente memoria. En ciertas realizaciones preferidas, cada oligonucleótido molde en la pluralidad de oligonucleótidos molde está presente en una cantidad sustancialmente equimolar, que en ciertas realizaciones y como se indicó anteriormente puede referirse a una composición en donde cada uno de los oligonucleótidos molde está presente a una concentración equimolar o a una concentración molar que se desvía de la equimolar en no más de 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, 50, 60, 70, 80, 90, 100 o 200% en una base molar, y que en ciertas otras realizaciones puede referirse a una composición en la que están presentes todos los oligonucleótidos molde a concentraciones molares que están dentro de un orden de magnitud entre sí. La pluralidad de moldes puede comprender al menos 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100 o más especies discretas de oligonucleótidos que tienen cada una una secuencia de nucleótidos distinta, que incluye cada valor entero intermedio entre ellas.

La composición molde descrita en la presente memoria comprende así una pluralidad de oligonucleótidos molde de

fórmula general:



5 en donde, brevemente y según se elabora con mayor detalle en otra parte de la presente memoria, de acuerdo con ciertas realizaciones preferidas:

10 V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias oligonucleotídicas molde V comprende una secuencia oligonucleotídica única;

15 J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias oligonucleotídicas molde J comprende una secuencia oligonucleotídica única;

20 U1 y U2 son cada uno o nada o comprenden un oligonucleótido que tiene, independientemente, una secuencia que se selecciona entre (i) una secuencia de oligonucleótido adaptador universal, y (ii) una secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a la secuencia de oligonucleótido adaptador universal;

25 B1, B2, B3, y B4 son cada uno independientemente nada o comprenden cada uno un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 nucleótidos contiguos (incluyendo todos valores enteros entre ellos), en donde en cada una de la pluralidad de secuencias oligonucleotídicas molde B comprende una secuencia oligonucleotídica única que identifica de manera única, o identifica como una combinación emparejada, (i) la secuencia oligonucleotídica V única del oligonucleótido molde y (ii) la secuencia oligonucleotídica única J del oligonucleótido molde; y

R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de V, J, U1, U2, B1, B2, B3 y B4.

30 En ciertas realizaciones preferidas, la composición molde comprende al menos un oligonucleótido molde con el que se puede reasociar cada cebador de amplificación de oligonucleótido en un conjunto de cebadores de amplificación. Es decir, en ciertas realizaciones preferidas, la composición molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar específicamente cada cebador oligonucleotídico del segmento V, y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de general fórmula (I) con la que puede hibridar específicamente cada cebador oligonucleotídico del segmento J.

40 De acuerdo con tales realizaciones, el conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica uno o una pluralidad de receptores inmunitarios adaptativos comprende una pluralidad *a* de cebadores oligonucleotídicos del segmento V únicos y una pluralidad *b* de cebadores oligonucleotídicos del segmento J únicos. Cada uno de la pluralidad de *a* cebadores oligonucleotídicos del segmento V es capaz independientemente de reasociarse o hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor inmunitario adaptativo o con el complemento del mismo, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico que codifica la región V del receptor inmunitario adaptativo. Cada uno de la pluralidad de *b* cebadores oligonucleotídicos del segmento J es capaz independientemente de reasociarse o hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor inmunitario adaptativo o con el complemento del mismo, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico que codifica la región J del receptor inmunitario adaptativo.

55 Por lo tanto, en ciertas realizaciones y como también se comenta en otra parte de la presente memoria, la presente composición molde puede utilizarse en reacciones de amplificación con cebadores de amplificación que están diseñados para amplificar todas las secuencias génicas que codifican el receptor inmunitario adaptativo reordenadas, incluyendo aquellas que no están expresadas, mientras que en otras realizaciones, la composición molde y los cebadores de amplificación se pueden diseñar de modo que no produzcan productos de amplificación de genes reordenados que no se expresan (p.ej., pseudogenes, orfones). Por lo tanto, se apreciará que en ciertas realizaciones solo un subconjunto de genes codificantes del receptor inmunitario adaptativo reordenados puede amplificarse deseablemente, de manera que se puedan diseñar y emplear subconjuntos de cebadores de amplificación adecuados para amplificar solo aquellas secuencias V-J reordenadas que sean de interés. En estas realizaciones y realizaciones relacionadas, correspondientemente, se puede utilizar una composición molde descrita en la presente memoria que comprende solo un subconjunto de interés de secuencias reordenadas V-J reordenadas, siempre que la composición molde contenga al menos un oligonucleótido molde con el que se pueda reasociar cada cebador de amplificación oligonucleotídico en un conjunto de cebadores de amplificación. El número



real de oligonucleótidos molde en la composición molde puede así variar considerablemente entre las realizaciones contempladas, como una función del conjunto de cebadores de amplificación que se va a utilizar.

Por ejemplo, en ciertas realizaciones relacionadas, en la composición molde, la pluralidad de oligonucleótidos molde puede tener una pluralidad de secuencias de fórmula general (I) que se selecciona entre (1) la pluralidad de secuencias de oligonucleótidos de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias V y J de TCRB expuestas en al menos un conjunto de 68 SEQ ID NO de V y J de TCRB, respectivamente, como se expone en la Figura 9 como Conjunto 1 de V/J de TCRB, Conjunto 2 de V/J de TCRB, Conjunto 3 de V/J de TCRB, Conjunto 4 de V/J de TCRB, Conjunto 5 de V/J de TCRB, Conjunto 6 de V/J de TCRB, Conjunto 7 de V/J de TCRB, Conjunto 8 de V/J de TCRB, Conjunto 9 de V/J de TCRB, Conjunto 10 de V/J de TCRB, Conjunto 11 de V/J de TCRB, Conjunto 12 de V/J de TCRB y Conjunto 13 de V/J de TCRB; (2) la pluralidad de secuencias oligonucleotídicas de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias V y J de TCRG expuestas en al menos un conjunto de 14 SEQ ID NO de V y J de TCRG, respectivamente, como se expone en la Figura 10 como conjunto 1 de V/J de TCRG, conjunto 2 de V/J de TCRG, conjunto 3 de V/J de TCRG, conjunto 4 de V/J de TCRG y conjunto 5 de V/J de TCRG; y (3) la pluralidad de secuencias oligonucleotídicas de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias V y J de IGH expuestas en al menos un conjunto de 127 SEQ ID NO de V y J de IGH, respectivamente, como se expone en la Figura 11 como conjunto 1 de V/J de IGH, conjunto 2 de V/J de IGH, conjunto 3 de V/J de IGH, conjunto 4 de V/J de IGH, conjunto 5 de V/J de IGH, conjunto 6 de V/J de IGH, conjunto 7 de V/J de IGH, conjunto 8 de V/J de IGH y conjunto 9 de V/J de IGH.

En ciertas realizaciones, V es una secuencia de polinucleótidos que codifica al menos 10-70 aminoácidos contiguos de una región V del receptor inmunitario adaptativo, o el complemento de la misma; J es una secuencia de polinucleótidos que codifica al menos 5-30 aminoácidos contiguos de una región J del receptor inmunitario adaptativo, o el complemento de la misma; U1 y U2 no son nada o comprenden un oligonucleótido que comprende una secuencia de nucleótidos que se selecciona entre (i) una secuencia de oligonucleótido adaptador universal, y (ii) una secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a la secuencia de oligonucleótido adaptador universal; B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación emparejada, (i) la secuencia oligonucleotídica V única y (ii) la secuencia oligonucleotídica J única; y R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de V, J, U1, U2, B1, B2, B3 y B4.

Un gran número de secuencias génica de la región (V) variable y de la región (J) de empalme del receptor inmunitario adaptativo se conocen como secuencias de nucleótidos y/o aminoácidos, incluyendo secuencias de ADN genómico no reordenadas de los loci de TCR e Ig, y secuencias de ADN reordenado productivamente en tales loci y sus productos codificados, y también incluyendo pseudogenes en estos loci, y también incluyendo los orfones relacionados. Véanse, p.ej., U.S.A.N. 13/217.126; U.S.A.N. 12/794.507; PCT/US2011/026373; PCT/US2011/049012. Estas y otras secuencias conocidas en la técnica se pueden utilizar de acuerdo con la presente descripción para el diseño y producción de oligonucleótidos molde para que sean incluidas en la composición molde proporcionada actualmente para normalizar la eficacia de amplificación de un conjunto de cebadores oligonucleotídicos, y para el diseño y producción del conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica cadenas de polipéptidos de TCR o de Ig, cuyo ADN reordenado puede estar presente en una muestra biológica que comprende ADN de células linfoides.

En la fórmula (I), V es una secuencia de polinucleótidos de al menos 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 350, 360, 380, 390, 400 o 450 y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica de la región variable (V) del receptor inmunitario adaptativo (p.ej., TCR o BCR), o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única. Las secuencias genómicas para los genes de la región V de TCR y BCR de seres humanos y otras especies son conocidas y están disponibles en bases de datos públicas tales como Genbank; las secuencias génicas de la región V incluyen secuencias de polinucleótidos que codifican los productos de genes de TCR y BCR reordenados expresados y también incluyen secuencias de polinucleótidos de pseudogenes que se han identificado en los loci de la región V. La longitud, la composición de nucleótidos (p.ej., contenido de GC), y la secuencia de polinucleótidos lineal real de las diversas secuencias de polinucleótidos V que se pueden incorporar a los moldes actualmente descritos de fórmula general (I) pueden variar ampliamente, y se sabe, por ejemplo, que incluyen "puntos calientes" o regiones hipervariables que exhiben diversidad de secuencia particular.

El polinucleótido V en la fórmula general (I) (o su complemento) incluye secuencias con las que se pueden reasociar específicamente los miembros de conjuntos de cebadores oligonucleotídicos específicos para genes de TCR o BCR. Los conjuntos de cebadores que son capaces de amplificar el ADN reordenado que codifica una pluralidad de TCR o BCR se describen, por ejemplo, en los documentos U.S.A.N. 13/217.126; U.S.A.N. 12/794.507;

PCT/US2011/026373; o PCT/US2011/049012; o similares; o como se describe allí, se pueden diseñar para que incluyan secuencias de oligonucleótidos que pueden hibridar específicamente con cada gen V único y con cada gen J en un locus del gen de TCR o BCR concreto (p.ej., TCR  $\alpha$ ,  $\beta$ ,  $\gamma$  o  $\delta$ , o IgH  $\mu$ ,  $\gamma$ ,  $\delta$ ,  $\alpha$  o  $\epsilon$ , o IgL  $\kappa$  o  $\lambda$ ). Por ejemplo, a modo de ilustración y no de limitación, un cebador oligonucleotídico de un conjunto de amplificación de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica uno o una pluralidad de TCR o BCR puede incluir típicamente una secuencia de nucleótidos de 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 o 40 nucleótidos contiguos, o más, y puede reasociarse específicamente con una secuencia complementaria de 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 o 40 nucleótidos contiguos de un polinucleótido V o J como se proporciona en la presente memoria. En ciertas realizaciones, los cebadores pueden comprender al menos 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 nucleótidos, y en cierta realización los cebadores pueden comprender secuencias de no más de 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 o 40 nucleótidos contiguos. Los cebadores y sitios de reasociación de cebadores de otras longitudes también se contemplan expresamente, como se describe en la presente memoria.

La secuencia de polinucleótidos completa de cada polinucleótido V en la fórmula general (I) puede consistir, pero no es necesario que consista exclusivamente en nucleótidos contiguos de cada gen V distinto. Por ejemplo y de acuerdo con ciertas realizaciones, en la composición molde descrita en la presente memoria, solo se necesita que cada polinucleótido V de fórmula (I) tenga al menos una región que comprenda una secuencia oligonucleotídica V única que se encuentre en un gen V y con la que se pueda reasociar específicamente un único cebador de la región V del conjunto de cebadores. Por lo tanto, el polinucleótido V de fórmula (I) puede comprender la totalidad o cualquier porción prescrita (p.ej., al menos 15, 20, 30, 60, 90, 120, 150, 180 o 210 nucleótidos contiguos, o cualquier valor entero entre ellos) de una secuencia génica V de origen natural (incluyendo una secuencia de pseudogén V) siempre que al menos se incluya al menos una única región de la secuencia oligonucleotídica V (el sitio de reasociación del cebador) que no esté incluida en ningún otro polinucleótido molde V.

Se puede preferir en ciertas realizaciones que la pluralidad de polinucleótidos V que está presente en la composición molde descrita en la presente memoria tenga longitudes que simulen las longitudes globales de secuencias de nucleótidos del gen V de origen natural, conocidas, incluso cuando las secuencias de nucleótidos específicas difieren entre la región V molde y cualquier gen V de origen natural. Las longitudes de la región V en los moldes descritos en la presente memoria pueden diferir de las longitudes de las secuencias del gen V que se producen de forma natural en no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 por ciento.

El polinucleótido V en la fórmula (I) puede, en ciertas realizaciones, comprender una secuencia de nucleótidos que tiene una longitud que es igual o similar a la longitud de un gen V típico desde su codón de inicio a su región codificante de CDR3 y puede incluir, pero no es necesario que incluya, una secuencia de nucleótidos que codifica la región CDR3. Las secuencias de nucleótidos que codifican CDR3 y las longitudes de secuencia pueden variar considerablemente y se han caracterizado por varios esquemas de numeración diferentes (p.ej., Lefranc, 1999 *Immunologist* 7:132; Kabat et al., 1991 en: *Sequences of Proteins of Immunological Interest*, Publicación NIH 91-3242; Chothia et al., 1987 *J. Mol. Biol.* 196:901; Chothia et al., 1989 *Nature* 342:877; Al-Lazikani et al., 1997 *J. Mol. Biol.* 273:927; véanse también, p.ej., Rock et al., 1994 *J. Exp. Med.* 179:323; Saada et al., 2007 *Immunol. Cell Biol.* 85:323).

Brevemente, la región CDR3 típicamente abarca la porción polipeptídica que se extiende desde un residuo de cisteína altamente conservado (codificado por el codón trinucleotídico TGY; Y = T o C) en el segmento V a un residuo de fenilalanina altamente conservado (codificado por TTY) en el segmento J de TCR, o a un triptófano altamente conservado (codificado por TGG) en IGH. Más de 90% de los reordenamientos productivos naturales en el locus TCRB tiene una longitud de codificación de CDR3 por este criterio de entre 24 y 54 nucleótidos, que corresponden a entre 9 y 17 aminoácidos codificados. Las longitudes de CDR3 de los oligonucleótidos molde sintéticos descritos en la actualidad, para cualquier locus de TCR o BCR dado, caen dentro del mismo intervalo que 95% de los reordenamientos naturales. Por lo tanto, por ejemplo, en una composición molde descrita en la presente memoria para la determinación del factor de amplificación, la porción codificante de CDR3 del polinucleótido V puede tener una longitud de 24 a 54 nucleótidos, incluyendo cada número entero entre ellos. Los esquemas de numeración para las regiones que codifican CDR3 descritas anteriormente denotan las posiciones de los codones conservados de cisteína, fenilalanina y triptófano, y estos esquemas de numeración también pueden aplicarse a pseudogenes en los que uno o más codones que codifican estos aminoácidos conservados pueden haber sido reemplazados por un codón que codifica un aminoácido diferente. Para los pseudogenes que no utilizan estos aminoácidos conservados, la longitud de CDR3 puede definirse con respecto a la posición correspondiente en la que se habría observado el residuo conservado sin la sustitución, de acuerdo con uno de los esquemas de numeración de posiciones de secuencia de CDR3 establecidos mencionados anteriormente.

También se puede preferir, en ciertas realizaciones, que la pluralidad de polinucleótidos V que están presentes en la composición molde descrita en la presente memoria tenga composiciones de nucleótidos (p.ej., porcentaje de contenido de GC) que simulen las composiciones nucleotídicas globales de secuencias de genes V de origen natural

conocidas, incluso cuando difieran las secuencias de nucleótidos específicas. Tales composiciones de nucleótidos de la región V molde pueden diferir de las composiciones de nucleótidos de secuencias del gen V de origen natural en no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 por ciento. Opcionalmente y de acuerdo con ciertas realizaciones, el polinucleótido V del oligonucleótido molde descrito en la presente memoria incluye un codón de terminación en o cerca del extremo 3' de V en la fórmula general (I).

En la fórmula (I) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única.

El polinucleótido J en la fórmula general (I) (o su complemento) incluye secuencias con las que se pueden reasociar específicamente miembros de los conjuntos de cebadores oligonucleotídicos específicos para genes de TCR o BCR. Los conjuntos de cebadores que son capaces de amplificar el ADN reordenado que codifica una pluralidad de TCR o BCR se describen, por ejemplo, en los documentos U.S.A.N. 13/217.126; U.S.A.N. 12/794,507; PCT/US2011/026373; o PCT/US2011/049012; o similares; o como se describe allí, se pueden diseñar para que incluyan secuencias de oligonucleótidos que pueden hibridar específicamente con cada gen V único y con cada gen J único en un locus del gen de TCR o BCR concreto (p.ej., TCR  $\alpha$ ,  $\beta$ ,  $\gamma$  o  $\delta$ , o IgH  $\mu$ ,  $\gamma$ ,  $\delta$ ,  $\alpha$  o  $\epsilon$ , o IgL  $\kappa$  o  $\lambda$ ).

La secuencia de polinucleótidos completa de cada polinucleótido J en la fórmula general (I) puede consistir, pero es necesario que consista exclusivamente en nucleótidos contiguos de cada gen J distinto. Por ejemplo y de acuerdo con ciertas realizaciones, solo es necesario que en la composición molde descrita en la presente memoria, cada polinucleótido J de fórmula (I) tenga al menos una región que comprenda una secuencia oligonucleotídica J única que se encuentra en un gen J y con la que se puede reasociar específicamente un único cebador de la región V en el conjunto de cebadores. Por lo tanto, el polinucleótido V de fórmula (I) puede comprender la totalidad o cualquier porción prescrita (p.ej., al menos 15, 20, 30, 60, 90, 120, 150, 180 o 210 nucleótidos contiguos, o cualquier valor entero entre ellos) de una secuencia génica V de origen natural (incluyendo una secuencia de pseudogén V) siempre que al menos se incluya una región de secuencia oligonucleotídica V única (el sitio de reasociación del cebador) que no esté incluida en ningún otro polinucleótido J molde.

Se puede preferir en ciertas realizaciones que la pluralidad de polinucleótidos J que están presentes en la composición molde descrita en la presente memoria tengan longitudes que simulan las longitudes globales de secuencias de nucleótidos del gen J conocidas, de origen natural, incluso cuando las secuencias de nucleótidos específicas difieran entre la región J molde y cualquier gen J de origen natural. Las longitudes de la región J en los moldes descritos en la presente memoria pueden diferir de las longitudes de las secuencias del gen J de origen natural en no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 por ciento.

El polinucleótido J en la fórmula (I) puede, por lo tanto, en ciertas realizaciones, comprender una secuencia de nucleótidos que tiene una longitud que es igual o similar a la longitud de un gen J de origen natural típico y puede incluir, pero no es necesario que incluya, una secuencia de nucleótidos que codifica la región CDR3, como se comentó anteriormente.

Las secuencias genómicas para los genes de la región J de TCR y BCR de seres humanos y otras especies son conocidas y están disponibles en bases de datos públicas tales como Genbank; las secuencias génicas de la región J incluyen secuencias de polinucleótidos que codifican los productos de genes de TCR y BCR reordenados expresados y no expresados. La longitud, la composición de nucleótidos (p.ej., contenido de GC), y la secuencia de polinucleótidos lineal real de las diversas secuencias de polinucleótidos J que se pueden incorporar a los moldes de fórmula general (I) descritos actualmente pueden variar ampliamente.

Las alternativas a las secuencias V y J descritas en la presente memoria, para su uso en la construcción de los oligonucleótidos molde y/o cebadores oligonucleotídicos del segmento V y el segmento J descritos en la presente memoria, pueden ser seleccionadas por un experto basándose en la presente descripción utilizando conocimientos en la técnica referentes a secuencias génicas publicadas para las regiones que codifican V y J de los genes para cada subunidad TCR e Ig. Las entradas de Genbank de referencia para secuencias de receptor inmunitario adaptativo humano incluyen: TCR $\alpha$ : (TCRA/D): NC\_000014.8 (chr14: 22090057..23021075); TCR $\beta$ : (TCRB): NC\_000007.13 (chr7: 141998851..142510972); TCR $\gamma$ : (TCRG): NC\_000007.13 (chr7: 38279625..38407656); cadena pesada de inmunoglobulina, IgH (IGH): NC\_000014.8 (chr14: 106032614..107288051); cadena ligera kappa de inmunoglobulina, IgLk (IGK): NC\_000002.11 (chr2: 89156874..90274235); y cadena ligera-lambda de inmunoglobulina, IgL $\lambda$  (IGL): NC\_000022.10 (chr22: 22380474..23265085). Las entradas de Genbank de referencia para secuencias de loci del receptor inmunitario adaptativo de ratón incluyen: TCR $\beta$ : (TCRB): NC\_000072.5 (chr6: 40841295..41508370) y cadena pesada de inmunoglobulina, IgH (IGH): NC\_000078.5 (chr12: 114496979..117248165).

Los análisis de diseño de moldes y cebadores y las consideraciones de selección del sitio diana pueden realizarse,

por ejemplo, utilizando el soporte lógico de análisis de cebadores OLIGO y/o el soporte lógico del algoritmo BLASTN 2.0.5 (Altschul et al., Nucleic Acids Res. 1997, 25(17):3389-402), u otros programas similares disponibles en la técnica.

