

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 777 529**

51 Int. Cl.:

C12Q 1/6881 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.03.2015 PCT/US2015/018967**

87 Fecha y número de publicación internacional: **22.10.2015 WO15160439**

96 Fecha de presentación y número de la solicitud europea: **05.03.2015 E 15779750 (7)**

97 Fecha y número de publicación de la concesión europea: **08.01.2020 EP 3132059**

54 Título: **Cuantificación de genomas de células inmunitarias adaptativas en una mezcla compleja de células**

30 Prioridad:

17.04.2014 US 201461981085 P
24.04.2014 US 201461983920 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.08.2020

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:

ROBINS, HARLAN SAUL;
CARLSON, CHRISTOPHER SCOTT;
RIEDER, MARK J.;
SHERWOOD, ANNA M. y
EMERSON, RYAN O.

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 777 529 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

Cuantificación de genomas de células inmunitarias adaptativas en una mezcla compleja de células

5 **Antecedentes**

El sistema inmunitario adaptativo protege a los organismos superiores contra las infecciones y otros hechos patológicos que pueden atribuirse a sustancias foráneas, usando receptores inmunitarios adaptativos, las proteínas de reconocimiento específicas del antígeno que son expresadas por las células hematopoyéticas del linaje linfocítico y que son capaces de distinguir las moléculas propias de las no propias en el hospedador. Estos linfocitos se pueden encontrar en la circulación y en los tejidos de un hospedador, y se ha descrito su recirculación entre la sangre y los vasos linfáticos, incluyendo su extravasación a través de las vénulas endoteliales altas de los ganglios linfáticos, así como en sitios de infección, inflamación, lesión tisular y otras lesiones clínicas. (Véase, por ejemplo, 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).

Por consiguiente, la naturaleza dinámica del movimiento de los linfocitos en todo el organismo del hospedador se refleja en los cambios en la calidad (por ejemplo, la especificidad por el antígeno del receptor inmunitario adaptativo expresado clónicamente (inmunoglobulina o receptor de linfocitos T), linfocito T frente a linfocito B, linfocito T auxiliar (T_h) frente a linfocito T regulador (T_{reg}), linfocito T efector frente a linfocito T de la memoria, etc.) y la distribución cuantitativa de los linfocitos entre tejidos, en función de los cambios en el estado inmunitario del hospedador.

Por ejemplo, numerosos estudios han encontrado una asociación entre (i) la presencia de linfocitos infiltrantes tumorales (TIL, *Tumor Infiltrating Lymphocytes*) en una variedad de tumores sólidos e (ii) el pronóstico del paciente y las tasas de supervivencia general. En algunos estudios, los linfocitos T infiltrantes tumorales que tienen un fenotipo específico (por ejemplo, linfocitos T CD8+ y CD4+ o linfocitos T reguladores) son predictores positivos o negativos de la supervivencia (por ejemplo, Jochems *et al.*, 2011 *Experimental Biol. Med.* 236:567-579). En determinados casos, sin embargo, el recuento de TIL solo es un predictor de la supervivencia a largo plazo (por ejemplo, Katz *et al.*, 2009 *Ann. Surg. Oncol.* 16:2524-2530). Por lo tanto, la determinación cuantitativa de los recuentos de TIL tiene un alto valor de pronóstico en una variedad de cánceres que incluyen el cáncer colorrectal, hepatocelular, de vesícula biliar, pancreático, esofágico, ovárico, endometrial, de cuello de útero, de vejiga y urotelial. Si bien se sabe más sobre la asociación de los linfocitos T infiltrantes tumorales, también se sabe que los linfocitos B se infiltran en los tumores, y los estudios han demostrado una asociación de los linfocitos B infiltrantes tumorales con una ventaja de supervivencia (por ejemplo, Ladányi, *et al.*, *Cancer Immunol. Immunother.* 60(12):1729-38, 21 de julio de 2011 (epub antes de la impresión).