5 Por consiguiente, basándose en la presente descripción y a la vista de estas secuencias de genes del receptor inmunitario adaptativo y de las metodologías de diseño de oligonucleótidos conocidas, para la inclusión en los presentes oligonucleótidos molde los expertos en la técnica pueden diseñar una pluralidad de secuencias de polinucleótidos específicas de la región V y específicas de la región J que contienen independientemente secuencias de oligonucleótidos que son únicas para un gen V y J dado, respectivamente. De forma similar, a partir de la  
10 presente descripción y a la vista de las secuencias del receptor inmunitario adaptativo conocidas, los expertos en la técnica también pueden diseñar un conjunto de cebadores que comprende una pluralidad de cebadores oligonucleotídicos específicos de la región V y específicos de la región J que son capaces cada uno independientemente de reasociarse con una secuencia específica que es única para un gen V y J dado, respectivamente, por lo que la pluralidad de cebadores es capaz de amplificar sustancialmente todos los genes V y sustancialmente todos los genes J en un locus que codifica el receptor inmunitario adaptativo dado (p.ej., un locus de TCR o IgH humano). Tales conjuntos de cebadores permiten la generación, en una PCR multiplexada (p.ej., utilizando múltiples pares de cebadores directos e inversos), de productos de amplificación que tienen un primer extremo que está codificado por un segmento génico que codifica la región V reordenada y un segundo extremo que está codificado por un segmento génico que codifica una región J.

20 Tales productos de amplificación pueden incluir una secuencia que codifica CDR3. Los cebadores pueden diseñarse preferiblemente para producir productos de amplificación que tienen porciones suficientes de secuencias V y J y/o secuencias código de barras V-J (B) como se describe en la presente memoria, de forma que secuenciando los productos (amplicones), es posible identificar basándose en secuencias que son únicas para cada segmento génico (i) el gen V concreto, y (ii) el gen J concreto en cuya proximidad el gen V experimenta reordenamiento para producir un gen que codifica el receptor inmunitario adaptativo funcional. Típicamente, y en realizaciones preferidas, los productos de amplificación por PCR no tendrán más de 600 pares de bases de tamaño, lo que de acuerdo con la teoría no limitante excluirá los productos de amplificación de los genes del receptor inmunitario adaptativo no reordenados. En ciertas otras realizaciones preferidas, los productos de amplificación tendrán un tamaño de no más de 500, 400, 300, 250, 200, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30 o 20 pares de bases, tal como puede proporcionar ventajosamente una cuantificación rápida y de alto rendimiento de amplicones separados de secuencia mediante lecturas cortas de secuencia.

35 En ciertas realizaciones preferidas, la pluralidad de oligonucleótidos molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar, preferiblemente hibridar específicamente, cada cebador oligonucleotídico del segmento V y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar, preferiblemente hibridar específicamente, cada cebador oligonucleotídico del segmento J.

40 En ciertas realizaciones, la pluralidad de oligonucleótidos molde comprende al menos 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, 51-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-800, 801-900, 901-1000, 1001 -1100, 1101-1200, 1201-1300, 1301-1400, 1401-1500, 1501-1600, 1601-1700, 1701-2000, o 2001-2500 secuencias de oligonucleótidos únicas.

45 En ciertas realizaciones, la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único. Se apreciará que debido a que los oligonucleótidos molde tienen una pluralidad de secuencias de oligonucleótidos de fórmula general (I), que incluye un polinucleótido V y que también incluye un polinucleótido J, la composición molde puede comprender menos que (axb) secuencias de oligonucleótidos únicas, pero en ciertas realizaciones comprenderá al menos la mayor de las secuencias de oligonucleotídicas únicas a o b.

50 Por consiguiente, la composición puede incorporar al menos una aparición de cada secuencia de polinucleótido V única y al menos una aparición de cada secuencia de polinucleótido J única, donde en algunos casos la al menos una aparición de un polinucleótido V único concreto estará presente en el mismo oligonucleótido molde en el que se puede encontrar la al menos una aparición de un polinucleótido J único concreto. Así, por ejemplo, "al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único" puede referirse en ciertos casos a un único oligonucleótido molde en el que están presentes un único polinucleótido V y un único polinucleótido J.

60 Como también se describe en otra parte de la presente memoria, en ciertas otras realizaciones preferidas, la composición molde comprende al menos un oligonucleótido molde con el que se puede reasociar cada cebador de amplificación de oligonucleótidos en un conjunto de cebadores de amplificación. Por lo tanto, la composición puede comprender menos de a o b secuencias únicas, por ejemplo, donde un conjunto de cebadores de amplificación

puede no incluir un cebador único para cada secuencia V y/o J posible.

Se observará que ciertas realizaciones contemplan una composición molde para la determinación del factor de amplificación, en donde la composición molde comprende una pluralidad de oligonucleótidos molde que tienen una pluralidad de secuencias oligonucleotídicas de fórmula general 5'-U1-B1-V-B2-R-B3-J-B4-U2-3' (I) como se describe en la presente memoria. De acuerdo con estas realizaciones y realizaciones relacionadas y también como se describe en otra parte de la presente memoria, el conjunto de cebadores de amplificación oligonucleotídicos que es capaz de amplificar el ADN reordenado productivamente puede excluir cualquier cebador oligonucleotídico que hibride específicamente con un pseudogén u orfón de la región V o con un pseudogen u orfón de la región J. Por lo tanto, en tales realizaciones, la composición molde excluirá deseablemente oligonucleótidos molde de fórmula general (I) en los que las secuencias de oligonucleótido V único y/o las secuencias de oligonucleótido J único son secuencias que son, respectivamente, únicas para un pseudogén u orfón de la región V o un pseudogen u orfón de la región J.

Se describe una composición molde de TCRB ilustrativa que comprende 858 oligonucleótidos molde distintos en la Lista de Secuencias en SEQ ID NO: 3157-4014. Otra composición molde de TCRB ilustrativa que comprende 871 oligonucleótidos molde distintos se describe en la Lista de secuencias en SEQ ID NO: 1-871. Otra composición molde de TCRB ilustrativa que comprende 689 oligonucleótidos molde distintos se describe en la Lista de Secuencias en SEQ ID NO: 872-1560.

Se describe una composición molde de TCRG ilustrativa que comprende 70 oligonucleótidos molde distintos en la Lista de Secuencias en SEQ ID NO: 4015-4084. También se describe una composición molde de TCRG ilustrativa que comprende 70 oligonucleótidos molde distintos en la Lista de Secuencias en SEQ ID NO: 1561-1630.

Se describe una composición molde de IGH ilustrativa que comprende 1116 oligonucleótidos molde distintos en la Lista de Secuencias en SEQ ID NO: 4085-5200. También se describe una composición molde de IGH ilustrativa que comprende 1116 oligonucleótidos molde distintos en la Lista de Secuencias en SEQ ID NO: 1805-2920.

También se describen en la presente memoria conjuntos ilustrativos de polinucleótidos V y J para su inclusión en los oligonucleótidos molde descritos en la presente memoria que tienen una pluralidad de secuencias de oligonucleótidos de fórmula general (I). Para TCRB, la pluralidad de oligonucleótidos molde puede tener una pluralidad de secuencias de oligonucleótidos de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias de V y J de TCRB expuestas en al menos un conjunto de 68 SEQ ID NO de V y J de TCRB, respectivamente, como se expone en la Figura 5 como Conjunto 1 de V/J de TCRB, Conjunto 2 de V/J de TCRB, Conjunto 3 de V/J de TCRB, Conjunto 4 de V/J de TCRB, Conjunto 5 de V/J de TCRB, Conjunto 6 de V/J de TCRB, Conjunto 7 de V/J de TCRB, Conjunto 8 de V/J de TCRB, Conjunto 9 de V/J de TCRB, Conjunto 10 de V/J de TCRB, Conjunto 11 de V/J de TCRB, Conjunto 12 de V/J de TCRB y Conjunto 13 de V/J de TCRB .

Para TCRG, la pluralidad de oligonucleótidos molde puede tener una pluralidad de secuencias de oligonucleótidos de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias de V y J de TCRG expuestas en al menos un conjunto de 14 SEQ ID NO de V y J de TCRG, respectivamente, como se expone en la Figura 6 como conjunto 1 de V/J de TCRG, conjunto 2 de V/J de TCRG, conjunto 3 de V/J de TCRG, conjunto 4 de V/J de TCRG y conjunto 5 de V/J de TCRG.

Para IGH, la pluralidad de oligonucleótidos molde puede tener una pluralidad de secuencias de oligonucleótidos de fórmula general (I) en la que los polinucleótidos V y J tienen las secuencias V y J de IGH expuestas en al menos un conjunto de 127 SEQ ID NO de V y J de IGH, respectivamente, como se expone en la Figura 7 como conjunto 1 de V/J de IGH, conjunto 2 de V/J de IGH, conjunto 3 de V/J de IGH, conjunto 4 de V/J de IGH, conjunto 5 de V/J de IGH, conjunto 6 de V/J de IGH, conjunto 7 de V/J de IGH, conjunto 8 de V/J de IGH y conjunto 9 de V/J de IGH.

ADAPTADORES. Los oligonucleótidos molde descritos en la presente memoria de fórmula general (I) también pueden comprender una primera (U1) y una segunda (U2) secuencias de oligonucleótido adaptador universal, o pueden carecer de una o ambas de U1 y U2. U1 por lo tanto puede comprender nada o un oligonucleótido que tiene una secuencia que se selecciona entre (i) una primera secuencia de oligonucleótido adaptador universal, y (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una primera secuencia de oligonucleótido adaptador universal, y U2 puede comprender nada o un oligonucleótido que tiene una secuencia que se selecciona entre (i) una segunda secuencia de oligonucleótido adaptador universal, y (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una segunda secuencia de oligonucleótido adaptador universal.

U1 y/o U2 pueden comprender, por ejemplo, secuencias de oligonucleótido adaptador universal y/o secuencias de oligonucleótidos específicas de la plataforma de secuenciación que son específicas de una tecnología de secuenciación de molécula única empleada, por ejemplo sistemas HiSeq™ o GeneAnalyzer™-2 (GA-2) (Illumina, Inc., San Diego, CA) u otro paquete de secuenciación adecuado de instrumentación, reactivos y soporte lógico. La

inclusión de tales secuencias adaptadoras específicas de la plataforma permite la secuenciación cuantitativa directa de la composición molde descrita actualmente, que comprende una pluralidad de diferentes oligonucleótidos molde de fórmula general (I), utilizando una metodología de secuenciación de nucleótidos tal como HiSeq™ o GA2 o equivalente. Por lo tanto, esta característica permite ventajosamente la caracterización cualitativa y cuantitativa de la composición molde.

En particular, la capacidad de secuenciar todos los componentes de la composición molde permite directamente la verificación de que cada oligonucleótido molde en la pluralidad de oligonucleótidos molde está presente en una cantidad sustancialmente equimolar. Por ejemplo, se puede generar un conjunto de los oligonucleótidos molde descritos actualmente que tienen secuencias adaptadoras universales en ambos extremos, de modo que las secuencias adaptadoras se pueden utilizar para incorporar adicionalmente oligonucleótidos específicos de la plataforma de secuenciación a cada extremo de cada molde.

Sin desear estar limitados por la teoría, los oligonucleótidos específicos de la plataforma se pueden añadir a los extremos de tales moldes modificados utilizando secuencias oligonucleótidos 5' (5' - secuencia de la plataforma - adaptador universal - secuencia 1 - 3') y 3' (5' - secuencia de la plataforma - adaptador universal - secuencia 2 - 3') en tan solo dos ciclos de desnaturalización, reasociación y extensión, de modo que la representación relativa en la composición molde de cada uno de los oligonucleótidos molde del componente no se altera cuantitativamente. Las secuencias identificadoras únicas (p.ej., las secuencias código de barras B que comprenden secuencias de oligonucleótidos V y B únicas que están asociadas con, y así identifican, respectivamente, regiones V y J individuales, como se describe en la presente memoria) se colocan adyacentes a las secuencias adaptadoras, permitiendo así la secuenciación cuantitativa en lecturas de secuencias cortas, para caracterizar la población de moldes mediante el criterio de la cantidad relativa de cada secuencia molde única que está presente.

Cuando tal secuenciación cuantitativa directa indica que uno o más oligonucleótidos concretos pueden estar representados de manera anormalmente alta o anormalmente baja en una preparación de la composición molde, el ajuste de la composición molde puede realizarse en consecuencia para obtener una composición molde en la que todos los oligonucleótidos están presentes en cantidades sustancialmente equimolares. La composición molde en donde todos los oligonucleótidos están presentes en cantidades sustancialmente equimolares se puede utilizar a continuación como un patrón de calibración para conjuntos de cebadores de amplificación, tal como en los métodos descritos actualmente para determinar y corregir el potencial de amplificación no uniforme entre miembros de un conjunto de cebadores.

Además de las secuencias adaptadoras descritas en los Ejemplos e incluidas en las secuencias molde ilustrativas en la Lista de Secuencias (p.ej., en los extremos 5' y 3' de SEQ ID NO: 1-1630), otras secuencias de oligonucleótidos que se pueden utilizar como secuencias adaptadoras universales serán conocidas por los expertos en la técnica a la vista de la presente descripción, incluyendo la selección de secuencias de oligonucleótidos adaptadores que son distintas de las secuencias encontradas en otras porciones de los moldes descritos en la presente memoria. Los ejemplos no limitantes de secuencias adaptadoras adicionales se muestran en la Tabla 3 y se exponen en SEQ ID NO: 1710-1731.

**Tabla 3. Secuencias de adaptadores ilustrativas**

Nombre del adaptador (cebador)	Secuencia	SEQ ID NO:
Promotor T7	AATACGACTCACTATAGG	1710
Terminador T7	GCTAGTTATTGCTCAGCGG	1711
T3	ATTAACCCTCACTAAAGG	1712
SP6	GATTTAGGTGACACTATAG	1713
M13F(-21)	TGTA AACGACGCCAGT	1714
M13F(-40)	GTTTTCCAGTCACGAC	1715
M13R Inverso	CAGAAACAGCTATGACC	1716
AOX1 Directo	GACTGGTTCCAATTGACAAGC	1717
AOX1 Inverso	GCAATGGCATTCTGACATCC	1718
pGEX Directo (GST 5, pGEX 5')	GGGCTGGCAAGCCAGTTTGGTG	1719
pGEX Inverso (GST 3, pGEX 3')	CCGGGAGCTGCATGTGTCAGAGG	1720
BGH Inverso	AACTAGAAGGCACAGTCGAGGC	1721
GFP (C' terminal, CFP, YFP o BFP)	CACTCTCGGCATGGACGAGC	1722

GFP Inverso	TGGTGCAGATGAACTTCAGG	1723
GAG	GTTCGACCCCGCCTCGATCC	1724
GAG Inverso	TGACACACATTCCACAGGGTC	1725
CYC1 Inverso	GCGTGAATGTAAGCGTGAC	1726
pFastBacF	5'-d(GGATTATTCATACCGTCCCA)-3'	1727
pFastBacR	5'-d(CAAATGTGGTATGGCTGATT)-3'	1728
pBAD Directo	5'-d(ATGCCATAGCATT TTTATCC)-3'	1729
pBAD Inverso	5'-d(GATTTAATCTGTATCAGG)-3'	1730
CMV-Directo	5'-d(CGCAAATGGGCGGTAGGCGTG)-3'	1731

5 CÓDIGOS DE BARRAS. Como se describe en la presente memoria, ciertas realizaciones contemplan el diseño de las secuencias oligonucleotídicas molde para que contengan secuencias de firmas cortas que permitan una identificación inequívoca de la secuencia molde, y por lo tanto de al menos un cebador responsable de amplificar ese molde, sin tener que secuenciar el producto de amplificación completo. En el molde descrito, los oligonucleótidos de fórmula general (I), B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 o más nucleótidos contiguos (incluyendo todos los valores enteros entre ellos), en donde en cada una de la pluralidad de secuencias oligonucleotídicas molde B comprende una secuencia oligonucleotídica única que identifica de manera única, como una combinación emparejada, (i) la secuencia oligonucleotídica V única del oligonucleótido molde y (ii) la secuencia oligonucleotídica J única del oligonucleótido molde.

15 Así, por ejemplo, los oligonucleótidos molde que tienen secuencias identificadoras código de barras pueden permitir lecturas de secuencia de producto de amplificación relativamente cortas, tales como lecturas de secuencia código de barras de no más de 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 55, 50, 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4 o menos nucleótidos, seguido del emparejamiento de esta información de secuencia código de barras con las secuencias V y J asociadas que se incorporan al molde que tiene el código de barras como parte del diseño del molde. Con este enfoque, se puede secuenciar parcialmente una gran cantidad de productos de amplificación de manera simultánea mediante secuenciación paralela de alto rendimiento, para identificar cebadores que son responsables del sesgo de amplificación en un conjunto de cebadores complejo.

25 Los códigos de barras ilustrativos pueden comprender un primer oligonucleótido código de barras de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 nucleótidos que identifica de forma única cada polinucleótido V en el molde y un segundo oligonucleótido código de barras de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 nucleótidos que identifican de forma única cada polinucleótido J en el molde, para proporcionar códigos de barras de, respectivamente, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 o 32 nucleótidos de longitud, pero no se pretende que estas realizaciones y realizaciones relacionadas sean limitantes. Los oligonucleótidos de códigos de barras pueden comprender secuencias de oligonucleótidos de cualquier longitud, siempre que se obtenga una longitud código de barras mínima que impida la aparición de una secuencia código de barras dada en dos o más oligonucleótidos molde que tengan otras secuencias diferentes (p.ej., secuencias V y J).

35 Por lo tanto, la longitud mínima del código de barras, para evitar dicha redundancia entre los códigos de barras que se utilizan para identificar de forma única diferentes pares de secuencias V-J, es  $4^X$  nucleótidos, donde  $4^X$  es mayor que la cantidad de especies moldes distintas que se deben diferenciar basándose en que tienen secuencias no idénticas. Por ejemplo, para el conjunto de 871 oligonucleótidos molde expuestos en la presente memoria como SEQ ID NO: 1-871, la longitud mínima del código de barras sería de cinco nucleótidos, lo que permitiría un total teórico de 1024 (es decir, más de 871) diferentes secuencias de pentanucleótidos posibles. En la práctica, las longitudes de lectura de la secuencia de oligonucleótido código de barras pueden limitarse únicamente por los límites de longitud de lectura de secuencia del aparato de secuenciación de nucleótidos que se vaya a emplear. Para ciertas realizaciones, diferentes oligonucleótidos de códigos de barras que distinguirán especies individuales de oligonucleótidos molde deben tener emparejamientos erróneos de al menos dos nucleótidos (p.ej., una distancia mínima de Hamming de 2) cuando se alinea para maximizar la cantidad de nucleótidos que coinciden en posiciones concretas en las secuencias de oligonucleótidos de códigos de barras.

45 En realizaciones preferidas, para cada especie de oligonucleótido molde distinta que tiene una secuencia única dentro de la composición molde de fórmula general (I), B1, B2, B3 y B4 serán idénticos.

El experto en la técnica estará familiarizado con el diseño, la síntesis y la incorporación a una construcción de oligonucleótidos o polinucleótidos más grande, de secuencias código de barras de oligonucleótidos de, por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 200, 300, 500 o más nucleótidos contiguos, incluyendo todos los valores enteros entre ellos. Para ejemplos no limitantes del diseño e implementación de estrategias de identificación de secuencia código de barras de oligonucleótidos, véanse, p.ej., De Career et al., 2011 Adv. Env. Microbiol. 77:6310; Parameswaran et al., 2007 Nucl. A.C. Res. 35(19):330; Roh et al., 2010 Trends Biotechnol. 28:291.

Por lo general, los códigos de barras se colocan en moldes en lugares donde no se encuentran de forma natural, es decir, los códigos de barras comprenden secuencias de nucleótidos que son distintas de cualquier secuencia oligonucleotídica de origen natural que pueda encontrarse en la proximidad de las secuencias adyacentes a las que están situados los códigos de barras (p.ej., secuencias V y/o J). Tales secuencias código de barras se pueden incluir, de acuerdo con ciertas realizaciones descritas en la presente memoria, como elementos B1, B2 y/o B3 del oligonucleótido molde de la fórmula general (I) descrito actualmente. Por consiguiente, algunos de los oligonucleótidos molde descritos en la presente memoria de fórmula general (I) también pueden comprender, en ciertas realizaciones, uno, dos o los tres códigos de barras B1, B2 y B3, mientras que en ciertas otras realizaciones algunos o todos estos códigos de barras pueden estar ausentes. En ciertas realizaciones, todas las secuencias código de barras tendrán un contenido de GC idéntico o similar (p.ej., difiriendo en el contenido de GC en no más de 20% o en no más de 19, 18, 17, 16, 15, 14, 13, 12, 11 o 10%).

En las composiciones molde de acuerdo con ciertas realizaciones descritas en la presente memoria, el elemento B que contiene el código de barras (p.ej., B1, B2, B3 y/o B4) comprende la secuencia oligonucleotídica que identifica de forma única una única combinación V-J emparejada. Opcionalmente y en ciertas realizaciones, el elemento B que contiene un código de barras también puede incluir un nucleótido aleatorio o una secuencia polinucleotídica aleatoria de al menos 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, 45, 50, 55, 60, 70, 80, 90, 100, 200, 300, 500 o más nucleótidos contiguos, situados aguas arriba y/o aguas abajo de la secuencia código de barras específica que identifica de forma única cada combinación VJ emparejada específica. Cuando están presentes tanto aguas arriba como o aguas abajo de la secuencia código de barras específica, la secuencia de nucleótidos aleatoria o de polinucleótidos aleatoria son independientes entre sí, es decir, pueden comprender, pero no es necesario que comprendan la misma secuencia nucleótidos o la misma secuencia de polinucleótidos.

**SITIOS PARA ENZIMAS DE RESTRICCIÓN.** De acuerdo con ciertas realizaciones descritas en la presente memoria, el oligonucleótido molde puede comprender un sitio de reconocimiento de endonucleasa de restricción (RE) que está situado entre las secuencias V y J y no se produce en ninguna otra parte de la secuencia oligonucleotídica molde. El sitio de reconocimiento de RE puede estar opcionalmente adyacente a un sitio código de barras que identifica la secuencia de la región V. El sitio RE se puede incluir para cualquiera de una serie de propósitos, incluyendo sin limitación, como característica estructural que puede explotarse para destruir moldes de forma selectiva contactándolos con la enzima de restricción apropiada. Puede ser deseable degradar los presentes oligonucleótidos molde selectivamente poniéndolos en contacto con un RE adecuado, por ejemplo, para eliminar oligonucleótidos molde de otras composiciones en las que pueden haberse introducido deliberada o accidentalmente. Alternativamente, el sitio RE puede explotarse útilmente en el transcurso de la secuenciación de oligonucleótidos molde en la composición molde, y/o como un marcador de secuencia posicional en una secuencia oligonucleotídica molde independientemente de si se escinde o no con una enzima de restricción. Un sitio RE ilustrativo es el motivo oligonucleotídico GTCGAC, que es reconocido por la enzima de restricción Sal I. Se conoce en la técnica un gran número de enzimas de restricción adicionales y sus respectivas secuencias de sitios de reconocimiento de RE y están disponibles comercialmente (p.ej., New England Biolabs, Beverly, MA). Estas incluyen, por ejemplo, EcoRI (GAATTC) y SphI (GCATGC). Aquellos que estén familiarizados con la técnica apreciarán que se puede incorporar cualquiera de una variedad de tales sitios de reconocimiento de RE a realizaciones concretas de los oligonucleótidos molde descritos actualmente.

### **Secuenciación**

La secuenciación se puede realizar utilizando cualquiera de una variedad de máquinas y sistemas de secuenciación de moléculas individuales de alto rendimiento disponibles. Los sistemas de secuencia ilustrativos incluyen sistemas de secuencia por síntesis tales como el analizador de genoma Illumina y aparatos asociados (Illumina, Inc., San Diego, CA), el sistema de análisis genético de Helicos (Helicos BioSciences Corp., Cambridge, MA), Pacific BioSciences PacBio RS (Pacific Biosciences, Menlo Park, CA) u otros sistemas con capacidades similares. La secuenciación se logra utilizando un conjunto de oligonucleótidos de secuenciación que hibridan con una región definida dentro de las moléculas de ADN amplificadas. Los oligonucleótidos de secuenciación se diseñan de manera que los segmentos génicos que codifican V y J puedan identificarse de forma única mediante las secuencias que se generan basándose en la presente descripción y a la vista de secuencias de genes del receptor inmunitario adaptativo conocidas que aparecen en bases de datos disponibles públicamente. Véanse, p.ej., los documentos US 2012/0058902 A1, US 2010/0330571 A1, WO2011/106738 o WO2012/027503. Los cebadores de secuenciación de



la región J de TCRB ilustrativos se exponen en la Tabla 4:

TABLA 4: CEBADORES DE SECUENCIACIÓN DE TCRBJ

CEBADOR	SECUENCIA	SEQ ID NO:
> Jseq1-1	ACAACACTGTGAGTCTGGTGCCTTGTCCAAAGAAA	1696
> Jseq1-2	ACAACGGTTAACCTGGTCCCCGAACCGAAGGTG	1697
> Jseq1-3	ACAACAGTGAGCCAACCTCCCTCTCCAAAATAT	1698
> Jseq1-4	AAGACAGAGAGCTGGGTTCCACTGCCAAAAAAC	1699
> Jseq1-5	AGGATGGAGAGTCGAGTCCCATCACAAAATGC	1700
> Jseq1-6	GTCACAGTGAGCCTGGTCCCGTTCCCAAAGTGG	1701
> Jseq2-1	AGCACGGTGAGCCGTGTCCCTGGCCCGAAGAAC	1702
> Jseq2-2	AGTACGGTCAGCCTAGAGCCTTCTCCAAAAAAC	1703
> Jseq2-3	AGCACTGTCAGCCGGGTGCCTGGGCCAAAATAC	1704
> Jseq2-4	AGCACTGAGAGCCGGGTCCCGGCGCCGAAGTAC	1705
> Jseq2-5	AGCACCAGGAGCCGCGTGCCTGGCCCGAAGTAC	1706
> Jseq2-6	AGCACGGTCAGCCTGCTGCCGGCCCCGAAAGTC	1707
> Jseq2-7	GTGACCGTGAGCCTGGTGCCCGGCCCGAAGTAC	1708

- 5 En ciertas realizaciones, los segmentos génicos que codifica la región J amplificada pueden tener cada uno una etiqueta identificadora definida por la secuencia única de 2, 3, 4, 5, 6, 7, 8, 9, 10 o aproximadamente 15, 20 o más nucleótidos, situados en una posición definida con respecto a un sitio RSS. Por ejemplo, se puede utilizar una etiqueta de cuatro bases, en el segmento que codifica la región J $\beta$  de las regiones que codifican CDR3 de TCR $\beta$  amplificadas, en las posiciones +11 a +14 aguas abajo del sitio RSS. Sin embargo, no es necesario que estas realizaciones y realizaciones relacionadas estén tan limitadas y también contemplan otras etiquetas identificadoras definidas por secuencia de nucleótidos relativamente cortas que pueden detectarse en segmentos génicos que codifican la región J y definidas basándose en sus posiciones relativas a un sitio RSS. Éstas pueden variar entre loci que codifican diferentes receptores inmunitarios adaptativos.
- 10
- 15 De acuerdo con esto, los oligonucleótidos de secuenciación pueden hibridar adyacentes a una etiqueta de cuatro bases dentro de los segmentos génicos que codifican J amplificados en las posiciones +11 a +14 aguas abajo del sitio RSS. Por ejemplo, los oligonucleótidos de secuenciación para TCRB se pueden diseñar para que se reasocian con un motivo de nucleótido consenso observado inmediatamente aguas abajo de esta "etiqueta", de modo que las primeras cuatro bases de una lectura de secuencia identificarán de forma única el segmento génico que codifica J (véase, p.ej. el documento WO/2012/027503).
- 20

La longitud promedio de la región codificante de CDR3, para el TCR, definida como los nucleótidos que codifican el polipéptido TCR entre la segunda cisteína conservada del segmento V y la fenilalanina conservada del segmento J, es de 35 +/- 3 nucleótidos. De acuerdo con esto y en ciertas realizaciones, la amplificación por PCR utilizando cebadores oligonucleotídicos del segmento V con cebadores oligonucleotídicos del segmento J que se inicia a partir de la etiqueta del segmento J de una región J de TCR o de IgH concreta (p.ej., J $\beta$  de TCR, J $\gamma$  de TCR o JH de IgH como se describe en la presente memoria) capturarán casi siempre la unión V-D-J completa en una lectura de 50 pares de bases. La longitud promedio de la región CDR3 de IgH, definida como los nucleótidos entre la cisteína conservada en el segmento V y la fenilalanina conservada en el segmento J, está menos restringida que en el locus TCR $\beta$ , pero típicamente estará entre aproximadamente 10 y aproximadamente 70 nucleótidos. Por consiguiente, y en ciertas realizaciones, la amplificación por PCR utilizando cebadores oligonucleotídicos del segmento V con cebadores oligonucleotídicos del segmento J que comienza en la etiqueta del segmento J de IgH capturarán la unión V-D-J completa en una lectura de 100 pares de bases.

25

30

Los cebadores de PCR que se reasocian y soportan la extensión de polinucleótidos en secuencias molde emparejadas erróneamente se denominan cebadores promiscuos. En ciertas realizaciones, los cebadores de PCR inversa del segmento J de TCR e Ig pueden diseñarse para minimizar el solapamiento con los oligonucleótidos de secuenciación, con el fin de minimizar el cebado promiscuo en el contexto de la PCR múltiple. En una realización, los cebadores inversos del segmento J de TCR e Ig pueden estar anclados al extremo 3' mediante reasociación con el motivo del sitio de empalme consenso, con solapamiento mínimo de los cebadores de secuenciación. En general, los cebadores del segmento V y J de TCR e Ig se pueden seleccionar para operar en la PCR a temperaturas de reasociación compatibles utilizando programas de diseño y análisis de secuencia/cebador conocidos bajo

35

40

parámetros por defecto.

Para la reacción de secuenciación, los cebadores de secuenciación de IGHJ ilustrativos se extienden tres nucleótidos a través de las secuencias de CAG conservadas como se describe en el documento WO/2012/027503.

5

### Determinación del factor de amplificación

Ciertas realizaciones contemplan el uso de las composiciones molde sintéticas descritas en la presente memoria para determinar los factores de amplificación para estimar la cantidad de secuencias codificantes de receptores inmunitarios adaptativos reordenados en una muestra. Estas realizaciones y realizaciones relacionadas se pueden utilizar para cuantificar la cantidad de secuencias codificantes del receptor inmunitario adaptativo en una muestra de ADN que se ha obtenido de células linfoides, incluyendo células linfoides que están presentes en una mezcla de células que comprende células en las que el ADN que codifica un receptor inmunitario adaptativo ha experimentado reordenamiento de ADN, pero en donde la muestra también contiene ADN de células en las que no se ha producido dicho reordenamiento (p.ej., células no linfoides, células linfoides inmaduras y/o no linfoides, células mesenquimales, células cancerosas, etc.).

10

15

20

25

El número total de miembros diferentes de una clase dada de receptores inmunitarios adaptativos (p.ej., TCR o IG) en un sujeto se pueden estimar mediante PCR multiplexada utilizando un conjunto completo de cebadores de amplificación V-J seguido de secuenciación cuantitativa de productos de amplificación. La amplificación multiplexada y la secuenciación de alto rendimiento de las secuencias de ADN que codifican TCR y BCR (IG) reordenadas son descritas, por ejemplo, por Robins et al., 2009 Blood 114, 4099; Robins et al., 2010 Sci. Translat. Med. 2:47ra64; Robins et al., 2011 J. Immunol. Meth. doi: 10.1016/j.jim.2011.09. 001; Sherwood et al. 2011 Sci. Translat. Med. 3:90ra61; Publicación de Estados Unidos Núm. 2012/0058902, Publicación de Estados Unidos Núm. 2010/0330571, documento WO/2010/151416, documento WO/2011/106738 y documento WO2012/027503.

25

30

35

40

45

50

55

Esta metodología típicamente implica tomar muestras de ADN de una subpoblación de células linfoides, tales como células linfoides que están presentes en una muestra de sangre, que también se sabe que contienen células nucleadas que carecen de ADN que codifica TCR o IG reordenado. Las presentes composiciones y métodos pueden permitir una mejora de la exactitud y la precisión en la determinación de la cantidad de moléculas de ADN que codifican TCR e IG reordenadas en dicha muestra. Como se describe en la presente memoria, por ejemplo, "incorporando" (añadiendo) la muestra de ADN con la presente composición molde, se proporciona un patrón molde de amplificación interno para evaluar las eficacias relativas a través del intervalo de cebadores oligonucleotídicos que están presentes en el conjunto de cebadores de amplificación multiplexado. Al evaluar de este modo los productos de amplificación de la presente composición molde artificial, que se añade a la reacción de amplificación en cantidades conocidas, se puede determinar un factor de amplificación (p.ej., un factor multiplicador, normalizador, de aumento o disminución a escala o geométrico, etc.) para el conjunto de cebadores de amplificación oligonucleotídicos y a continuación se puede utilizar para calcular la cantidad de moldes de ADN natural en la muestra.

40

45

50

55

60

Como otro ejemplo, estas realizaciones y realizaciones relacionadas permiten la cuantificación de enfermedad residual mínima (ERM) en linfoma o leucemia, mediante la detección cuantitativa de ADN que codifica TCR reordenado en muestras obtenidas de preparaciones mixtas de células linfoides y no linfoides, incluyendo células de linfoma o leucemia persistentes. Los métodos anteriores determinan la ERM como el número de células malignas que son detectables como una proporción del número total de células en una muestra. Por el contrario, los presentes métodos permiten la estimación del número total de células en una muestra que tiene ADN que codifica TCR que ha experimentado reordenamiento, de modo que las células malignas (p.ej., aquellas que tienen un reordenamiento de TCR concreto, tal como un clonotipo) se pueden cuantificar como una proporción de tales células reordenadas en lugar de como una proporción de todas las células. A modo de teoría no limitante, se cree que debido a que la representación de todas las células reordenadas en una muestra clínica de un sujeto que tiene o se sospecha que tiene ERM es típicamente muy baja, los presentes métodos mejorarán drásticamente la sensibilidad con la que se puede detectar ERM, incluyendo la mejora de tal sensibilidad al aumentar la razón señal/ruido.

50

55

60

65

En la presente memoria se describe un método para cuantificar moléculas de ADN reordenado que codifican uno o una pluralidad de receptores inmunitarios adaptativos en una muestra biológica que comprende ADN de células linfoides de un sujeto, comprendiendo cada receptor inmunitario adaptativo una región variable y una región de empalme. En resumen, el método comprende las etapas de:

55

60

(A) en una reacción de amplificación multiplexada que utiliza el conjunto de cebadores de amplificación oligonucleotídicos descrito que es capaz de amplificar sustancialmente todas las combinaciones codificantes de V-J para un receptor inmunitario adaptativo dado, amplificar el ADN de la muestra a la que se ha añadido una cantidad conocida de la composición molde descrita en la presente memoria para la determinación del factor de amplificación, para obtener productos de amplificación;

(B) secuenciar cuantitativamente los productos de amplificación de (A) para cuantificar (i) productos de

amplificación del molde, que son productos de amplificación de la composición molde descrita en la presente memoria y serán identificables porque contienen al menos una secuencia de oligonucleótido código de barras, y (ii) productos de amplificación de las secuencias de ADN que codifican el receptor inmunitario adaptativo reordenadas en la muestra, que serán identificables porque contienen secuencias V y J específicas pero

carecen de una secuencia de oligonucleótido código de barras;  
 (C) calcular un factor de amplificación basado en información cuantitativa obtenida en la etapa (B); y  
 (D) utilizar el factor de amplificación de (C) para determinar, mediante cálculo, la cantidad de moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra.

Sin desear estar limitados por la teoría, de acuerdo con estos métodos y con métodos relacionados, se mide la cantidad de moléculas de ADN que codifican TCR o IG reordenadas que se muestrean en una reacción de amplificación multiplexada. Para hacerlo, se determina y se promedia un valor de cobertura de secuencia, p.ej., la cantidad de lecturas de secuencia de salida que se determinan para cada molécula de entrada (molde) a través del número completo de oligonucleótidos molde diferentes que están presentes, para obtener un valor de cobertura de secuencia promedio. Al dividir (i) la cantidad de lecturas que se obtienen para una secuencia dada por (ii) el valor de cobertura de secuencia promedio, se puede calcular la cantidad de moléculas reordenadas que están presentes como moldes al comienzo de la reacción de amplificación.

Así, por ejemplo, para calcular el valor de cobertura de la secuencia, a cada amplificación por PCR se le añade una cantidad conocida de un conjunto de moléculas sintéticas de la composición molde descrita actualmente, teniendo los moldes sintéticos la estructura básica de fórmula (I) 5' U-B1-V-B2-R-(B3)-J-B4-U 3' donde cada V es un segmento de 300 pares de bases que tiene una secuencia que coincide con una secuencia del gen V de TCR o IG y J es un segmento de 100 pares de bases que tiene una secuencia que coincide con un gen J de TCR o IG. B2 es una secuencia de oligonucleótido código de barras única que identifica cada par VJ y que también diferencia los productos de amplificación de los moldes sintéticos de ADN (que contendrán la secuencia código de barras) de los productos de amplificación de moléculas de ADN molde biológicas de origen natural que son aportadas por la muestra de ADN linfocitario (que carecerá de la secuencia código de barras). En este ejemplo, B3 de fórmula (I) no es nada. Después de la amplificación y secuenciación por PCR, se cuenta el número de cada molécula sintética secuenciada (es decir, los productos de amplificación que contienen la secuencia código de barras). La cobertura de la secuencia de las moléculas sintéticas se calcula a continuación basándose en el número conocido de moléculas molde sintéticas de partida utilizadas para incorporarlas a la reacción de amplificación.

Por ejemplo, se puede añadir a la reacción de amplificación un conjunto de 5000 moléculas molde sintéticas que contienen código de barras que comprenden 4-5 copias de cada una de las 1100 secuencias oligonucleotídicas molde sintéticas únicas (que representan cada posible par VJ). Si los productos de amplificación incluyen 50.000 secuencias que coinciden con las moléculas molde sintéticas, se ha obtenido un valor de cobertura de secuencia de 10X y el factor de amplificación es 10. Para estimar la cantidad de moléculas molde VDJ reordenadas naturales en el ADN obtenido de la muestra, el número de productos de amplificación de los moldes naturales (es decir, los productos de amplificación que carecen de cualquier secuencia código de barras) se divide por el factor de amplificación. Para mayor precisión, debido a que en este ejemplo las 5000 moléculas sintéticas son un conjunto complejo de 1100 moléculas que representan cada par VJ, el factor de amplificación para cada par VJ se puede calcular individualmente. El factor de amplificación se puede promediar a través de todas las moléculas sintéticas (Fig. 7). La precisión y robustez del método se muestran en la Fig. 8 y los detalles se describen a continuación en el Ejemplo 2.

Por consiguiente, en estos casos, el método comprende:

(A) amplificar ADN en una reacción en cadena de polimerasa múltiple (PCR) que comprende: (1) ADN de la muestra biológica que comprende células linfocitarias del sujeto, (2) la composición molde para la determinación del factor de amplificación de fórmula general (I) como se describe en la presente memoria, en la que está presente un número conocido de cada uno de la pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única, (3) un conjunto de cebadores de amplificación oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica uno o una pluralidad de receptores inmunitarios adaptativos en el ADN de la muestra biológica, comprendiendo el conjunto de cebadores: (a) en cantidades sustancialmente equimolares, una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V de receptor inmunitario adaptativo o con el complemento del mismo, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico que codifica la región V del receptor inmunitario adaptativo y en donde la pluralidad de cebadores del segmento V hibrida específicamente con todos los segmentos génicos funcionales que codifican la región V del receptor inmunitario adaptativo que están presentes en la composición molde, b) en cantidades sustancialmente equimolares, una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor

inmunitario adaptativo o con el complemento del mismo, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica la región J del receptor inmunitario adaptativo y en donde la pluralidad de cebadores del segmento J hibridan específicamente sustancialmente con todos los segmentos génicos funcionales que codifican la región J del receptor inmunitario adaptativo que están presentes en la composición molde, en donde los cebadores oligonucleotídicos del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción de cadena de polimerasa múltiple (PCR) de (i) sustancialmente todos los oligonucleótidos molde en la composición molde para producir una multiplicidad de moléculas de ADN molde amplificadas, siendo suficiente dicha multiplicidad de moléculas de ADN molde amplificadas para cuantificar la diversidad de oligonucleótidos molde en la composición molde, y (ii) sustancialmente todas las moléculas de ADN reordenado que codifican receptores inmunitarios adaptativos en la muestra biológica para producir una multiplicidad de moléculas de ADN reordenado amplificado, siendo suficiente dicha multiplicidad de ADN reordenado amplificado las moléculas son suficientes para cuantificar la diversidad de las moléculas de ADN reordenado en el ADN de la muestra biológica, y en donde cada molécula de ADN amplificado en la multiplicidad de moléculas de ADN molde amplificado y en la multiplicidad de moléculas de ADN reordenado amplificado es menor que 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80 o 70 nucleótidos de longitud;

(B) secuenciar cuantitativamente toda o una porción suficiente de cada una de dichas moléculas de ADN molde amplificado y cada una de dichas moléculas de ADN reordenado amplificado para cuantificar (i) una cantidad de producto molde de moléculas de ADN molde amplificado que contienen al menos una secuencia código de barras de oligonucleótido, y (ii) una cantidad de producto reordenado de moléculas de ADN reordenado amplificado que carecen de una secuencia código de barras de oligonucleótido;

(C) calcular un factor de amplificación dividiendo la cantidad de producto molde de (B) (i) por la cantidad conocida de cada uno de la pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única de (A) (2); y

(D) dividir la cantidad de producto reordenado de (B)(ii) por el factor de amplificación calculado en (C) para cuantificar las moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra.

No se pretende que las realizaciones contempladas se limiten al método descrito anteriormente, de manera que a partir de la presente descripción, el experto en la técnica apreciará las variaciones que pueden emplearse. Un enfoque alternativo, por ejemplo, puede no utilizar la composición molde sintética descrita en la presente memoria como un molde de control incorporado a la amplificación por PCR multiplexada de una muestra de ADN que contiene TCR de células linfoides reordenado y/o ADN que codifica IG, así como ADN no reordenado. En cambio, de acuerdo con una de tales alternativas, a la reacción de amplificación que utiliza cebadores de amplificación de V y J se le puede añadir un conjunto conocido de cebadores de amplificación oligonucleotídicos que amplifican una región de secuencia genómica distinta, altamente conservada. Estos cebadores de control genómico pueden amplificar cada genoma que está presente en la muestra de ADN independientemente de si contiene o no secuencias codificantes de TCR y/o IG, mientras que los cebadores de V y J pueden amplificar productos solo de genomas con una región VDJ reordenada. La razón entre estas dos clases de moléculas de producto de amplificación permite la estimación del número total de genomas de células B en la muestra.