Por lo tanto, la determinación cuantitativa de la presencia de células inmunitarias adaptativas (por ejemplo, linfocitos T y B) en tejidos enfermos puede proporcionar información útil para el diagnóstico, pronóstico y otros fines, tales como en el cáncer, la infección, la inflamación, la lesión tisular y otras afecciones.

El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígenos de linfocitos T y B con suficiente diversidad para reconocer el universo de posibles patógenos. Los linfocitos B maduran para expresar anticuerpos (inmunoglobulinas, Ig) que se producen como heterodímeros de un polipéptido de cadena pesada (H) y ligera (L), mientras que los linfocitos T expresan receptores de linfocitos T (TCR, *T Cell Receptors*) heterodiméricos. La capacidad de los linfocitos T para reconocer el universo de antígenos asociados con los diferentes cánceres u organismos infecciosos es conferida por su receptor de antígeno de linfocitos T (TCR), que está constituido tanto por una cadena α (alfa) como por una cadena β (beta) o tanto por una cadena γ (gamma) como por una cadena δ (delta). Las proteínas que forman estas cadenas están codificadas por ADN, que emplea un mecanismo único para generar la tremenda diversidad de los TCR. Este receptor de reconocimiento inmunitario de múltiples subunidades se asocia con el complejo CD3 y se une a los péptidos presentados por proteínas del complejo principal de histocompatibilidad (MHC) de clase I o de clase II sobre la superficie de las células presentadoras de antígeno (APC, *Antigen-Presenting Cells*). La unión del TCR con el péptido antigénico en la APC es el hecho central en la activación de los linfocitos T, que se produce en una sinapsis inmunitaria en el punto de contacto entre el linfocito T y la APC.

Cada péptido TCR contiene regiones variables determinantes de complementariedad (CDR, *Complementarity Determining Regions*), así como regiones marco conservadas (FR, *Framework Regions*) y una región constante. La diversidad de secuencias de los linfocitos T $\alpha\beta$ está determinada en gran medida por la secuencia de aminoácidos de los bucles de la tercera región determinante de complementariedad (CDR3) de los dominios variables de la cadena α y β , cuya diversidad es el resultado de la recombinación entre los segmentos génicos de la región variable ($V\beta$), de diversidad ($D\beta$) y de unión ($J\beta$) en el locus de la cadena β , y entre los segmentos génicos $V\alpha$ y $J\alpha$ en el locus de la cadena α , respectivamente. La existencia de múltiples de dichos segmentos génicos en los locus de las cadenas α y β del TCR permite que se codifique un gran número de secuencias de CDR3 distintas. La diversidad de secuencias de CDR3 se aumenta aún más mediante la adición y eliminación independientes de nucleótidos en las uniones $V\beta$ - $D\beta$, $D\beta$ - $J\beta$ y $V\alpha$ - $J\alpha$ durante el proceso de reordenación génica del TCR. A este respecto, la

inmunocompetencia se refleja en la diversidad de los TCR.

El TCR $\gamma\delta$ se distingue del TCR $\alpha\beta$ en codifica un receptor que interactúa estrechamente con el sistema inmunitario innato. TCR $\gamma\delta$ se expresa temprano en el desarrollo, tiene una distribución anatómica especializada, tiene un único patógeno y especificidades con moléculas pequeñas, así como un amplio espectro de interacciones celulares innatas y adaptativas. Un patrón sesgado de la expresión del segmento V y J de TCR γ se establece temprano en la ontogenia como los subconjuntos restringidos de células TCR $\gamma\delta$ pueblan la boca, la piel, el intestino, la vagina y los pulmones prenatalmente. Por consiguiente, el repertorio diverso de TCR γ en tejidos adultos es el resultado de una extensa expansión periférica tras la estimulación por la exposición ambiental a patógenos y moléculas tóxicas.