Se contempla adicionalmente para estas y otras realizaciones relacionadas de cualquiera de los métodos descritos en la presente memoria que dicho método pueda comprender adicionalmente la secuenciación de las moléculas de ADN que codifican el receptor inmunitario amplificado que se producen. Las composiciones y los métodos para la secuenciación de las secuencias génicas del receptor inmunitario adaptativo reordenado y para la determinación del clonotipo del receptor inmunitario adaptativo son descritos por Robins et al., 2009 Blood 114, 4099; Robins et al., 2010 Sci. Translat. Med. 2: 47ra64; Robins et al., 2011 J. Immunol. Meth. doi: 10.1016/j.jim.2011.09. 001; Sherwood et al. 2011 Sci. Translat. Med. 3:90ra61; Publicación de Estados Unidos Núm. 2012/0058902, Publicación de Estados Unidos Núm. 2010/0330571, documento WO/2010/151416, documento WO/2011 /106738 y documento WO2012/027503, incorporados a la presente memoria como referencia. También se pueden encontrar detalles sobre la secuenciación de productos de amplificación por PCR, procesamiento de datos de secuenciación y usos de mediciones de diversidad de receptores inmunitarios adaptativos, todos los cuales se pueden emplear para su uso de acuerdo con los métodos descritos en la presente memoria.

De acuerdo con una teoría no limitante, estas realizaciones explotan la comprensión actual en la técnica (también descrita anteriormente) de que una vez que un linfocito T ha reordenado sus genes que codifican TCR, sus células de progeie poseen el mismo reordenamiento del gen que codifica el receptor inmunitario adaptativo, dando lugar así a una población clonal que puede identificarse de manera única por la presencia en la misma de segmentos de genes V y J que codifican CDR3 reordenados que pueden amplificarse mediante una combinación de pares específica de cebadores oligonucleotídicos específicos de V y J como se describe en la presente memoria.

La práctica de ciertas realizaciones de la presente invención empleará, a menos que se indique específicamente lo contrario, métodos convencionales en técnicas de microbiología, biología molecular, bioquímica, genética molecular, biología celular, virología e inmunología que están dentro del conocimiento práctico de la técnica, y referencia a

varios de los cuales se hace a continuación con el propósito de ilustración. Tales técnicas se explican completamente en la bibliografía. Véanse, por ejemplo, Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (3ª Edición, 2001); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2ª Edición, 1989); Maniatis et al., *Molecular Cloning: A Laboratory Manual* (1982); Ausubel et al., *Current Protocols in Molecular Biology* (John Wiley and Sons, actualizado en julio de 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates y Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford Univ. Press USA, 1985); *Current Protocols in Immunology* (Editado por: John E. Coligan, Ada M. Kruisbeek, David H. Margulies, Ethan M. Shevach, Warren Strober 2001 John Wiley & Sons, NY, NY); *Real-Time PCR: Current Technology y Applications*, Edited by Julie Logan, Kirstin Edwards y Nick Saunders, 2009, Caister Academic Press, Norfolk, UK; Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); Guthrie y Fink, *Guide to Yeast Genetics y Molecular Biology* (Academic Press, New York, 1991); *Oligonucleotide Synthesis* (N. Gait, Ed., 1984); *Nucleic Acid Hybridization* (B. Hames y S. Higgins, Eds., 1985); *Transcription y Translation* (B. Hames & S. Higgins, Eds., 1984); *Animal Cell Culture* (R. Freshney, Ed., 1986); Perbal, *A Practical Guide to Molecular Cloning* (1984); *Next-Generation Genome Sequencing* (Janitz, 2008 Wiley-VCH); *PCR Protocols (Methods in Molecular Biology)* (Park, Ed., 3ª Edición, 2010 Humana Press); *Immobilized Cells And Enzymes* (IRL Press, 1986); el tratado, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller y M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Harlow y Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998); *Immunochemical Methods In Cell And Molecular Biology* (Mayer y Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir y CC Blackwell, eds., 1986); Riott, *Essential Immunology*, 6th Edition, (Blackwell Scientific Publications, Oxford, 1988); *Embryonic Stem Cells: Methods y Protocols (Methods in Molecular Biology)* (Kurstad Turksen, Ed., 2002); *Embryonic Stem Cell Protocols: Volume I: Isolation y Characterization (Methods in Molecular Biology)* (Kurstad Turksen, Ed., 2006); *Embryonic Stem Cell Protocols: Volume II: Differentiation Models (Methods in Molecular Biology)* (Kurstad Turksen, Ed., 2006); *Human Embryonic Stem Cell Protocols (Methods in Molecular Biology)* (Kursad Turksen Ed., 2006); *Mesenchymal Stem Cells: Methods y Protocols (Methods in Molecular Biology)* (Darwin J. Prockop, Donald G. Phinney, y Bruce A. Bunnell Eds., 2008); *Hematopoietic Stem Cell Protocols (Methods in Molecular Medicine)* (Christopher A. Klug, y Craig T. Jordan Eds., 2001); *Hematopoietic Stem Cell Protocols (Methods in Molecular Biology)* (Kevin D. Bunting Ed., 2008) *Neural Stem Cells: Methods y Protocols (Methods in Molecular Biology)* (Leslie P. Weiner Ed., 2008).

A menos que se proporcionen definiciones específicas, la nomenclatura utilizada con respecto a, y los procedimientos y técnicas de laboratorio de, biología molecular, química analítica, química orgánica sintética y química médica y farmacéutica descritos en la presente memoria son los bien conocidos y comúnmente utilizados en la técnica. Se pueden utilizar técnicas convencionales para tecnología recombinante, biología molecular, microbiología, síntesis química, análisis químicos, preparación farmacéutica, formulación y suministro, y tratamiento de pacientes.

### Sistema informático

La FIG. 12 es un diagrama de bloques de alto nivel que ilustra una vista funcional de un sistema informático típico 1200 descrito en la presente memoria. Se ilustran al menos un procesador 1202 acoplado a un bus 1204. También acoplado al bus 1204 hay una memoria 1206, un dispositivo de almacenamiento 1208, un teclado 1210, un adaptador de gráficos 1212, un dispositivo señalador 1214 y un adaptador de red 1216. Una pantalla 1218 está acoplada al adaptador de gráficos 1212.

El procesador 1202 puede ser cualquier procesador de propósito general tal como una CPU compatible INTEL x86, SUN MICROSYSTEMS SPARC o con POWERPC. El dispositivo de almacenamiento 1208 es, en un caso, una unidad de disco duro, pero también puede ser cualquier otro dispositivo capaz de almacenar datos, tal como un disco compacto (CD) o DVD grabables, o un dispositivo de memoria de estado sólido. La memoria 1206 puede ser, por ejemplo, soporte lógico inalterable "firmware", memoria de solo lectura (ROM), memoria de acceso aleatorio no volátil (NVRAM) y/o RAM, y contiene instrucciones y datos utilizados por el procesador 202. El dispositivo señalador 1214 puede ser un ratón, bola de seguimiento u otro tipo de dispositivo señalador, y se utiliza combinado con el teclado 1210 para introducir datos en el sistema informático 1200. El adaptador gráfico 1212 muestra imágenes y otra información en la pantalla 1218. El adaptador de red 1216 acopla el sistema informático 1200 a la red.

Como se conoce en la técnica, el sistema informático 1200 está adaptado para ejecutar módulos de programas informáticos. Según se utiliza en la presente memoria, el término "módulo" se refiere a la lógica y/o los datos del programa informático para proporcionar la funcionalidad especificada. Se puede implementar un módulo en el equipo, soporte lógico inalterable y/o soporte lógico. En un caso, los módulos se almacenan en el dispositivo de almacenamiento 1208, se cargan en la memoria 1206 y se ejecutan en el procesador 1202. Los diversos métodos y etapas ilustrados en la presente memoria pueden ejecutarse por medio de los componentes del sistema informático 1200.

Las descripciones y representaciones algorítmicas son comúnmente utilizadas por los expertos en las técnicas de

procesamiento de datos para transmitir la esencia de su trabajo de manera eficaz a otros expertos en la materia. Se entiende que estas operaciones, aunque se describen funcionalmente, computacionalmente o lógicamente, se implementan mediante programas informáticos o circuitos eléctricos equivalentes, microcódigo o similares. Además, también ha resultado conveniente a veces, referirse a estas disposiciones de operaciones como módulos, sin pérdida de generalidad. Las operaciones descritas y sus módulos asociados pueden incorporarse en soporte lógico, soporte lógico inalterable, o equipo o cualquier combinación de los mismos.

Cualquiera de los etapas, operaciones o procedimientos descritos en la presente memoria puede realizarse o implementarse con uno o más módulos de equipo o soporte lógico, solos o en combinados con otros dispositivos. En un caso, se implementa un módulo de soporte lógico con un producto de programa informático que comprende un medio legible por ordenador que contiene código de programa informático, que puede ejecutarse mediante un procesador informático para realizar cualquiera o todas las etapas, operaciones o procedimientos descritos.

En la presente memoria se describe un aparato para realizar las operaciones de la presente memoria. Este aparato puede construirse especialmente para los fines requeridos, y/o puede comprender un dispositivo informático de propósito general activado o reconfigurado selectivamente por un programa informático almacenado en el ordenador. Tal programa informático puede almacenarse en un medio de almacenamiento legible por ordenador tangible o cualquier tipo de medio adecuado para almacenar instrucciones electrónicas, y acoplado a un bus de sistema informático. Además, todos los sistemas informáticos a los que se hace referencia en la memoria descriptiva pueden incluir un único procesador o pueden ser arquitecturas que emplean diseños de múltiples procesadores para aumentar la capacidad informática.

En la presente memoria se describe una señal de datos informáticos incorporada en una onda portadora, en donde la señal de datos informáticos incluye cualquier producto de programa informático u otra combinación de datos descrita en la presente memoria. La señal de datos del ordenador es un producto que se presenta en un medio tangible u onda portadora y se modula o codifica de otro modo en la onda portadora, que es tangible, y se transmite de acuerdo con cualquier método de transmisión adecuado.

Además, el lenguaje utilizado en la memoria descriptiva se ha seleccionado principalmente para fines de legibilidad y explicativos, y puede no haberse seleccionado para delinear o circunscribir el tema de la invención. Por lo tanto, se pretende que el alcance de la invención esté limitado no por esta descripción detallada, sino por las reivindicaciones. En consecuencia, la descripción de las realizaciones de la invención pretende ser ilustrativa, pero no limitante, del alcance de la invención.

A menos que el contexto requiera lo contrario, a lo largo de la presente memoria descriptiva y las reivindicaciones, la palabra "comprender" y sus variaciones, tales como, "comprende" y "que comprende" deben interpretarse en un sentido abierto e inclusivo, es decir, como "incluyendo, pero sin limitarse a". Por "que consiste en" se entiende que incluye, y típicamente se limita a, lo que sigue a la frase "que consiste en". Por "que consiste esencialmente en" se entiende que incluye cualquier elemento enumerado después de la frase, y se limita a otros elementos que no interfieren ni contribuyen a la actividad o acción especificada en la descripción para los elementos enumerados. Por lo tanto, la frase "que consiste esencialmente en" indica que los elementos enumerados son requeridos u obligatorios, pero que no se requieren otros elementos y pueden o no estar presentes dependiendo de si afectan o no a la actividad o acción de los elementos enumerados.

En esta memoria descriptiva y en las reivindicaciones adjuntas, las formas singulares "un", "uno", "una", "el" y "la" incluyen referencias en plural a menos que el contenido indique claramente lo contrario. Según se utiliza en la presente memoria, en realizaciones concretas, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 5%, 6%, 7%, 8% o 9%. En otras realizaciones, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 10%, 11%, 12%, 13% o 14%. En otras realizaciones más, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 15%, 16%, 17%, 18%, 19% o 20%.

La referencia a lo largo de esta memoria descriptiva a "una realización" o "una realización" o "un aspecto" significa que se incluye un rasgo, estructura o característica concretos descritos en relación con la realización en al menos una realización de la presente invención. De este modo, las entradas de las frases "en una realización" o "en una realización" en varios lugares a lo largo de esta memoria descriptiva no necesariamente se refieren todas a la misma realización. Además, los rasgos, estructuras o características concretos se pueden combinar de cualquier manera adecuada en una o más realizaciones.

## Ejemplos

### Ejemplo 1

### Caracterización de la clonalidad de células T y detección de ERM en muestras de T-ALL

En este ejemplo, se obtuvieron muestras de sangre por pares de leucemia linfoblástica aguda (T-ALL) pediátrica de células T de 43 pacientes (Children's Oncology Group AALL0434) en el momento del diagnóstico y el día 29 después de la terapia de inducción. La región determinante de complementariedad (CDR3) que codifica las regiones de los genes *TCRB* y *TCRG* se secuenciaron para los 86 especímenes mediante secuenciación de alto rendimiento (HTS) utilizando una plataforma Illumina GA2 (Illumina, San Diego, CA) como se describió previamente (véanse Robins et al., Blood, 114(19):4099-4107, 2009; y Robins et al., Sci Transl Med. 3(90):90ra61, 2011). Se utilizaron muestras de pretratamiento para identificar secuencias de ADN que codifican CDR3 de TCR únicas reordenadas para el clon leucémico, y las muestras post-tratamiento se evaluaron para determinar la frecuencia de cada secuencia de ADN que codifica CDR3 de TCR reordenada como un porcentaje del total. La frecuencia de cada secuencia de ADN que codifica CDR3 de TCR reordenada también se enumeró en muestras posteriores al tratamiento de todos los demás pacientes para evaluar la especificidad. Estos resultados se compararon con los resultados de ERM obtenidos mediante citometría de flujo de 9 colores, por protocolo de prueba.

Treinta y una de 43 muestras de pretratamiento (72,1%) tenían una población clonal de células T detectable basándose en análisis de secuencia de *TCRB*, y 27 de ellas también tenían una secuencia de *TCRG* clonal detectable. Cinco muestras exhibieron una secuencia que codificaba CDR3 de *TCRG* reordenada única adicional compatible con el reordenamiento de ambos loci de *TCRG* o con la presencia de dos subpoblaciones clonales. Cinco de 12 casos sin una secuencia del gen de *TCRB* clonal detectable tenían el inmunofenotipo de T-ALL del precursor tímico temprano (ETP), compatible con el hecho de que portaban genes *TCRB* y *TCRG* de línea germinal no reordenada. Ningún otro caso fue ETP. Seis de los 12 casos sin una secuencia del gen de *TCRB* clonal detectable tenían el inmunofenotipo de T-ALL próximo al del precursor tímico temprano (n-ETP). La muestra restante sin una secuencia del gen de *TCRB* clonal detectable, tenía una secuencia de gen de *TCRG* clonal detectable. Por el contrario, de los 31 casos en los que se identificó un reordenamiento de *TCRB* clonal en el momento del diagnóstico, ningún caso fue ETP, tres fueron próximos a ETP, y los 28 restantes fueron de inmunofenotipo T-ALL no ETP. En esta cohorte, hubo una asociación de la ausencia de una secuencia de *TCRB* clonal completa en la muestra de pretratamiento con inmunofenotipo ETP o próximo a ETP mediante mpFC ( $p < 0,0001$ , Prueba exacta de Fischer).

Las 37 muestras restantes con una secuencia de TCR clonal detectable antes del tratamiento se utilizaron para comparar mpFlow con HTS. La secuenciación de alto rendimiento utilizada para detectar las secuencias de *TCRB* clonales originales del paciente el día 29 revelaron tres subgrupos de pacientes en esta cohorte: 1) aquellos para los que no se detectó ERM mediante HTS o mpFC, 9 casos; 2) aquellos para los que se detectó ERM solo mediante HTS pero no mediante mpFC, 10 casos; y 3) aquellos en los que no se detectó ERM mediante HTS ni mpFC, 12 casos (Figura 1B). De los diez casos para los que HTS detectó ERM y mpFC no lo hizo, la ERM fue 10 a 100 veces menor que para los 12 casos para los que tanto mpFC como HTS detectaron ERM. Estos resultados sugieren que HTS ofrece una detección de ERM de sensibilidad superior. Las frecuencias de secuencia de fondo eran muy bajas ( $0-10^{-5}$ ) en otras muestras de pacientes postratamiento, siendo ligeramente más alta para *TCRG* que para *TCRB*, compatible con la diversidad de la secuencia de línea germinal. La fuerte asociación del estado de ETP con la carencia de una secuencia de TCR clonal detectable mediante HTS en el momento del diagnóstico también sugirió la utilidad de HTS de TCR para identificar este subconjunto de malos resultados de T-ALL.

### Materiales y métodos

**Preparación de la muestra.** Se recogió sangre (muestras de 5 ml) antes del tratamiento (Día 0) y 29 días después del tratamiento inicial (Día 29) de 43 individuos diagnosticados de T-ALL. La sangre recolectada se subdividió en una muestra para la secuenciación de alto rendimiento de TCR (HTS) y una muestra para el análisis de citometría de flujo de mpFC. Las muestras fueron codificadas a ciegas y enviadas para la HTS. Todos los pacientes dieron su consentimiento para el uso de sus muestras como parte de la prueba COG.

**Citometría de flujo.** Se realizó citometría de flujo multiparamétrica de 9 colores (mpFC) en la Universidad de Washington (Seattle, WA) como parte de la evaluación de rutina para ERM (17). Los datos de flujo se revisaron, y las muestras se clasificaron como T-ALL de precursor tímico temprano (ETP) o no ETP, como se describió previamente (15). Los casos en los que el inmunofenotipo fue similar pero no suficiente para cumplir los criterios de ETP se designaron como próximos a ETP (nETP). Los datos se analizaron utilizando el soporte lógico Woodlist versión 2.7 (B. Wood, Univ. of Washington, Seattle, WA). Se informa sobre los resultados como el porcentaje de linfoblastos de células T totales, células mononucleares totales y celularidad total para su comparación con estudios moleculares.

**Secuenciación de alto rendimiento (HTS)/Secuenciación de regiones CDR3:** las regiones de CDR3 de *TCRG* y *TCRB* se amplificaron y secuenciaron a partir de 400 ng de muestras del Día 0 y 1200 ng del Día 29, o toda la muestra restante de ADN extraído. La amplificación y secuenciación de las regiones CDR3 de *TCRβ* se llevaron a cabo como se describió previamente (9) y amplificación y secuenciación de las regiones CDR3 de *TCRG* se llevaron

como se describió previamente (11). Las secuencias para las regiones CDR3 de TCR $\beta$  y TCR $\gamma$  se delinearón de acuerdo con la definición establecida por la colaboración internacional ImMunoGeneTics (18). Las secuencias que no coincidían con las secuencias de CDR3 se eliminaron del análisis. Se utilizó un algoritmo convencional para identificar qué segmentos V, D y J contribuyeron a cada secuencia CDR3 de TCR $\beta$  y qué segmentos V y J contribuyeron a cada secuencia de CDR3 de TCR $\gamma$  (18). Las secuencias de CDR3 reordenadas se clasificaron como no productivas si se identificaron inserciones o deleciones que daban lugar a desplazamientos de marco o codones de parada prematuros.

Identificación de secuencias de CDR3: se aislaron diez ml de sangre de seis controles sanos. Se aislaron células mononucleares de sangre periférica (PBMC), se extrajo el ADN y se utilizaron 1200 ng de ADN extraído para amplificar las secuencias de TCRB utilizando el mismo protocolo que las muestras de T-ALL del día 29. La frecuencia de los clones reactivos de células T  $\alpha\beta$  más comunes en PBMC de estos individuos sanos promedió el 3% del repertorio total con una desviación típica de 3%. Debido a que las células T  $\alpha\beta$  portaban ambas secuencias codificantes de CDR3 de cadena TCRG y TCRB (11) reordenadas, se supuso que ambas frecuencias de las cadenas de CDR3 de TCRB y TCRG de mayor número de copias representaban los linfoblastos T neoplásicos. El día 0, las muestras con una frecuencia diez veces mayor de las secuencias de CDR3 más comunes en comparación con los controles, >30%, se consideraron población clonal. Para las muestras en las que las dos secuencias de TCR más comunes eran de frecuencia comparable, se consideró que ambas secuencias eran el clon de cáncer probable. La ERM a los 29 días del tratamiento se escrutó buscando secuencias de CDR3 que coincidieran de forma idéntica con la secuencia clonal identificada el día 0, requiriendo un emparejamiento de 60 pares de bases. Se observaron tanto la presencia como la frecuencia de ERM con respecto a el repertorio total de TCR. Para determinar si la secuencia clonal era específica, todas las muestras del día 29 se escrutaron para determinar la presencia y la frecuencia de todas las secuencias clonales de CDR3 de TCR identificadas para determinar la especificidad cruzada del paciente.

## Resultados

Utilizando 43 pares de muestras de T-ALL emparejadas derivadas de pacientes inscritos en la prueba AALL0434 del Children's Oncology Group (Universidad de Washington, Seattle, WA), se secuenciaron las regiones determinantes de complementariedad (CDR3) de TCRB (Tabla 5) y TCRG (Tablas 5 y 6) mediante HTS utilizando aproximadamente 75.000 genomas el día 0 y 150.000 genomas el día 29. Estas secuencias del día 0 permitieron la definición, para cada paciente, de las secuencias únicas del gen de TCR recombinado que representan los linfoblastos T neoplásicos clonales de cada paciente (Tablas 5 y 6). Los datos normalizados de mpFC (blast/células CD7+) también se muestran en las Tablas 5 y 6.

Tabla 5. % de células T que eran Blast el día 0 y el día 29: caracterización mediante mpFC y HTS de TCRB

Paciente	Día 0 Blast/Células T mpFC	Día 0 HTS de TCRB	Día 29 Blast/Células T mediante mpFC	Día 29 HTS de TCRB
PT_44	0,97672065	0,99434804	0	0
PT_63	0,971875	0,99433591	0	0
PT_48	0,98097252	0,98802835	0	0
PT 9	0,97037794	0,98506196	0	0
PT 7	0,9726997	0,98471692	0	0
PT_15	0,97247706	0,96641846	0	0
PT_59	0,92966361	0,93608697	0	0
PT_36	0,63565891	0,77634724	0	0
PT_33	0,36514523	0,38978913	0	0
PT_53	0,96843177	0,98636256	0	4,6305E-06
PT_52	0,87757202	0,97285037	0	5,0976E-06
PT_5	0,88681319	0,80910021	0	1,4101E-05
PT_50	0,98088531	0,99470926	0	3,4997E-05
PT_1	0,93128964	0,86986261	0	4,4024E-05
PT_32	0,87001287	0,89047715	0	0,00013614
PT 8	0,97985901	0,63830248	0	0,0004481



ES 2 683 037 T3

Paciente	Día 0 Blast/Células T mpFC	Día 0 HTS de TCRB	Día 29 Blast/Células T mediante mpFC	Día 29 HTS de TCRB
PT_30	0,67403315	0,48984069	0	0,00048091
PT_49	0,94558522	0,93357269	0	0,00070619
PT_66	0,71351351	0,86920201	0	0,00348305
PT_39	1	0,98803323	0,0002907	0,00201475
PT_65	0,9563795	0,84025047	0,00114679	7,6499E-05
PT_41	0,90253807	0,92867626	0,00152838	0,00292659
PT_61	0,93349456	0,62550401	0,00653266	0,02535341
PT_51	0,95571576	0,94895541	0,00797546	0,0047952
PT 6	0,98189135	0,95963063	0,01690647	0,01741715
PT_47	0,83670295	0,58431047	0,02276786	0,03573035
PT_4	0,99295775	0,96190309	0,04795737	0,1614654
PT_58	0,68115942	0,58647625	0,05369128	0,01921525
PT_2	0,81477516	0,74794199	0,08156607	0,02893032
PT_40	0,5815508	0,70912279	0,0990566	0,2422609
PT_14	0,91898148	0,88730182	0,14479026	0,02124786
PT_55	0,98183653	0,12006615	0,43255814	0
PT_57	0,95523906	0,02088572	0,6459144	0
PT_3	0,98694779	0,01755708	0,00202703	0
PT_12	0,89263804	0,01493205	0,05456989	0
PT_54	0,97082495	0,09110046	0,069869	0,28020422
PT_45	0,99296482	0,07754443	0,2345679	0
PT_60	0,90118153	0,03709322	0,05792683	0
PT_68	0,9874477	0,02719658	0,00849858	0,0424174
PT_56	0,86721144	0,01765149	0,1245283	0,0053763
PT_13	0,95947426	0,06090199	0,76954315	0
PT_42	0,91920252	0,08504223	0,0019544	0,01346556
PT_11	0,96177686	0,00694597	0	0
		AVG ETP	0,18478833	
		AVG otro	0,01613064	
		AVG All	0,06712017	

Las muestras PT\_13, 39, 42, 45, 56, 60 y 68 se clasificaron como ETP. Las muestras PT\_3, 12, 54, 55 y 57 se clasificaron como nETP (Figuras 2 y 3).

5 Tabla 6. % de células T que fueron Blast el día 0 y caracterización del día 29 mediante mpFC y HTS de TCRB/TCRG

Muestra	Día 0			Día 29		
	mpFC	HTS		mpFC	HTS	
	Blast/Células T	TCRB	TCRG	Blast/Células T	TCRB	TCRG
1	0,931289641	0,869862606	52,5254048	0	4,4024E-05	0,03937815
2	0,814775161	0,747941989	82,3288372	0,08156607	0,02893032	4,63888072

ES 2 683 037 T3

Muestra	Día 0			Día 29		
	mpFC	HTS		mpFC	HTS	
	Blast/Células T	TCRB	TCRG	Blast/Células T	TCRB	TCRG
3	0,986947791	0,017557079	12,3924143	0,00202703	0	
4	0,992957746	0,961903091	54,7257337	0,04795737	0,1614654	7,45339935
5	0,886813187	0,809100213	41,6389335	0	1,4101E-05	0,01490627
6	0,981891348	0,95963063	58,557176	0	0,01420907	1,25544372
7	0,972699697	0,984716924	0,59258927	0	0	0
8	0,979859013	0,638302477	70,6051915	0	0,0004481	0
9	0,970377937	0,98506196	50	0	0	0
11	0,96177686	0,006945971	94,6793227	0	0	19,9611535
12	0,892638037	0,014932055	4,16576441	0,05456989	0	
13	0,959474261	0,050950488	0,05179646	0,76954315	0	
14	0,918981481	0,88730182	84,245218	0,14479026	0,0154863	7,92210056
15	0,972477064	0,966418465	80,8011935	0	0	0,04827268
30	0,67403315	0,489840689	17,1898345		0,00048091	
32	0,87001287	0,863124126	64,0647828		0,00013614	0,08466056
33	0,36514523	0,389789133	6,58138203		0	
36	0,63565891	0,776347236	55,128848		0	
39	1	0,98803323	78,3268267		0,00201475	7,67E-04
40	0,5815508	0,515729332	35,0639536	0,0990566	0,09248403	
41	0,90253807	0,928676259	59,2559701	0,00152838	0,00292659	0,1406089
42	0,91920252	0,08504223	ND	0,0019544	0,01346556	
44	0,97672065	0,994348041	96,9885863	0	0	
45	0,99296482	0,077544426	10,8482546	0,2345679	0	
47	0,83670295	0,58431047	27,5213337	0,02276786	0,03573035	
48	0,98097252	0,988028346	12,2479173	0	0	
49	0,94558522	0,933572687	72,0934554	0	0,00070619	
50	0,98088531	0,994709258			3,4997E-05	
51	0,95571576	0,948955408	61,973728	0,00797546	0,0047952	
52	0,87757202	0,97285037	13,5297526		5,0976E-06	
53	0,96843177	0,986362564	8,64156011		4,6305E-06	
54	0,97082495	0,107187084	71,5454996	0,069869	0,08321639	
55	0,98183653	0,120066146	4,8143757	0,43255814	0	
56	0,86721144	0,017651493	33,7748913	0,1245283	0,00564494	
57	0,95523906	0,020885717	71,2907132	0,6459144	0,00838002	0,16145989
58	0,68115942	0,586476251	14,9081924	0,05369128	0,03701839	3,40390897
59	0,92966361	0,936086965	ND	0	0	
60	0,90118153	0,037093219	ND	0,05792683	0	
61	0,93349456	0,625504008	56,3484994	0,00653266	0,02535341	0,03928545

Muestra	Día 0			Día 29		
	mpFC	HTS		mpFC	HTS	
	Blast/Células T	TCRB	TCRG	Blast/Células T	TCRB	TCRG
63	0,971875	0,994335907	76,362186		0	
65	0,9563795	0,840250471	92,723888	0,00114679	7,6499E-05	0,03433687
66	0,71351351	0,869202008	11,7433739		0,00276186	
68	0,9874477	0,024076739	ND	0,00849858	0,0424174	

ND; no determinado.

En general, cada una de estas secuencias clonales representaba al menos el 15% del repertorio total de células T del sujeto en el momento del diagnóstico, una proporción que era mayor que tres desviaciones típicas de la media de los clones de células T reactivas en individuos normales (Tabla 7 y Figura 5). El día 0, se identificó fácilmente una población clonal de *TCRB* en 31 de 43 casos (Figura 1A, Figura 4). Por el contrario, 12 muestras de pacientes no tenían un secuencia de *TCRB* clonal identificable en el momento del diagnóstico, definida como la que comprendía menos de 15% de las secuencias totales. De los 12 pacientes en los que no se identificó secuencia de *TCRB* clonal, cuatro tenían un reordenamiento de *TCRG* clonal en el análisis concurrente (Tabla 6).

Tabla 7. Frecuencia de aparición de clones de mayor frecuencia en sujetos de control normal

Control sano	Clon más frecuente
AS	2,1975
BC	0,6709
CD	1,6264
HR	7,7823
JA	1,6008
RL	8,7205
Promedio	3,7664
SD	3,520896205
Promedio + 3DT	14,32908862

Con el conocimiento de las secuencias de CDR específicas de la población clonal del paciente desde el día 0, se examinó a continuación la capacidad de HTS para identificar la secuencia clonal el día 29, y estos hallazgos se compararon con los resultados de mpFC que se realizaron como parte del protocolo actual en la Universidad de Washington en la prueba AALL0434 del Children's Oncology Group. A la luz de las secuencias de *TCRB* y *TCRG* clonales originales del paciente en el momento del diagnóstico, la secuenciación de alto rendimiento de estas muestras de pacientes el día 29 reveló tres subgrupos de pacientes en esta cohorte: 1) aquellos para quienes no se detectó ERM mediante HTS o mpFC, 9 casos; 2) aquellos para quienes se detectó ERM tanto mediante HTS como mediante mpFC, 12 casos; y 3) un subconjunto para el que la ERM era detectable mediante HTS, pero no mediante mpFC, 10 casos (Figura 1B). En general, hubo un aumento de aproximadamente 100 veces en la sensibilidad con la que la ERM podría detectarse mediante la metodología HTS descrita en la presente memoria, en comparación con mpFC.

De los casos para los que no se identificó población de *TCRB* clonal el día 0, la evaluación posterior mostró que estos pacientes generalmente tenían niveles altos de ERM el día 29. La revisión del inmunofenotipo de los linfoblastos T de los pacientes el día 0 mostró que de los 12 casos, 5 tenían un inmunofenotipo compatible con un subtipo precursor tímico temprano (ETP), mientras que 7 tenían un inmunofenotipo próximo a ETP, en donde se cumplieron algunos, pero no todos, los criterios inmunofenotípicos para ETP (Tabla 5) (15). La asociación de enfermedad residual mínima más alta en pacientes con linfoblastos de inmunofenotipo ETP ( $x = 0,1848$ ) versus T-ALL no ETP (media = 0,01613, estadística de Fisher) fue compatible con informes previos que identificaron este subgrupo por tener un curso clínico más agresivo con aumento de la propensión a la recaída de la enfermedad. Basándose en estos datos, el grupo próximo a ETP apareció en sus características para ocupar una posición a lo largo de un continuo entre los subtipos no ETP y ETP de T-ALL.

Para evaluar la especificidad de HTS para la evaluación de ERM, se realizó una evaluación de la frecuencia de

aparición de una secuencia clonal específica de TCR de un paciente que está presente en las otras 41 muestras de pacientes el día 29. Este análisis demostró 15 "falsos positivos" de 1.722 comparaciones de genes *TCRG* ( $= N^2 - N = 42^2 - 42 = 1.722$ ) para *TCRG* y 3 "falsos positivos" de 1.722 comparaciones para *TCRB*. Cabe destacar que el nivel de estos clones falsos positivos fue mucho menor, en un orden de magnitud, en comparación con los verdaderos positivos de ERM.

## Discusión

Datos recientes del estudio AIEOP-BFM-ALL 2000, una prueba clínica prospectiva multiinstitucional que involucró a 464 pacientes con T-ALL confirmó la importancia de la enfermedad residual mínima (ERM) para categorizar la enfermedad normal, intermedia y de alto riesgo basándose en ERM en dos puntos temporales después de la terapia (2). Pacientes con ERM detectable de más de  $10^3$  células en el segundo punto temporal, TP2 (día 78) no tuvieron un buen resultado, de modo que se concluyó que la ERM persistente en TP2 era un factor predictivo importante para la recaída de la T-ALL. La detección de este nivel de enfermedad se encontraba fácilmente dentro de la sensibilidad de la presente tecnología HTS como se muestra en la presente memoria, por lo que la determinación de la clonalidad se logró sin los requisitos laboriosos de PCR individualizada o diseño de oligonucleótidos específico de alelos de enfoque anteriores para ERM. La HTS era fácilmente aplicable a escala, menos compleja desde el punto de vista organizativo que las metodologías anteriores, y se adaptaba fácilmente a cualquier entorno de laboratorio clínico.

En el presente ejemplo, la secuenciación de *TCRG* fue adecuada para detectar la clonalidad en una proporción ligeramente mayor de casos en comparación con *TCRB*, pero *TCRB* también se consideró de gran utilidad para la evaluación de la ERM debido a la mayor diversidad de líneas germinales de *TCRB* con respecto a *TCRG*. En análisis de muestras del día 29, se encontró que las secuencias de *TCR* clonales eran altamente específicas con la identificación coincidente de secuencias cruzadas de pacientes, es decir, la identificación de la secuencia clonal de un paciente en los otros 41 pacientes secuenciados en este estudio, siendo generalmente poco frecuentes. Los protocolos moleculares previamente descritos para la detección de ERM utilizando cebadores personalizados específicos para el paciente (7, 8) no evaluaron de forma rutinaria la especificidad cruzada del paciente, puesto que la evaluación de la sensibilidad (p.ej., % de los sujetos correctamente identificados como ERM+ según los criterios dados) y la falta de especificidad (p.ej., % de los sujetos identificados correctamente como ERM- según los criterios dados) fue el único requisito para los ensayos previos. Como tales, los presentes métodos HTS también permitieron la identificación de un solapamiento raro y coincidente de las secuencias del gen de TCR de un paciente a otro. Notablemente, el nivel de estas secuencias compartidas detectables en tales comparaciones falsas positivas entre pacientes fue mucho más bajo que el observado en casos reales de pacientes con ERM. Es importante destacar que, para ERM de T-ALL, los reordenamientos del gen de TCR parecían ser estables, con evidencia limitada (menos del 5%) para la evolución clonal.

La metodología HTS actualmente descrita contribuyó al pronóstico de la enfermedad en el momento del diagnóstico para T-ALL para un subconjunto de casos, al identificar muestras en las que no había reordenamiento de genes *TCRB* en el momento del diagnóstico, a pesar de la evidencia de una alta proporción de linfoblastos T mediante mpFC. Estos casos tendían a tener un inmunofenotipo similar al precursor tímico temprano, o características inmunofenotípicas similares que se describieron en la presente memoria como cercana a ETP (15). Se ha propuesto que un subtipo de ETP de T-ALL se deriva de un subconjunto inmunogenéticamente temprano de células T antes del reordenamiento génico del *TCR*, aunque un estudio reciente sugiere que dicho subconjunto también puede haber representado la reversión del estado doblemente positivo de células T inmaduras a un inmunofenotipo más inmaduro (16). La comparación con los resultados de citometría de flujo de ERM el día 29 mostró que los 12 casos, para los cuales no se identificó secuencia *TCRB* clonal en el pretratamiento, tenía un aumento de ERM al día 29, lo que concuerda con un curso clínico más agresivo de estos tumores de tipo ETP.

La utilidad de HTS para la secuenciación de los genes que codifican el receptor inmunitario adaptativo linfoide en T-ALL es extensible al seguimiento clínico rutinario mediante HTS de los genes que codifican el receptor inmunitario adaptativo linfoide en leucemia/linfomas linfoblásticos de células B (B-ALL). La determinación de ERM en B-ALL, mediante técnicas moleculares y mpFC, ha demostrado ser importante para el pronóstico del paciente (3). La B-ALL comúnmente muestra reordenamientos de linaje cruzado (B a T) de *TCRB* y/o *TCRG*, por lo que el enfoque descrito en la presente memoria es directamente aplicable a un subconjunto de casos de B-ALL (1) De manera similar, la HTS de reordenamientos de genes de cadena pesada de inmunoglobulina también se puede realizar fácilmente basándose en las enseñanzas proporcionadas en la presente memoria.

## Referencias

1. Flohr et al., *Leukemia* 22, 771-782 (2008).
2. Schrappe et al., *Blood* 118, 2077-2084).
3. van Dongen et al., *Lancet* 352, 1731-1738 (1998).

4. D. Campana, Hematology Am Soc Hematol Educ Program 2010, 7-12).
5. Campana, Curr Hematol Malig Rep 5, 169-176 (2010).
6. Roshal et al., Cytometry B Clin Cytom 78, 139-146).
7. van der Velden et al., Leukemia 21, 706-713 (2007).
- 5 8. van der Velden et al., Leukemia 21, 604-611 (2007).
9. Robins et al., Blood 114, 4099-4107 (2009).
10. Robins et al., Sci Transl Med 2, 47ra64 (2010).
11. Sherwood et al., Sci Transl Med 3, 90ra61 (2011).
12. Kalos et al., Sci Transl Med 3, 95ra73 (2011).
- 10 13. Boyd et al., Sci Transl Med 1, 12ra23 (2009).
14. Robins, Desmarais, Matthis, Livingston, Andriesen, Reijonen, Carlson, Nepom, Yee, Cerosaletti, Ultra-sensitive detection of rare T cell clones. J Immunol Methods doi:10.1016/j.jim.2011.09.001 (10 de sept. de 2011, publicación electrónica antes de impresión PMID 21945395).
15. Coustan-Smith et al., Lancet Oncol 10, 147-156 (2009).
16. Berquam-Vrieze et al., Blood 118, 4646-4656).
17. Wood et al., Arch Pathol Lab Med 130, 680-690 (2006).
18. Yousfi Monod et al., Bioinformatics 20 Supl 1, i379-385 (2004).

## Ejemplo 2

### 20 **Uso de la composición molde para determinar el factor de amplificación**

Este ejemplo describe la cuantificación de moléculas de ADN reordenado que codifican una pluralidad de moléculas de IG, utilizando la composición de oligonucleótido molde actualmente descrita como molde sintético "incorporado" a una amplificación por PCR multiplexada de una muestra de ADN que contiene células B y ADN de fibroblastos.

ADN molde biológico: Se utilizaron como fuentes de ADN molde ocho muestras biológicas, con cada muestra biológica que contenía la misma cantidad de ADN genómico total (ADNg), 300 ng, pero en una proporción diferente de (i) ADN extraído de células B con respecto a (ii) ADN extraído de células de fibroblastos humanos, un tipo de célula en el que los genes que codifican IG y TCR no se reordenan. Las muestras contenían 0, 0,07, 0,3, 1, 4, 18, 75 o 300 ng de ADNc de células B, suministrando el ADNg de fibroblastos el resto de cada preparación de ADNg de 300 ng. Se realizaron cuatro réplicas de cada muestra.

ADN de molde sintético: A cada reacción de PCR (más abajo) se le añadieron 5000 moléculas (4-5 moléculas de cada secuencia) de una composición molde de oligonucleótido que comprendía un conjunto de 1116 moléculas de oligonucleótido molde de IGH sintéticas (SEC ID NO: 4085-5200). (También se describe una composición molde de IGH que comprende un conjunto de 1116 oligonucleótidos molde en la Lista de Secuencias como SEQ ID NO: 1805-2920).

Reacción de PCR: La reacción de PCR utilizó la mezcla maestra de PCR QIAGEN múltiplex Plus™ (número de componente QIAGEN 206152, Qiagen, Valencia, CA), solución Q al 10% (QIAGEN) y 300 ng de ADN molde biológico (descrito anteriormente). Los cebadores de amplificación agrupados se añadieron de modo que la reacción final tenía una concentración agregada de cebador directo 2 µM y una concentración agregada de cebador inverso 2 µM. Los cebadores directos (SEQ ID NO: 5201-5286) incluían 86 cebadores que tenían en el extremo 3' un segmento de aproximadamente 20 pb que se reasociaba con la secuencia que codifica el segmento V de IGH y en el extremo 5' un cebador universal de aproximadamente 20 pb pGEXf. Los cebadores inversos (SEQ ID NO: 5287-5293) incluían un conjunto de cebadores específicos del segmento J que en el extremo 3' tenían un segmento de aproximadamente 20 pb que se reasociaba con la secuencia que codificaba el segmento J de IGH y en el extremo 5' de los cebadores de J había un cebador universal pGEXr. Las siguientes condiciones de ciclos térmicos se utilizaron en un termociclador C100 (Bio-Rad Laboratories, Hercules, CA, EE.UU.): un ciclo a 95°C durante 10 minutos, 30 ciclos a 94°C durante 30 segundos, 63°C durante 30 segundos, y 72°C durante un minuto, seguido de un ciclo a 72°C durante 10 minutos. Cada reacción se ejecutó por cuadruplicado.

Para la secuenciación, se incorporaron los adaptadores Illumina (Illumina Inc., San Diego, CA), que también incluían una etiqueta de 8 pb y un conjunto aleatorio de 6 pb de nucleótidos, en los extremos de los productos de reacción de

5 PCR en una reacción de PCR de 7 ciclos. Los reactivos y condiciones de la PCR fueron los descritos anteriormente, excepto por las condiciones de termociclación, que fueron: 95°C durante 5 minutos, seguidos de 7 ciclos de 95° durante 30 segundos, 68° durante 90 segundos y 72° durante 30 segundos. Después de termociclación, las reacciones se mantuvieron durante 10 minutos a 72° y los cebadores fueron los cebadores de cola del adaptador Illumina (SEQ ID NO: 5387 - 5578). Las muestras se secuenciaron en un secuenciador Illumina MiSEQ™ utilizando el cebador Illumina\_PE\_RD2.

10 Resultados. Se obtuvieron datos de secuencia para cada muestra y los productos de amplificación de moldes sintéticos se identificaron por la presencia de la secuencia de oligonucleótido código de barras. Para cada muestra, el número de productos molde se dividió por la cantidad de secuencias oligonucleotídicas molde sintéticas únicas (1116) para llegar a un factor de amplificación de la muestra. El número total de productos de amplificación de los moldes biológicos para cada muestra se dividió a continuación por el factor de amplificación para calcular la cantidad de moléculas moldeadas del molde biológico (p.ej., recombinaciones VDJ) en la reacción de amplificación inicial como una estimación de la cantidad de moldes de genoma de células B únicos. Los valores promedio con las desviaciones típicas se representaron gráficamente frente al número conocido de moléculas molde biológicas reordenadas basándose en la entrada de células B (Figura 8). En la Figura 8, los puntos representan el factor de amplificación promedio y las barras representan la desviación típica a través de las cuatro repeticiones. El uso de factores de amplificación calculados como se describe en la presente memoria para estimar la cantidad de moléculas que codifican IG con VJ reordenadas (como un valor proxy para el número de células B) arrojó 15 determinaciones que eran compatibles con las cantidades de células B conocidas al menos hasta una entrada de 20 100 células B. Los valores estimados del factor de amplificación y el factor de amplificación observado estaban altamente correlacionados (Fig. 8,  $R^2 = 0,9988$ ).