Las Ig expresadas por linfocitos 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_2L_2 . Cada par de cadenas H y L contiene un dominio hipervariable, que consiste en una región V_L y V_H , y un dominio constante. Las cadenas H de las Ig son de varios tipos, μ , δ , γ , α y β . La diversidad de las Ig dentro de un individuo está determinada principalmente por el dominio hipervariable. Al igual que el TCR, el dominio V de las cadenas H es creado por la unión combinatoria de los segmentos génicos V_H , D_H y J_H . La diversidad de secuencias del dominio hipervariable se aumenta aún más mediante la adición y eliminación independientes de nucleótidos en las uniones V_H-D_H , D_H-J_H y V_H-J_H durante el proceso de reordenación génica de las Ig. A este respecto, la inmunocompetencia se refleja en la diversidad de las Ig.

La caracterización cuantitativa de las células inmunitarias adaptativas basándose en la presencia en dichas células de genes codificantes de Ig y TCR reordenados funcionalmente que dirigen la expresión productiva de los receptores inmunitarios adaptativos se ha logrado usando muestras biológicas de las que se pueden aislar fácilmente células inmunitarias adaptativas 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, por ejemplo, el documento US 2010/0330571; véase también, por ejemplo, Murphy, "Janeway's Immunobiology" (8ª Ed.), 2011 Garland Science, NY, Apéndice I, pág. 717-762. Los métodos anteriores incluyen la cuantificación de la representación relativa de las células inmunitarias adaptativas en una muestra mediante la amplificación de polipéptidos de la región V, polipéptidos de la región J y un gen de control interno de la muestra, y la comparación del número de células que contienen polipéptidos de la región V y J con el número de células que contienen el gen de control interno. Véase, por ejemplo, el documento US2013/0288237. Sin embargo, este método no permite la cuantificación absoluta de las células inmunitarias adaptativas en la muestra. Aunque se puede determinar una representación relativa de las células inmunitarias adaptativas, los métodos actuales no permiten la determinación del número absoluto de células inmunitarias adaptativas en la muestra de entrada.

Existe la necesidad de un método que permita la cuantificación absoluta exacta de las células inmunitarias adaptativas en una muestra biológica compleja. También existe la necesidad de un método mejorado de cuantificación de una representación relativa de células inmunitarias adaptativas en dicha muestra biológica compleja.

Sumario de la invención

La invención proporciona un método de determinación de una proporción de linfocitos T o linfocitos B en una muestra en relación con el número total de genomas de entrada contenidos en dicha muestra que comprende:

A) amplificar por PCR multiplexada y secuenciar:

i) secuencias oligonucleotídicas de CDR3 reordenadas de locus de receptores de linfocitos T (TCR) de linfocitos T o locus de inmunoglobulinas (Ig) de linfocitos B en dicha muestra para obtener un número total de secuencias biológicas de salida, comprendiendo cada secuencia oligonucleotídica un segmento V y un segmento J;

ii) un primer conjunto de moldes sintéticos que representan esencialmente todas las combinaciones posibles de segmentos V y segmentos J, donde cada molde sintético del primer conjunto de moldes sintéticos comprende:

(a) una combinación única de un segmento V y un segmento J o C, donde el segmento V es un segmento V de TCR o Ig, y donde el segmento J o C es un segmento J o C de TCR o Ig; y

(b) un código de barras único que identifica dicho molde sintético como sintético;

B) determinar un factor de amplificación para cada molde sintético que comprende una combinación única de un segmento V y un segmento J o C, donde dicho factor de amplificación está representado por el número total de primeros moldes sintéticos amplificados y secuenciados en la etapa A (ii) dividido entre el número total de entrada de los primeros moldes sintéticos únicos introducidos en la etapa A (ii);

C) determinar el número total de linfocitos T o linfocitos B de la muestra dividiendo el número total de secuencias biológicas de salida observadas en la etapa A (i) entre el factor de amplificación de la etapa B;

D) amplificar por PCR multiplexada y secuenciar:

i) una o más regiones de control genómico del ADN obtenido de dicha muestra para obtener un número total de secuencias biológicas de salida para cada región de control genómico; e