## REIVINDICACIONES

1. Un método para detectar enfermedad residual mínima para un tumor maligno hematológico linfoide en un sujeto después de la terapia, comprendiendo el método

- 5 (a) amplificación de ADN extraído de una primera muestra en una reacción en cadena de polimerasa múltiple (PCR) utilizando cebadores del segmentos V y J para producir una multiplicidad de moléculas de ADN reordenado amplificado que comprenden cada una una región que codifica CDR3 de TCR, en donde la primera muestra se ha obtenido del sujeto antes de la terapia y comprende una pluralidad de células hematopoyéticas linfoides,
- 10 (b) secuenciación de alto rendimiento de dicha multiplicidad de moléculas de ADN reordenado amplificado para determinar las frecuencias relativas de regiones que codifican CDR3 de TCR reordenadas de manera única en la primera muestra, identificando así la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas,
- 15 (c) amplificación de ADN extraído de una segunda muestra en una reacción en cadena de polimerasa múltiple (PCR) utilizando los cebadores del segmento V y J para producir una multiplicidad de moléculas de ADN reordenado amplificado comprendiendo cada una una región que codifica CDR3 de TCR, en donde la segunda muestra se ha obtenido del sujeto después de la terapia y comprende una pluralidad de células hematopoyéticas linfoides, y
- 20 (d) secuenciación de alto rendimiento de dicha multiplicidad de moléculas de ADN reordenado amplificado producidas a partir de la segunda muestra para determinar la frecuencia relativa de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas en la segunda muestra, en donde la presencia de la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas a una frecuencia relativa de al menos una frecuencia de uno de cada  $10^6$  a uno de cada  $10^5$  regiones que codifican CDR3 de TCR indica enfermedad residual mínima para el tumor maligno hematológico linfoide;

en donde la PCR múltiple comprende adicionalmente una pluralidad de oligonucleótidos molde que tiene una pluralidad de secuencias de oligonucleótidos de fórmula general

30  $5'-U1-B1-V-B2-R-B3-J-B4-U2-3'$  [I]

en donde

- 35 (i) V es un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180 o 210, y no más de 1000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos V comprende una secuencia oligonucleotídica única,
- 40 (ii) J es un polinucleótido que comprende al menos 15-30, 31-60, 61-90, 91-120 o 120-150, y no más de 600, 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica que codifica la región de empalme (J) del receptor inmunitario adaptativo, o el complemento de la misma, y en cada una de la pluralidad de secuencias de oligonucleótidos J comprende una secuencia oligonucleotídica única,
- 45 (iii) U1 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (1) una primera secuencia de oligonucleótido adaptador universal, y (2) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a la primera secuencia de nucleótido adaptador universal,
- (iv) U2 no es nada o comprende un oligonucleótido que tiene una secuencia que se selecciona entre (1) una segunda secuencia de oligonucleótido adaptador universal, y (2) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está conectada y posicionada 5' con respecto a una
- 50 segunda secuencia de oligonucleótido adaptador universal,
- (v) B1, B2, B3 y B4 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia código de barras de oligonucleótido de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25, nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias de oligonucleótidos B comprende una secuencia oligonucleotídica única que identifica de forma
- 55 única, como una combinación pareada, (1) la secuencia oligonucleotídica V única de (i) y (2) la secuencia oligonucleotídica J única de (ii), y
- (vi) R no es nada o comprende un sitio de reconocimiento para enzimas de restricción que comprende una secuencia oligonucleotídica que está ausente de (i) - (v).

60 2. El método de la reivindicación 1, en donde la región que codifica CDR3 de TCR reordenada de manera única de los uno o más clones de células T malignas se identifica por su presencia en la primera muestra a una frecuencia relativa de al menos 15% de las regiones que codifican CDR3 de TCR reordenadas.

3. El método de la reivindicación 1 en donde la PCR múltiple comprende:

(i) una pluralidad de cebadores oligonucleotídicos del segmento V que son capaces cada uno independientemente de hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región V del receptor de células T (TCR) humano, en donde cada cebador del segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica V $\gamma$  de TCR o que codifica V $\beta$  de TCR y en donde la pluralidad de cebadores del segmento V hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican V $\gamma$  de TCR o que codifican V $\beta$  de TCR que están presentes en la muestra; y

(ii) una pluralidad de cebadores oligonucleotídicos del segmento J que son capaces cada uno independientemente hibridar específicamente con al menos un polinucleótido que codifica un polipéptido de la región J del receptor de células T (TCR) humano, en donde cada cebador del segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico funcional que codifica J $\gamma$  de TCR o que codifica J $\beta$  de TCR y en donde la pluralidad de cebadores del segmento J hibrida específicamente sustancialmente con todos los segmentos génicos funcionales que codifican J $\gamma$  de TCR o que codifican J $\beta$  de TCR que están presentes en la muestra.

4. El método de la reivindicación 3, en donde los cebadores del segmento V y del segmento J son capaces de promover la amplificación en dicha reacción en cadena de la polimerasa múltiple (PCR) de sustancialmente todas las regiones que codifican CDR3 de TCR $\gamma$  reordenadas o sustancialmente todas las regiones que codifican CDR3 de TCR $\beta$  en la muestra para producir una multiplicidad de moléculas de ADN reordenado amplificado de una población de células T en la muestra, siendo suficiente dicha multiplicidad de moléculas de ADN reordenado amplificado para cuantificar la diversidad de la región que codifica CDR3 de TCR $\gamma$  o CDR3 de TCR $\beta$  en la población de células T, y en donde cada molécula de ADN reordenado amplificado en la multiplicidad de moléculas de ADN reordenado amplificado tiene menos de 600 nucleótidos de longitud.

5. El método de la reivindicación 3 o 4:

(a) en donde cada segmento génico funcional que codifica V de TCR comprende una secuencia señal de recombinación (RSS) del gen V y cada segmento génico funcional que codifica J de TCR comprende una RSS del gen J, y en donde cada molécula de ADN reordenado amplificado comprende (i) al menos 10, 20, 30 o 40 nucleótidos contiguos de una hebra efectora del segmento génico que codifica V de TCR, estando situados dichos al menos 10, 20, 30 o 40 nucleótidos contiguos 5' respecto a la RSS del gen V y (ii) al menos 10, 20 o 30 nucleótidos contiguos de una hebra efectora del segmento génico que codifica J de TCR, estando situados dichos al menos 10, 20 o 30 nucleótidos contiguos 3' con respecto a la RSS del gen J;

(b) la pluralidad de cebadores oligonucleotídicos del segmento V y la pluralidad de cebadores oligonucleotídicos del segmento J comprenden al menos una de (1) las secuencias expuestas en SEQ ID NO: 5579-5643, (2) las secuencias expuestas en SEQ. ID NOs: 5644-5792, (3) las secuencias expuestas en SEQ ID NO: 5793-5816, (4) las secuencias expuestas en SEQ ID NO: 5817-6123, (5) las secuencias expuestas en SEQ ID NO : 6124-6127 y 6212-6215, (6) las secuencias expuestas en SEQ ID NO: 6128-621 1 y 6216- 6221, (7) las secuencias expuestas en SEQ ID NO: 6222-6286 y (8) las secuencias expuestas en SEQ ID NO: 6222-6351;

(c) uno o ambos de:

(i) los cebadores oligonucleotídicos del segmento V comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO: 5579-5630, 5644-5779, 5799-5816, 6124-6127 y 6222-6273, y

(ii) los cebadores del segmento J comprenden uno o una pluralidad de oligonucleótidos que exhiben al menos 90% de identidad de secuencia con una o más de las secuencias de nucleótidos expuestas en SEQ ID NO: 5631-5643, 5780-5798, 6212-6215, y 6274-6286.

6. El método de la reivindicación 1, en donde al menos uno de:

(i) la pluralidad de oligonucleótidos molde comprende al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar específicamente cada cebador oligonucleotídico del segmento V y al menos un oligonucleótido molde que tiene una secuencia oligonucleotídica de fórmula general (I) con la que puede hibridar específicamente cada cebador oligonucleotídico del segmento J; o

(ii) la pluralidad de oligonucleótidos molde comprende al menos a o al menos b secuencias de oligonucleótidos únicas, la que sea mayor, donde a es la cantidad de segmentos génicos que codifican la región V del receptor inmunitario adaptativo único en el sujeto y b es la cantidad de segmentos génicos que codifican la región J del receptor inmunitario adaptativo único en el sujeto, y la composición comprende al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único.



7. El método de la reivindicación 1, que comprende adicionalmente cuantificar moléculas de ADN reordenado que codifican uno o una pluralidad de receptores inmunitarios adaptativos en el ADN extraído de la muestra biológica mediante las etapas de:

- 5 (i) secuenciar cuantitativamente toda o una porción suficiente de cada una de dichas moléculas de ADN molde amplificado y cada una de dichas moléculas de ADN reordenado amplificado para cuantificar (a) una cantidad de producto molde de moléculas de ADN molde amplificado que contienen al menos una secuencia código de barras de oligonucleótido, y (b) una cantidad de producto reordenado de moléculas de ADN reordenado amplificado que carecen de una secuencia código de barras de oligonucleótido;
- 10 (ii) calcular un factor de amplificación dividiendo la cantidad de producto molde de (i)(a) por la cantidad conocida de cada uno de la pluralidad de oligonucleótidos molde que tienen una secuencia oligonucleotídica única; y
- (iii) dividir la cantidad de producto reordenado de (i)(b) por el factor de amplificación calculado en (ii) para cuantificar las moléculas de ADN que codifican el receptor inmunitario adaptativo único en la muestra.

15 8. El método de una cualquiera de las reivindicaciones precedentes, en donde dicho tumor maligno hematológico linfoide se selecciona entre: leucemia linfoblástica de células T aguda (T-ALL), leucemia linfoblástica de células  $\beta$  aguda (B-ALL), mieloma múltiple, plasmacitoma, macroglobulinemia, leucemia linfocítica crónica (CLL), leucemia linfoblástica aguda (ALL), linfoma de Hodgkin, linfoma no Hodgkin, linfoma cutáneo de células T, linfoma de células del manto, linfoma periférico de células T, leucemia de células pilosas, linfoma prolinfocítico de células T, linfoma angioinmunoblástico de células T, leucemia/linfoma de células T linfoblástico, leucemia/linfoma de células T adultas, micosis fungoide, síndrome de Sezary, neoplasia mieloproliferativa y síndrome mielodisplásico.

20

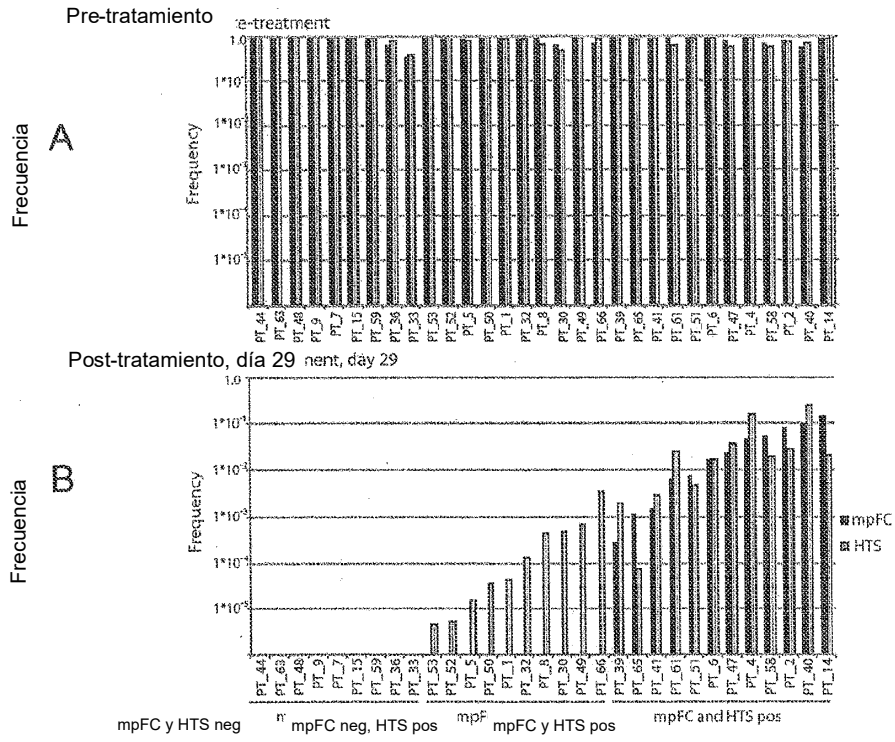


Fig. 1

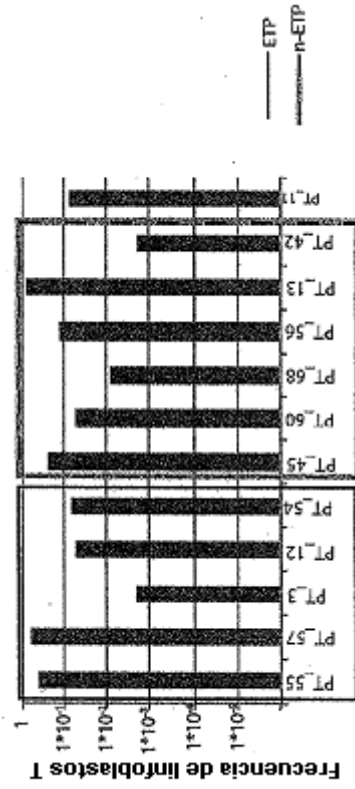


Fig. 2

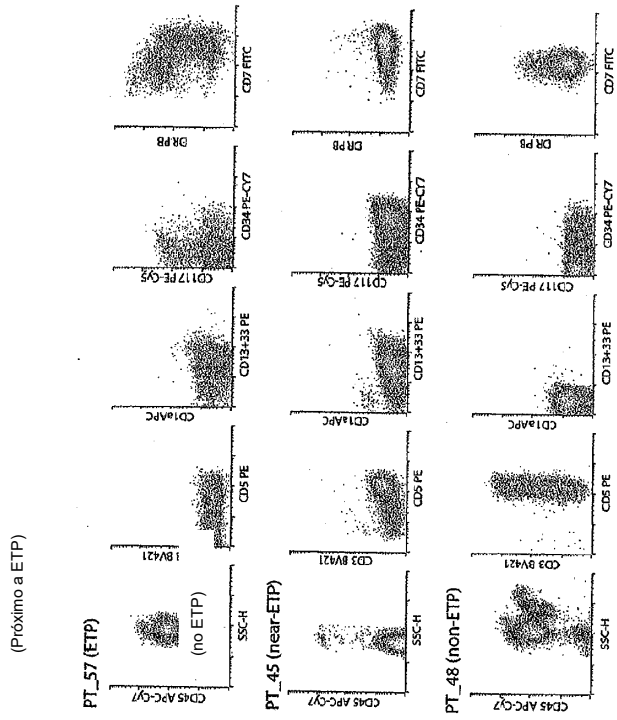


Fig. 3

ID	Secuencia CDR3 de TCRB	%	Gen V	Gen J	SEQ ID NO:
PT 50	TGTGCCAGCAGTTACTC <u>CGCCGA</u> AAAATGAAAACT	99,5	6	TRBJ1-4	774
PT 44	TGTGCCAGCAGCC <u>CCGGTCGCAG</u> CCAAGTGTCTGGGGCCAACGTCC T	99,4	11	TRBJ2-6	775
PT 63	TGTGCCAGCAGTTACTC <u>CGCCTC</u> CGAACCTAATGAGCA	99,4	6	TRBJ2-1	776
PT 39	TGTGCCAGCAGCTTAAGCGGGTA CCTTAAAGAGACCCA	98,9	TREB7-9	TRBJ2-5	777
PT 48	TGCGCCAGCAC <u>CCCCGAGAGGAG</u> AGCGGGAGCCAGATACGCA	98,8	TREB5-1	TRBJ2-3	778
PT 53	TGCAGTGCAGAG <u>GGCCTATGTT</u> CCGGGACCGATCCTCCTACGAGC A	98,6	TREB20-1	TRBJ2-7	779
PT 7	TGTGCCAGCAGCCAAGAT <u>TTTACA</u> GGTGCAGATTCTGGGGCCAACGT CCT	98,5	TREB4-2	TRBJ2-6	780
PT 8	TGTGCCAGCAGTGAAT <u>TGAAAGA</u> CGAGAAGGGTGGTTTTTTTGCCA CTGAAGC	98,5	2	TRBJ1-1	781
PT 9	TGCGCCAGCAGCCAAGATAGAGA GCGGGAGGATCGCACAGATACGC A	98,5	TREB4-1	TRBJ2-3	782
PT 52	TGTCCAGT <u>CATT</u> ACCCA	97,3	TREB3-1	TRBJ2-5	783
PT 4	TGTGCCAGCAGTTTTAT <u>AGGAGA</u> GGAGACCCA	96,2	TREB28	TRBJ2-5	784
PT 6	TGTGCCAGCAGCTTT <u>TGTCTGTCT</u> GGGGCCAACGTCTCT	96	TREB5-4	TRBJ2-6	785
PT 51	TGTGCCAGCAGTGAAT <u>TACACAGG</u> GGAAGC	94,9	TREB2	TRBJ1-1	786
PT 59	TGTGCCAGCAGCCTATTTACTAAA AGGGGTTGTTGGATGAGCA	93,6	7	TRBJ2-1	787
PT 49	TGTGCCAGCAGTTTT <u>CGAACTCTT</u> TCCCCGGGGGCGCAATTCGGCC CAAAAACATTCA	93,4	12	TRBJ2-4	788
PT 41	TGTGCCAGCAGTTCACAGATAC GCA	92,9	TREB28	TRBJ2-3	789

Fig. 4A

ID	Secuencia CDR3 de TCRB	%	Gen V	Gen J	SEQ ID NO:
PT 15	<u>TGTGCCAGCAGCCCTGACAGGGG</u> <u>ACGAGCCTGGAAACAC</u>	90,5	TRBV7-3	TRBJ1-3	790
PT 32	<u>TGTGCCAGCAGCCACGGGCGG</u> <u>ATAAAGATGAAAACT</u>	89,1	TRBV9	TRBJ1-4	791
PT 14	<u>TGTGCCAGCAGTTTGAACACAAA</u> <u>CCGGCAGTGGGCGGAGCTAACT</u> ATGGCTA	88,7	5	TRBJ1-2	792
PT 1	<u>TGTGCCAGCATTATTTAGGGTAC</u> <u>CCTAAACTATGGCTA</u>	87,1	TRBV2	TRBJ1-2	793
PT 66	<u>TGTGCCAGCAGTTACTCGCGCTC</u> <u>CGAACCTAATGAGCA</u>	86,9	6	TRBJ2-1	794
PT 65	<u>TGCAGTGCTAGAGATCTACTCCCC</u> <u>CTGGGGGGGACGCGTGAAACA</u> CCAT	84,0	TRBV20-1	TRBJ1-3	795
PT 5	<u>TGTGCCAGTAGTCACCCTCACACA</u> GATACGCA	82,8	TRBV19	TRBJ2-3	796
PT 2	<u>TGTGCCAGCAGTTACATTTGAGGC</u> <u>ITAACTATGGCTA</u>	80,9	6	TRBJ2-7	797
PT 36	<u>TGTGCCAGCAGTTTAATAACAGTI</u> <u>CTTCTCGGGGAGCT</u>	77,5	12	TRBJ2-2	798
PT 40	<u>TGTGCCAGCAGTGAAGGCGGGAG</u> <u>AGCGCTTGGGAGCA</u>	70,9	TRBV2	TRBJ2-1	799
PT 61	TGCAGTGCTAGAGAGGGTAAGAG ACCCA	62,6	TRBV20-1	TRBJ2-5	800
PT 58	<u>TGTGCCAGCAGTTCTATGACAG</u> <u>GGGATCCATTCAGCCCCA</u>	58,6	6	TRBJ1-5	801
PT 47	<u>TGTGCCAGCAGTGATGAAAGGCG</u> <u>GGGCTCCTACAATGAGCA</u>	58,4	TRBV2	TRBJ2-1	802
PT 30	<u>TGTGCCAGCAGTCTAACCGGACA</u> <u>GCGGTATAATGAAAACT</u>	49	TRBV27	TRBJ1-4	803
PT 33	<u>TGTGCCAGCAGCTTATGGGAAAC</u> CGGGGAGCT	39	TRBV11-2	TRBJ2-2	804