5 ii) un segundo conjunto de moldes sintéticos, donde cada molde sintético del segundo conjunto de moldes sintéticos comprende la secuencia de una o más de dichas secuencias de control genómico, un código de barras único y un tramo de ácidos nucleicos aleatorios, y donde cada molde sintético del segundo conjunto de moldes sintéticos se representa solo una vez;

10 E) determinar un factor de amplificación para cada una de dichas regiones de control genómico dividiendo el número total de segundos moldes sintéticos amplificados y secuenciados en la etapa D (ii) entre el número total de entrada de los segundos moldes sintéticos únicos introducidos en la etapa D (ii);

F) determinar el número total de genomas de entrada dividiendo el número total de secuencias biológicas de salida para cada región de control genómico de la etapa D (i) entre el factor de amplificación correspondiente para esa región de control genómico de la etapa E; y

15 G) determinar la proporción de linfocitos T o linfocitos B contenidos en la muestra con respecto al número de genomas totales de la muestra dividiendo el número total de linfocitos T o linfocitos B obtenidos en la etapa C entre el número total de genomas de entrada obtenidos en la etapa F.

20 En el presente documento, se divulgan métodos de cuantificación de una serie de células inmunitarias adaptativas en una muestra biológica, y métodos de cuantificación de una representación relativa de células inmunitarias adaptativas en una muestra biológica que comprende una mezcla de células que comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas.

25 En un caso, se proporciona un método de determinación de la proporción de linfocitos T o B en una muestra en relación con el número total de genomas de entrada. En un caso, el método proporciona la amplificación por PCR multiplexada y la secuenciación de locus de receptores de linfocitos T (TCR) reordenados de linfocitos T o locus de inmunoglobulina (Ig) de células V en una muestra para obtener un número total de secuencias biológicas de salida. En un caso adicional, se proporcionan métodos de amplificación mediante PCR multiplexada y secuenciación de un primer conjunto de moldes sintéticos, comprendiendo cada uno de ellos un segmento V de TCR o Ig y un segmento J o C de TCR o Ig, y un código de barras único que identifica dicho molde sintético como sintético. En un caso, cada molde sintético comprende una combinación única de segmentos V y J o C. En otro caso, el método proporciona la determinación de un factor de amplificación que está representado por el número total de primeros moldes sintéticos amplificados y secuenciados dividido entre el número de entrada total de los primeros moldes sintéticos únicos. En una realización, el método proporciona la determinación del número total de linfocitos T o linfocitos B de la muestra dividiendo el número total de secuencias biológicas de salida entre el factor de amplificación.

40 En un caso, el método proporciona además la amplificación mediante PCR multiplexada y secuenciación de una o más regiones de control genómico a partir de ADN obtenido de una muestra para obtener el número total de secuencias biológicas de salida para cada región de control genómico. En otro caso, se proporcionan métodos de amplificación mediante PCR multiplexada y secuenciación de un segundo conjunto de moldes sintéticos, comprendiendo cada uno la secuencia de una o más de dichas regiones de control genómico, un código de barras único y un tramo de ácidos nucleicos aleatorios. En un caso, cada molde sintético del segundo conjunto de moldes sintéticos se representa solo una vez. En un caso adicional, el método proporciona la determinación de un factor de amplificación para cada una de las regiones de control genómico dividiendo el número total de segundos moldes sintéticos amplificados y secuenciados entre el número de entrada total de segundos moldes sintéticos únicos. En un caso, el método proporciona además un método de determinación del número total de genomas de entrada dividiendo el número total de secuencias biológicas de salida de cada región de control genómico entre el factor de amplificación correspondiente para esa región de control genómico. En otro caso más, los métodos proporcionan la determinación de la proporción de linfocitos T o linfocitos B contenidos en la muestra en relación con los genomas totales dividiendo el número total de linfocitos T o linfocitos B de la muestra entre el número total de genomas de entrada de la muestra.

55 En una realización, la muestra se obtiene de un sujeto mamífero. En otra realización, la muestra comprende una mezcla de células que comprenden linfocitos T y/o linfocitos B y células que no son linfocitos T ni/o linfocitos B.