Fig. 4B

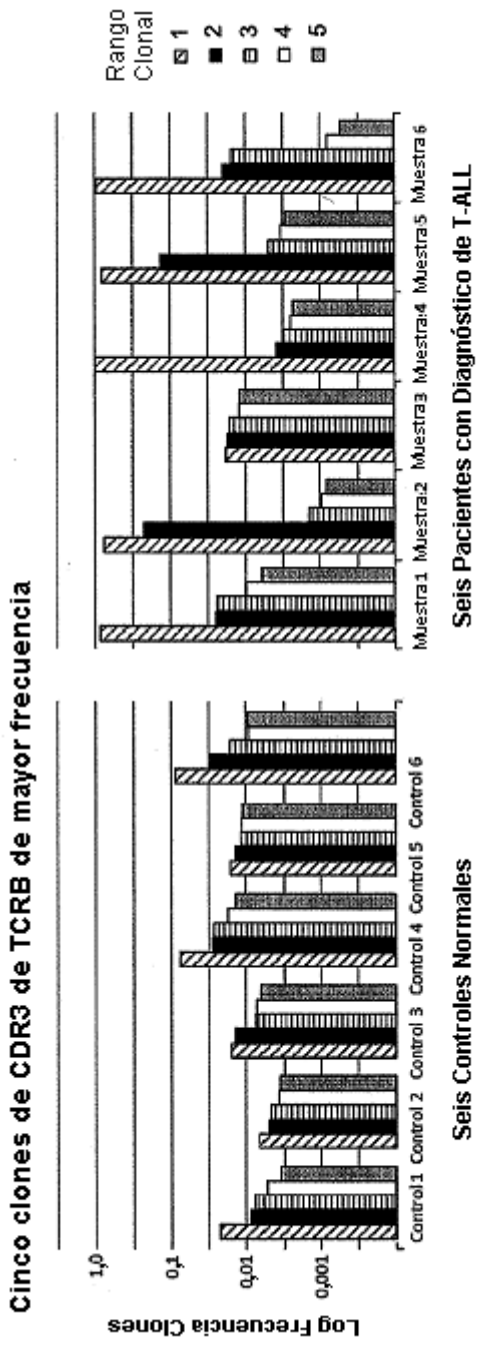


Fig. 5



Fig. 6



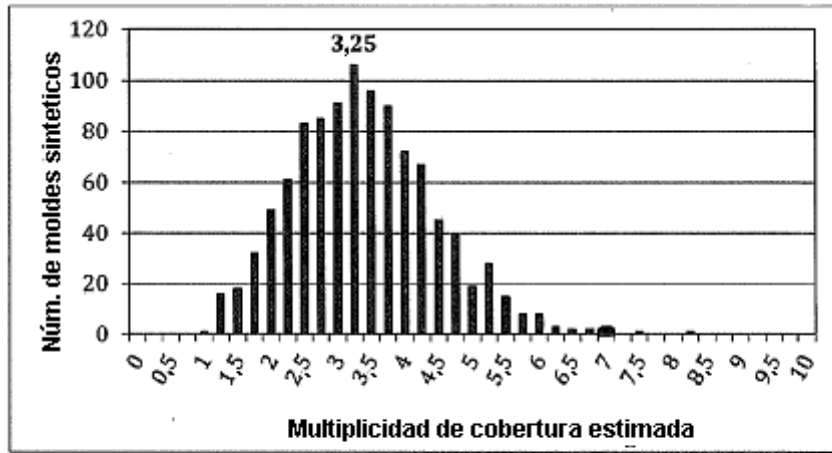


Fig. 7

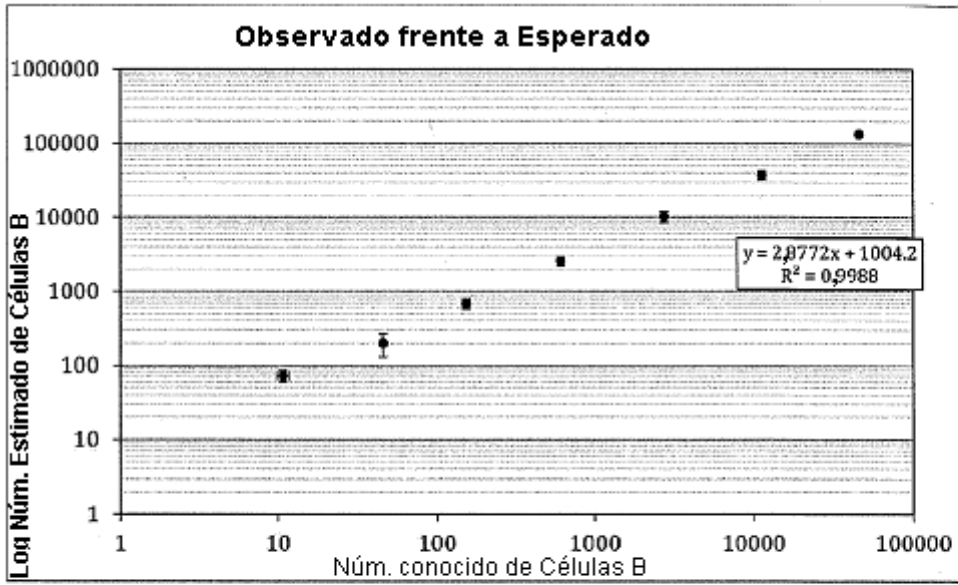


Fig. 8

TCRB V/J CONJUNTO 1		TCRB V/J CONJUNTO 2		TCRB V/J CONJUNTO 3	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
2921	2989	2921	2990	2921	2991
2922	2990	2922	2991	2922	2992
2923	2991	2923	2992	2923	2993
2924	2992	2924	2993	2924	2994
2925	2993	2925	2994	2925	2995
2926	2994	2926	2995	2926	2996
2927	2995	2927	2996	2927	2997
2928	2996	2928	2997	2928	2998
2929	2997	2929	2998	2929	2999
2930	2998	2930	2999	2930	3000
2931	2999	2931	3000	2931	3001
2932	3000	2932	3001	2932	2989
2933	3001	2933	2989	2933	2990
2934	2989	2934	2990	2934	2991
2935	2990	2935	2991	2935	2992
2936	2991	2936	2992	2936	2993
2937	2992	2937	2993	2937	2994
2938	2993	2938	2994	2938	2995
2939	2994	2939	2995	2939	2996
2940	2995	2940	2996	2940	2997
2941	2996	2941	2997	2941	2998
2942	2997	2942	2998	2942	2999
2943	2998	2943	2999	2943	3000
2944	2999	2944	3000	2944	3001
2945	3000	2945	3001	2945	2989
2946	3001	2946	2989	2946	2990
2947	2989	2947	2990	2947	2991

Fig. 9A

TCRB V/J CONJUNTO 1		TCRB V/J CONJUNTO 2		TCRB V/J CONJUNTO 3	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2948	2990	2948	2991	2948	2992
2949	2991	2949	2992	2949	2993
2950	2992	2950	2993	2950	2994
2951	2993	2951	2994	2951	2995
2952	2994	2952	2995	2952	2996
2953	2995	2953	2996	2953	2997
2954	2996	2954	2997	2954	2998
2955	2997	2955	2998	2955	2999
2956	2998	2956	2999	2956	3000
2957	2999	2957	3000	2957	3001
2958	3000	2958	3001	2958	2989
2959	3001	2959	2989	2959	2990
2960	2989	2960	2990	2960	2991
2961	2990	2961	2991	2961	2992
2962	2991	2962	2992	2962	2993
2963	2992	2963	2993	2963	2994
2964	2993	2964	2994	2964	2995
2965	2994	2965	2995	2965	2996
2966	2995	2966	2996	2966	2997
2967	2996	2967	2997	2967	2998
2968	2997	2968	2998	2968	2999
2969	2998	2969	2999	2969	3000
2970	2999	2970	3000	2970	3001
2971	3000	2971	3001	2971	2989
2972	3001	2972	2989	2972	2990
2973	2989	2973	2990	2973	2991
2974	2990	2974	2991	2974	2992
2975	2991	2975	2992	2975	2993
2976	2992	2976	2993	2976	2994
2977	2993	2977	2994	2977	2995
2978	2994	2978	2995	2978	2996

Fig. 9B

TCRB V/J CONJUNTO 1		TCRB V/J CONJUNTO 2		TCRB V/J CONJUNTO 3	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2979	2995	2979	2996	2979	2997
2980	2996	2980	2997	2980	2998
2981	2997	2981	2998	2981	2999
2982	2998	2982	2999	2982	3000
2983	2999	2983	3000	2983	3001
2984	3000	2984	3001	2984	2989
2985	3001	2985	2989	2985	2990
2986	2989	2986	2990	2986	2991
2987	2990	2987	2991	2987	2992
2988	2991	2988	2992	2988	2993

TCRB V/J CONJUNTO 4		TCRB V/J CONJUNTO 5		TCRB V/J CONJUNTO 6	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2921	2992	2921	2993	2921	2994
2922	2993	2922	2994	2922	2995
2923	2994	2923	2995	2923	2996
2924	2995	2924	2996	2924	2997
2925	2996	2925	2997	2925	2998
2926	2997	2926	2998	2926	2999
2927	2998	2927	2999	2927	3000
2928	2999	2928	3000	2928	3001
2929	3000	2929	3001	2929	2989
2930	3001	2930	2989	2930	2990
2931	2989	2931	2990	2931	2991
2932	2990	2932	2991	2932	2992
2933	2991	2933	2992	2933	2993
2934	2992	2934	2993	2934	2994
2935	2993	2935	2994	2935	2995
2936	2994	2936	2995	2936	2996
2937	2995	2937	2996	2937	2997

Fig. 9C

TCRB V/J CONJUNTO 4		TCRB V/J CONJUNTO 5		TCRB V/J CONJUNTO 6	
SEQ. ID NO.:V	SEQ. ID NO.:J	SEQ. ID NO.:V	SEQ. ID NO.:J	SEQ. ID NO.:V	SEQ. ID NO.:J
2938	2996	2938	2997	2938	2998
2939	2997	2939	2998	2939	2999
2940	2998	2940	2999	2940	3000
2941	2999	2941	3000	2941	3001
2942	3000	2942	3001	2942	2989
2943	3001	2943	2989	2943	2990
2944	2989	2944	2990	2944	2991
2945	2990	2945	2991	2945	2992
2946	2991	2946	2992	2946	2993
2947	2992	2947	2993	2947	2994
2948	2993	2948	2994	2948	2995
2949	2994	2949	2995	2949	2996
2950	2995	2950	2996	2950	2997
2951	2996	2951	2997	2951	2998
2952	2997	2952	2998	2952	2999
2953	2998	2953	2999	2953	3000
2954	2999	2954	3000	2954	3001
2955	3000	2955	3001	2955	2989
2956	3001	2956	2989	2956	2990
2957	2989	2957	2990	2957	2991
2958	2990	2958	2991	2958	2992
2959	2991	2959	2992	2959	2993
2960	2992	2960	2993	2960	2994
2961	2993	2961	2994	2961	2995
2962	2994	2962	2995	2962	2996
2963	2995	2963	2996	2963	2997
2964	2996	2964	2997	2964	2998
2965	2997	2965	2998	2965	2999
2966	2998	2966	2999	2966	3000
2967	2999	2967	3000	2967	3001
2968	3000	2968	3001	2968	2989

Fig. 9D

TCRB V/J CONJUNTO 4		TCRB V/J CONJUNTO 5		TCRB V/J CONJUNTO 6	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2969	3001	2969	2989	2969	2990
2970	2989	2970	2990	2970	2991
2971	2990	2971	2991	2971	2992
2972	2991	2972	2992	2972	2993
2973	2992	2973	2993	2973	2994
2974	2993	2974	2994	2974	2995
2975	2994	2975	2995	2975	2996
2976	2995	2976	2996	2976	2997
2977	2996	2977	2997	2977	2998
2978	2997	2978	2998	2978	2999
2979	2998	2979	2999	2979	3000
2980	2999	2980	3000	2980	3001
2981	3000	2981	3001	2981	2989
2982	3001	2982	2989	2982	2990
2983	2989	2983	2990	2983	2991
2984	2990	2984	2991	2984	2992
2985	2991	2985	2992	2985	2993
2986	2992	2986	2993	2986	2994
2987	2993	2987	2994	2987	2995
2988	2994	2988	2995	2988	2996

TCRB V/J CONJUNTO 7		TCRB V/J CONJUNTO 8		TCRB V/J CONJUNTO 9	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2921	2995	2921	2996	2921	2997
2922	2996	2922	2997	2922	2998
2923	2997	2923	2998	2923	2999
2924	2998	2924	2999	2924	3000
2925	2999	2925	3000	2925	3001
2926	3000	2926	3001	2926	2989
2927	3001	2927	2989	2927	2990

Fig. 9E

TCRB V/I CONJUNTO 7		TCRB V/I CONJUNTO 8		TCRB V/I CONJUNTO 9	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2928	2989	2928	2990	2928	2991
2929	2990	2929	2991	2929	2992
2930	2991	2930	2992	2930	2993
2931	2992	2931	2993	2931	2994
2932	2993	2932	2994	2932	2995
2933	2994	2933	2995	2933	2996
2934	2995	2934	2996	2934	2997
2935	2996	2935	2997	2935	2998
2936	2997	2936	2998	2936	2999
2937	2998	2937	2999	2937	3000
2938	2999	2938	3000	2938	3001
2939	3000	2939	3001	2939	2989
2940	3001	2940	2989	2940	2990
2941	2989	2941	2990	2941	2991
2942	2990	2942	2991	2942	2992
2943	2991	2943	2992	2943	2993
2944	2992	2944	2993	2944	2994
2945	2993	2945	2994	2945	2995
2946	2994	2946	2995	2946	2996
2947	2995	2947	2996	2947	2997
2948	2996	2948	2997	2948	2998
2949	2997	2949	2998	2949	2999
2950	2998	2950	2999	2950	3000
2951	2999	2951	3000	2951	3001
2952	3000	2952	3001	2952	2989
2953	3001	2953	2989	2953	2990
2954	2989	2954	2990	2954	2991
2955	2990	2955	2991	2955	2992
2956	2991	2956	2992	2956	2993
2957	2992	2957	2993	2957	2994
2958	2993	2958	2994	2958	2995

Fig. 9F



TCRB V/J CONJUNTO 7		TCRB V/J CONJUNTO 8		TCRB V/J CONJUNTO 9	
SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J	SEQ. ID NO:V	SEQ. ID NO:J
2959	2994	2959	2995	2959	2996
2960	2995	2960	2996	2960	2997
2961	2996	2961	2997	2961	2998
2962	2997	2962	2998	2962	2999
2963	2998	2963	2999	2963	3000
2964	2999	2964	3000	2964	3001
2965	3000	2965	3001	2965	2989
2966	3001	2966	2989	2966	2990
2967	2989	2967	2990	2967	2991
2968	2990	2968	2991	2968	2992
2969	2991	2969	2992	2969	2993
2970	2992	2970	2993	2970	2994
2971	2993	2971	2994	2971	2995
2972	2994	2972	2995	2972	2996
2973	2995	2973	2996	2973	2997
2974	2996	2974	2997	2974	2998
2975	2997	2975	2998	2975	2999
2976	2998	2976	2999	2976	3000
2977	2999	2977	3000	2977	3001
2978	3000	2978	3001	2978	2989
2979	3001	2979	2989	2979	2990
2980	2989	2980	2990	2980	2991
2981	2990	2981	2991	2981	2992
2982	2991	2982	2992	2982	2993
2983	2992	2983	2993	2983	2994
2984	2993	2984	2994	2984	2995
2985	2994	2985	2995	2985	2996
2986	2995	2986	2996	2986	2997
2987	2996	2987	2997	2987	2998
2988	2997	2988	2998	2988	2999

Fig. 9G

TCRB V/J CONJUNTO 10		TCRB V/J CONJUNTO 11		TCRB V/J CONJUNTO 12	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
2921	2998	2921	2999	2921	3000
2922	2999	2922	3000	2922	3001
2923	3000	2923	3001	2923	2989
2924	3001	2924	2989	2924	2990
2925	2989	2925	2990	2925	2991
2926	2990	2926	2991	2926	2992
2927	2991	2927	2992	2927	2993
2928	2992	2928	2993	2928	2994
2929	2993	2929	2994	2929	2995
2930	2994	2930	2995	2930	2996
2931	2995	2931	2996	2931	2997
2932	2996	2932	2997	2932	2998
2933	2997	2933	2998	2933	2999
2934	2998	2934	2999	2934	3000
2935	2999	2935	3000	2935	3001
2936	3000	2936	3001	2936	2989
2937	3001	2937	2989	2937	2990
2938	2989	2938	2990	2938	2991
2939	2990	2939	2991	2939	2992
2940	2991	2940	2992	2940	2993
2941	2992	2941	2993	2941	2994
2942	2993	2942	2994	2942	2995
2943	2994	2943	2995	2943	2996
2944	2995	2944	2996	2944	2997
2945	2996	2945	2997	2945	2998
2946	2997	2946	2998	2946	2999
2947	2998	2947	2999	2947	3000
2948	2999	2948	3000	2948	3001
2949	3000	2949	3001	2949	2989
2950	3001	2950	2989	2950	2990

Fig. 9H

TCRB V/J CONJUNTO 10		TCRB V/J CONJUNTO 11		TCRB V/J CONJUNTO 12	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
2951	2989	2951	2990	2951	2991
2952	2990	2952	2991	2952	2992
2953	2991	2953	2992	2953	2993
2954	2992	2954	2993	2954	2994
2955	2993	2955	2994	2955	2995
2956	2994	2956	2995	2956	2996
2957	2995	2957	2996	2957	2997
2958	2996	2958	2997	2958	2998
2959	2997	2959	2998	2959	2999
2960	2998	2960	2999	2960	3000
2961	2999	2961	3000	2961	3001
2962	3000	2962	3001	2962	2989
2963	3001	2963	2989	2963	2990
2964	2989	2964	2990	2964	2991
2965	2990	2965	2991	2965	2992
2966	2991	2966	2992	2966	2993
2967	2992	2967	2993	2967	2994
2968	2993	2968	2994	2968	2995
2969	2994	2969	2995	2969	2996
2970	2995	2970	2996	2970	2997
2971	2996	2971	2997	2971	2998
2972	2997	2972	2998	2972	2999
2973	2998	2973	2999	2973	3000
2974	2999	2974	3000	2974	3001
2975	3000	2975	3001	2975	2989
2976	3001	2976	2989	2976	2990
2977	2989	2977	2990	2977	2991
2978	2990	2978	2991	2978	2992
2979	2991	2979	2992	2979	2993
2980	2992	2980	2993	2980	2994
2981	2993	2981	2994	2981	2995

Fig. 9I

TCRB V/J CONJUNTO 10		TCRB V/J CONJUNTO 11		TCRB V/J CONJUNTO 12	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
2982	2994	2982	2995	2982	2996
2983	2995	2983	2996	2983	2997
2984	2996	2984	2997	2984	2998
2985	2997	2985	2998	2985	2999
2986	2998	2986	2999	2986	3000
2987	2999	2987	3000	2987	3001
2988	3000	2988	3001	2988	2989

TCRB V/J CONJUNTO 13	
SEQ ID NO:V	SEQ ID NO:J
2921	3001
2922	2989
2923	2990
2924	2991
2925	2992
2926	2993
2927	2994
2928	2995
2929	2996
2930	2997
2931	2998
2932	2999
2933	3000
2934	3001
2935	2989
2936	2990
2937	2991
2938	2992
2939	2993
2940	2994

Fig. 9J

TCRB V/J CONJUNTO 13	
SEQ. ID NO:V	SEQ. ID NO:J
2941	2995
2942	2996
2943	2997
2944	2998
2945	2999
2946	3000
2947	3001
2948	2989
2949	2990
2950	2991
2951	2992
2952	2993
2953	2994
2954	2995
2955	2996
2956	2997
2957	2998
2958	2999
2959	3000
2960	3001
2961	2989
2962	2990
2963	2991
2964	2992
2965	2993
2966	2994
2967	2995
2968	2996
2969	2997
2970	2998
2971	2999

Fig. 9K

TCRB V/I CONJUNTO 13	
SEQ.ID NO:V	SEQ.ID NO:J
2972	3000
2973	3001
2974	2989
2975	2990
2976	2991
2977	2992
2978	2993
2979	2994
2980	2995
2981	2996
2982	2997
2983	2998
2984	2999
2985	3000
2986	3001
2987	2989
2988	2990

Fig. 9L

TCRG V/J CONJUNTO 1		TCRG V/J CONJUNTO 2		TCRG V/J CONJUNTO 3	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
3002	3016	3002	3017	3002	3018
3003	3017	3003	3018	3003	3019
3004	3018	3004	3019	3004	3020
3005	3019	3005	3020	3005	3016
3006	3020	3006	3016	3006	3017
3007	3016	3007	3017	3007	3018
3008	3017	3008	3018	3008	3019
3009	3018	3009	3019	3009	3020
3010	3019	3010	3020	3010	3016
3011	3020	3011	3016	3011	3017
3012	3016	3012	3017	3012	3018
3013	3017	3013	3018	3013	3019
3014	3018	3014	3019	3014	3020
3015	3019	3015	3020	3015	3016

TCRG V/J CONJUNTO 4		TCRG V/J CONJUNTO 5	
SEQ ID NO:V	SEQ ID NO:J	SEQ ID NO:V	SEQ ID NO:J
3002	3019	3002	3020
3003	3020	3003	3016
3004	3016	3004	3017
3005	3017	3005	3018
3006	3018	3006	3019
3007	3019	3007	3020
3008	3020	3008	3016
3009	3016	3009	3017

Fig. 10A

TCRG V/J CONJUNTO 4		TCRG V/J CONJUNTO 5	
SEQ. ID NO.'V	SEQ. ID NO.'J	SEQ. ID NO.'V	SEQ. ID NO.'J
3010	3017	3010	3018
3011	3018	3011	3019
3012	3019	3012	3020
3013	3020	3013	3016
3014	3016	3014	3017
3015	3017	3015	3018

Fig. 10B



IGH V/J CONJUNTO 1		IGH V/J CONJUNTO 2		IGH V/J CONJUNTO 3	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3021	3148	3021	3149	3021	3150
3022	3149	3022	3150	3022	3151
3023	3150	3023	3151	3023	3152
3024	3151	3024	3152	3024	3153
3025	3152	3025	3153	3025	3154
3026	3153	3026	3154	3026	3155
3027	3154	3027	3155	3027	3156
3028	3155	3028	3156	3028	3148
3029	3156	3029	3148	3029	3149
3030	3148	3030	3149	3030	3150
3031	3149	3031	3150	3031	3151
3032	3150	3032	3151	3032	3152
3033	3151	3033	3152	3033	3153
3034	3152	3034	3153	3034	3154
3035	3153	3035	3154	3035	3155
3036	3154	3036	3155	3036	3156
3037	3155	3037	3156	3037	3148
3038	3156	3038	3148	3038	3149
3039	3148	3039	3149	3039	3150
3040	3149	3040	3150	3040	3151
3041	3150	3041	3151	3041	3152
3042	3151	3042	3152	3042	3153
3043	3152	3043	3153	3043	3154
3044	3153	3044	3154	3044	3155
3045	3154	3045	3155	3045	3156
3046	3155	3046	3156	3046	3148
3047	3156	3047	3148	3047	3149

Fig. 11A

IGH V/J CONJUNTO 1		IGH V/J CONJUNTO 2		IGH V/J CONJUNTO 3	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3048	3148	3048	3149	3048	3150
3049	3149	3049	3150	3049	3151
3050	3150	3050	3151	3050	3152
3051	3151	3051	3152	3051	3153
3052	3152	3052	3153	3052	3154
3053	3153	3053	3154	3053	3155
3054	3154	3054	3155	3054	3156
3055	3155	3055	3156	3055	3148
3056	3156	3056	3148	3056	3149
3057	3148	3057	3149	3057	3150
3058	3149	3058	3150	3058	3151
3059	3150	3059	3151	3059	3152
3060	3151	3060	3152	3060	3153
3061	3152	3061	3153	3061	3154
3062	3153	3062	3154	3062	3155
3063	3154	3063	3155	3063	3156
3064	3155	3064	3156	3064	3148
3065	3156	3065	3148	3065	3149
3066	3148	3066	3149	3066	3150
3067	3149	3067	3150	3067	3151
3068	3150	3068	3151	3068	3152
3069	3151	3069	3152	3069	3153
3070	3152	3070	3153	3070	3154
3071	3153	3071	3154	3071	3155
3072	3154	3072	3155	3072	3156
3073	3155	3073	3156	3073	3148
3074	3156	3074	3148	3074	3149
3075	3148	3075	3149	3075	3150
3076	3149	3076	3150	3076	3151
3077	3150	3077	3151	3077	3152
3078	3151	3078	3152	3078	3153

Fig. 11B

IGH V/J CONJUNTO 1		IGH V/J CONJUNTO 2		IGH V/J CONJUNTO 3	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3079	3152	3079	3153	3079	3154
3080	3153	3080	3154	3080	3155
3081	3154	3081	3155	3081	3156
3082	3155	3082	3156	3082	3148
3083	3156	3083	3148	3083	3149
3084	3148	3084	3149	3084	3150
3085	3149	3085	3150	3085	3151
3086	3150	3086	3151	3086	3152
3087	3151	3087	3152	3087	3153
3088	3152	3088	3153	3088	3154
3089	3153	3089	3154	3089	3155
3090	3154	3090	3155	3090	3156
3091	3155	3091	3156	3091	3148
3092	3156	3092	3148	3092	3149
3093	3148	3093	3149	3093	3150
3094	3149	3094	3150	3094	3151
3095	3150	3095	3151	3095	3152
3096	3151	3096	3152	3096	3153
3097	3152	3097	3153	3097	3154
3098	3153	3098	3154	3098	3155
3099	3154	3099	3155	3099	3156
3100	3155	3100	3156	3100	3148
3101	3156	3101	3148	3101	3149
3102	3148	3102	3149	3102	3150
3103	3149	3103	3150	3103	3151
3104	3150	3104	3151	3104	3152
3105	3151	3105	3152	3105	3153
3106	3152	3106	3153	3106	3154
3107	3153	3107	3154	3107	3155
3108	3154	3108	3155	3108	3156
3109	3155	3109	3156	3109	3148

Fig. 11C

IGH V/J CONJUNTO 1		IGH V/J CONJUNTO 2		IGH V/J CONJUNTO 3	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3110	3156	3110	3148	3110	3149
3111	3148	3111	3149	3111	3150
3112	3149	3112	3150	3112	3151
3113	3150	3113	3151	3113	3152
3114	3151	3114	3152	3114	3153
3115	3152	3115	3153	3115	3154
3116	3153	3116	3154	3116	3155
3117	3154	3117	3155	3117	3156
3118	3155	3118	3156	3118	3148
3119	3156	3119	3148	3119	3149
3120	3148	3120	3149	3120	3150
3121	3149	3121	3150	3121	3151
3122	3150	3122	3151	3122	3152
3123	3151	3123	3152	3123	3153
3124	3152	3124	3153	3124	3154
3125	3153	3125	3154	3125	3155
3126	3154	3126	3155	3126	3156
3127	3155	3127	3156	3127	3148
3128	3156	3128	3148	3128	3149
3129	3148	3129	3149	3129	3150
3130	3149	3130	3150	3130	3151
3131	3150	3131	3151	3131	3152
3132	3151	3132	3152	3132	3153
3133	3152	3133	3153	3133	3154
3134	3153	3134	3154	3134	3155
3135	3154	3135	3155	3135	3156
3136	3155	3136	3156	3136	3148
3137	3156	3137	3148	3137	3149
3138	3148	3138	3149	3138	3150
3139	3149	3139	3150	3139	3151
3140	3150	3140	3151	3140	3152

Fig. 11D

IGH V/J CONJUNTO 1		IGH V/J CONJUNTO 2		IGH V/J CONJUNTO 3	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3141	3151	3141	3152	3141	3153
3142	3152	3142	3153	3142	3154
3143	3153	3143	3154	3143	3155
3144	3154	3144	3155	3144	3156
3145	3155	3145	3156	3145	3148
3146	3156	3146	3148	3146	3149
3147	3148	3147	3149	3147	3150

IGH V/J CONJUNTO 4		IGH V/J CONJUNTO 5		IGH V/J CONJUNTO 6	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3021	3151	3021	3152	3021	3153
3022	3152	3022	3153	3022	3154
3023	3153	3023	3154	3023	3155
3024	3154	3024	3155	3024	3156
3025	3155	3025	3156	3025	3148
3026	3156	3026	3148	3026	3149
3027	3148	3027	3149	3027	3150
3028	3149	3028	3150	3028	3151
3029	3150	3029	3151	3029	3152
3030	3151	3030	3152	3030	3153
3031	3152	3031	3153	3031	3154
3032	3153	3032	3154	3032	3155
3033	3154	3033	3155	3033	3156
3034	3155	3034	3156	3034	3148
3035	3156	3035	3148	3035	3149
3036	3148	3036	3149	3036	3150
3037	3149	3037	3150	3037	3151
3038	3150	3038	3151	3038	3152
3039	3151	3039	3152	3039	3153
3040	3152	3040	3153	3040	3154

Fig. 11E

IGH V/J CONJUNTO 4		IGH V/J CONJUNTO 5		IGH V/J CONJUNTO 6	
SEQ. ID NO.: V	SEQ. ID NO.: J	SEQ. ID NO.: V	SEQ. ID NO.: J	SEQ. ID NO.: V	SEQ. ID NO.: J
3041	3153	3041	3154	3041	3155
3042	3154	3042	3155	3042	3156
3043	3155	3043	3156	3043	3148
3044	3156	3044	3148	3044	3149
3045	3148	3045	3149	3045	3150
3046	3149	3046	3150	3046	3151
3047	3150	3047	3151	3047	3152
3048	3151	3048	3152	3048	3153
3049	3152	3049	3153	3049	3154
3050	3153	3050	3154	3050	3155
3051	3154	3051	3155	3051	3156
3052	3155	3052	3156	3052	3148
3053	3156	3053	3148	3053	3149
3054	3148	3054	3149	3054	3150
3055	3149	3055	3150	3055	3151
3056	3150	3056	3151	3056	3152
3057	3151	3057	3152	3057	3153
3058	3152	3058	3153	3058	3154
3059	3153	3059	3154	3059	3155
3060	3154	3060	3155	3060	3156
3061	3155	3061	3156	3061	3148
3062	3156	3062	3148	3062	3149
3063	3148	3063	3149	3063	3150
3064	3149	3064	3150	3064	3151
3065	3150	3065	3151	3065	3152
3066	3151	3066	3152	3066	3153
3067	3152	3067	3153	3067	3154
3068	3153	3068	3154	3068	3155
3069	3154	3069	3155	3069	3156
3070	3155	3070	3156	3070	3148
3071	3156	3071	3148	3071	3149

Fig. 11F

IGH V/J CONJUNTO 4		IGH V/J CONJUNTO 5		IGH V/J CONJUNTO 6	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3072	3148	3072	3149	3072	3150
3073	3149	3073	3150	3073	3151
3074	3150	3074	3151	3074	3152
3075	3151	3075	3152	3075	3153
3076	3152	3076	3153	3076	3154
3077	3153	3077	3154	3077	3155
3078	3154	3078	3155	3078	3156
3079	3155	3079	3156	3079	3148
3080	3156	3080	3148	3080	3149
3081	3148	3081	3149	3081	3150
3082	3149	3082	3150	3082	3151
3083	3150	3083	3151	3083	3152
3084	3151	3084	3152	3084	3153
3085	3152	3085	3153	3085	3154
3086	3153	3086	3154	3086	3155
3087	3154	3087	3155	3087	3156
3088	3155	3088	3156	3088	3148
3089	3156	3089	3148	3089	3149
3090	3148	3090	3149	3090	3150
3091	3149	3091	3150	3091	3151
3092	3150	3092	3151	3092	3152
3093	3151	3093	3152	3093	3153
3094	3152	3094	3153	3094	3154
3095	3153	3095	3154	3095	3155
3096	3154	3096	3155	3096	3156
3097	3155	3097	3156	3097	3148
3098	3156	3098	3148	3098	3149
3099	3148	3099	3149	3099	3150
3100	3149	3100	3150	3100	3151
3101	3150	3101	3151	3101	3152
3102	3151	3102	3152	3102	3153

Fig. 11G

IGH V/J SECUENCIA 4		IGH V/J SECUENCIA 6		IGH V/J SECUENCIA 6	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3103	3152	3103	3153	3103	3154
3104	3153	3104	3154	3104	3155
3105	3154	3105	3155	3105	3156
3106	3155	3106	3156	3106	3148
3107	3156	3107	3148	3107	3149
3108	3148	3108	3149	3108	3150
3109	3149	3109	3150	3109	3151
3110	3150	3110	3151	3110	3152
3111	3151	3111	3152	3111	3153
3112	3152	3112	3153	3112	3154
3113	3153	3113	3154	3113	3155
3114	3154	3114	3155	3114	3156
3115	3155	3115	3156	3115	3148
3116	3156	3116	3148	3116	3149
3117	3148	3117	3149	3117	3150
3118	3149	3118	3150	3118	3151
3119	3150	3119	3151	3119	3152
3120	3151	3120	3152	3120	3153
3121	3152	3121	3153	3121	3154
3122	3153	3122	3154	3122	3155
3123	3154	3123	3155	3123	3156
3124	3155	3124	3156	3124	3148
3125	3156	3125	3148	3125	3149
3126	3148	3126	3149	3126	3150
3127	3149	3127	3150	3127	3151
3128	3150	3128	3151	3128	3152
3129	3151	3129	3152	3129	3153
3130	3152	3130	3153	3130	3154
3131	3153	3131	3154	3131	3155
3132	3154	3132	3155	3132	3156
3133	3155	3133	3156	3133	3148

Fig. 11H



IGH V/J CONJUNTO 4		IGH V/J CONJUNTO 5		IGH V/J CONJUNTO 6	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3134	3156	3134	3148	3134	3149
3135	3148	3135	3149	3135	3150
3136	3149	3136	3150	3136	3151
3137	3150	3137	3151	3137	3152
3138	3151	3138	3152	3138	3153
3139	3152	3139	3153	3139	3154
3140	3153	3140	3154	3140	3155
3141	3154	3141	3155	3141	3156
3142	3155	3142	3156	3142	3148
3143	3156	3143	3148	3143	3149
3144	3148	3144	3149	3144	3150
3145	3149	3145	3150	3145	3151
3146	3150	3146	3151	3146	3152
3147	3151	3147	3152	3147	3153

IGH V/J CONJUNTO 7		IGH V/J CONJUNTO 8		IGH V/J CONJUNTO 9	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3021	3154	3021	3155	3021	3156
3022	3155	3022	3156	3022	3148
3023	3156	3023	3148	3023	3149
3024	3148	3024	3149	3024	3150
3025	3149	3025	3150	3025	3151
3026	3150	3026	3151	3026	3152
3027	3151	3027	3152	3027	3153
3028	3152	3028	3153	3028	3154
3029	3153	3029	3154	3029	3155
3030	3154	3030	3155	3030	3156
3031	3155	3031	3156	3031	3148
3032	3156	3032	3148	3032	3149
3033	3148	3033	3149	3033	3150

Fig. 111

IGH V/J CONJUNTO 7		IGH V/J CONJUNTO 8		IGH V/J CONJUNTO 9	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3034	3149	3034	3150	3034	3151
3035	3150	3035	3151	3035	3152
3036	3151	3036	3152	3036	3153
3037	3152	3037	3153	3037	3154
3038	3153	3038	3154	3038	3155
3039	3154	3039	3155	3039	3156
3040	3155	3040	3156	3040	3148
3041	3156	3041	3148	3041	3149
3042	3148	3042	3149	3042	3150
3043	3149	3043	3150	3043	3151
3044	3150	3044	3151	3044	3152
3045	3151	3045	3152	3045	3153
3046	3152	3046	3153	3046	3154
3047	3153	3047	3154	3047	3155
3048	3154	3048	3155	3048	3156
3049	3155	3049	3156	3049	3148
3050	3156	3050	3148	3050	3149
3051	3148	3051	3149	3051	3150
3052	3149	3052	3150	3052	3151
3053	3150	3053	3151	3053	3152
3054	3151	3054	3152	3054	3153
3055	3152	3055	3153	3055	3154
3056	3153	3056	3154	3056	3155
3057	3154	3057	3155	3057	3156
3058	3155	3058	3156	3058	3148
3059	3156	3059	3148	3059	3149
3060	3148	3060	3149	3060	3150
3061	3149	3061	3150	3061	3151
3062	3150	3062	3151	3062	3152
3063	3151	3063	3152	3063	3153
3064	3152	3064	3153	3064	3154

Fig. 11J

IGH V/J: CONJUNTO 7		IGH V/J: CONJUNTO 8		IGH V/J: CONJUNTO 9	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3065	3153	3065	3154	3065	3155
3066	3154	3066	3155	3066	3156
3067	3155	3067	3156	3067	3148
3068	3156	3068	3148	3068	3149
3069	3148	3069	3149	3069	3150
3070	3149	3070	3150	3070	3151
3071	3150	3071	3151	3071	3152
3072	3151	3072	3152	3072	3153
3073	3152	3073	3153	3073	3154
3074	3153	3074	3154	3074	3155
3075	3154	3075	3155	3075	3156
3076	3155	3076	3156	3076	3148
3077	3156	3077	3148	3077	3149
3078	3148	3078	3149	3078	3150
3079	3149	3079	3150	3079	3151
3080	3150	3080	3151	3080	3152
3081	3151	3081	3152	3081	3153
3082	3152	3082	3153	3082	3154
3083	3153	3083	3154	3083	3155
3084	3154	3084	3155	3084	3156
3085	3155	3085	3156	3085	3148
3086	3156	3086	3148	3086	3149
3087	3148	3087	3149	3087	3150
3088	3149	3088	3150	3088	3151
3089	3150	3089	3151	3089	3152
3090	3151	3090	3152	3090	3153
3091	3152	3091	3153	3091	3154
3092	3153	3092	3154	3092	3155
3093	3154	3093	3155	3093	3156
3094	3155	3094	3156	3094	3148
3095	3156	3095	3148	3095	3149

Fig. 11K

IGH V/J CONJUNTO 7		IGH V/J CONJUNTO 8		IGH V/J CONJUNTO 9	
SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J	SEQ. ID NO: V	SEQ. ID NO: J
3096	3148	3096	3149	3096	3150
3097	3149	3097	3150	3097	3151
3098	3150	3098	3151	3098	3152
3099	3151	3099	3152	3099	3153
3100	3152	3100	3153	3100	3154
3101	3153	3101	3154	3101	3155
3102	3154	3102	3155	3102	3156
3103	3155	3103	3156	3103	3148
3104	3156	3104	3148	3104	3149
3105	3148	3105	3149	3105	3150
3106	3149	3106	3150	3106	3151
3107	3150	3107	3151	3107	3152
3108	3151	3108	3152	3108	3153
3109	3152	3109	3153	3109	3154
3110	3153	3110	3154	3110	3155
3111	3154	3111	3155	3111	3156
3112	3155	3112	3156	3112	3148
3113	3156	3113	3148	3113	3149
3114	3148	3114	3149	3114	3150
3115	3149	3115	3150	3115	3151
3116	3150	3116	3151	3116	3152
3117	3151	3117	3152	3117	3153
3118	3152	3118	3153	3118	3154
3119	3153	3119	3154	3119	3155
3120	3154	3120	3155	3120	3156
3121	3155	3121	3156	3121	3148
3122	3156	3122	3148	3122	3149
3123	3148	3123	3149	3123	3150
3124	3149	3124	3150	3124	3151
3125	3150	3125	3151	3125	3152
3126	3151	3126	3152	3126	3153

Fig. 11L

IGH V/J CONJUNTO 7		IGH V/J CONJUNTO 8		IGH V/J CONJUNTO 9	
SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J	SEQ ID NO: V	SEQ ID NO: J
3127	3152	3127	3153	3127	3154
3128	3153	3128	3154	3128	3155
3129	3154	3129	3155	3129	3156
3130	3155	3130	3156	3130	3148
3131	3156	3131	3148	3131	3149
3132	3148	3132	3149	3132	3150
3133	3149	3133	3150	3133	3151
3134	3150	3134	3151	3134	3152
3135	3151	3135	3152	3135	3153
3136	3152	3136	3153	3136	3154
3137	3153	3137	3154	3137	3155
3138	3154	3138	3155	3138	3156
3139	3155	3139	3156	3139	3148
3140	3156	3140	3148	3140	3149
3141	3148	3141	3149	3141	3150
3142	3149	3142	3150	3142	3151
3143	3150	3143	3151	3143	3152
3144	3151	3144	3152	3144	3153
3145	3152	3145	3153	3145	3154
3146	3153	3146	3154	3146	3155
3147	3154	3147	3155	3147	3156

Fig. 11M

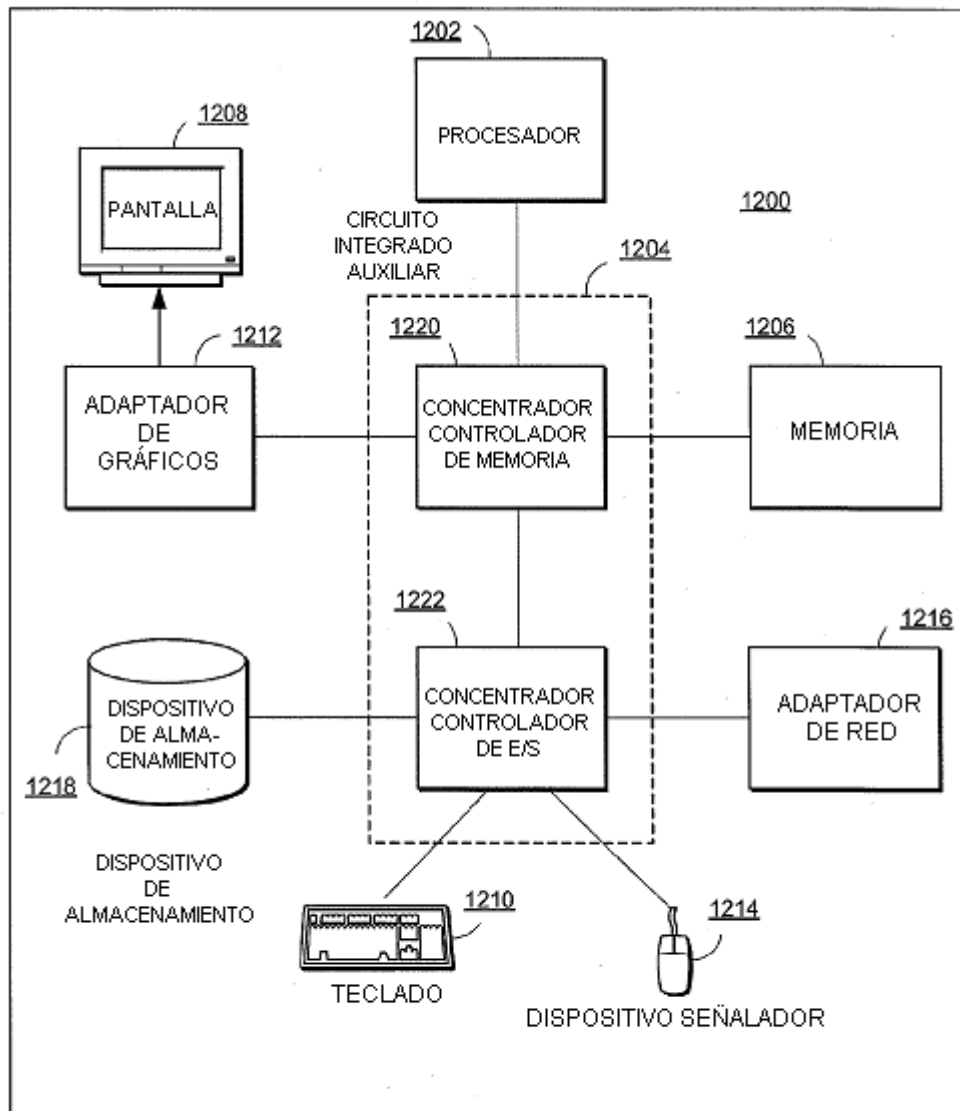


Fig. 12