60 En una realización, el método incluye moldes sintéticos que comprenden la secuencia, 5'-U1-B1-V-B2-J-B3-U2-3'. En una realización, V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica codificante de una región variable (V) de TCR o Ig, o su complemento. En una realización, cada molde sintético comprende una secuencia oligonucleotídica de la región V única. En una realización, J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica codificante de una región (J) de unión de TCR o Ig, o su complemento. En una realización, U1 comprende una secuencia oligonucleotídica que es una primera secuencia adaptadora universal o una primera secuencia oligonucleotídica de plataforma de secuenciación que está unida y posicionada 5' con respecto a una primera secuencia oligonucleotídica adaptadora universal. En una realización adicional, U2 comprende una secuencia oligonucleotídica que es una secuencia adaptadora universal o una segunda secuencia oligonucleotídica de plataforma de secuenciación que está unida y posicionada 5' con respecto a

- una primera secuencia oligonucleotídica adaptadora universal. En una realización, B1, B2 y B3 son cada uno independientemente, ya sea nada o una secuencia de código de barras oligonucleotídica de 3-25 ácidos nucleicos que identifica de manera única como una combinación de pares una secuencia oligonucleotídica única de la región V y un oligonucleótido único de la región J. En una realización, al menos un B1, B2 y B3 está presente en cada molde sintético. En una realización, al menos dos de entre B1, B2 y B3 están presentes en cada molde sintético. En una realización, los tres B1, B2 y B3 están presentes en cada molde sintético. En una realización, los moldes sintéticos comprenden además un número de oligonucleótidos aleatorios (aleatorizadores). En una realización, la cadena de oligonucleótidos aleatorios comprende de aproximadamente 4 a aproximadamente 50 nucleótidos. En una realización, el tramo aleatorio de oligonucleótidos comprende aproximadamente 8 oligonucleótidos.
- En una realización, el número total de moldes sintéticos del primer conjunto de moldes sintéticos sometido a amplificación se usa para la determinación, usando una dilución limitante de dichos moldes sintéticos, comprendiendo cada uno una región única V y J o C de TCR o Ig, de modo que el número de moldes sintéticos únicos observados permite la inferencia del número total de moléculas molde sintéticas de entrada. En una realización, cada molde sintético se encuentra solo en una sola copia.
- En una realización, el número total de moldes sintéticos del primer conjunto de moldes sintéticos sometidos a amplificación se determina contando el número de moldes sintéticos únicos basándose en nucleótidos aleatorios únicos contenidos en cada molde sintético.
- En una realización, el método proporciona la amplificación de dos o más regiones de control genómico. En otra realización, el método proporciona la amplificación de tres o más regiones de control genómico. En otra realización más, el método proporciona la amplificación de cuatro o más regiones de control genómico. Aún en otra realización, el método proporciona la amplificación de cinco o más regiones de control genómico. En una realización, el método proporciona la amplificación de cinco regiones de control genómico y el cálculo de factores de amplificación para cada una. En una realización, el factor de amplificación medio se determina tomando la media de los factores de amplificación para cada región de control genómico. En una realización, el factor de amplificación de la región de control genómico más alto y más bajo se descarta antes de tomar una media. En una realización, las regiones de control genómico son una o más de entre ACTB, B2M, Clorf34, CHMP2A, GPI, GUSB, HMBS, HPRT1, PSMB4, RPL13A, RPLP0, SDHA, SNRPD#, UBC, VCP, VPS29, PPIA, PSMB2, RAB7A, UBC, VCP, REEP5 y EMC7. En una realización, las regiones de control genómico son PSMB2, RAB7A, PPIA, REEP5 y EMC7.
- En una realización, la PCR multiplexada y la secuenciación de los locus de TCR o Ig reordenados y los primeros moldes sintéticos se realizan en una reacción de PCR multiplexada, mientras que la amplificación de las regiones de control genómico y del segundo conjunto de moldes sintéticos se realizan en una segunda reacción de PCR multiplexada. En otra realización, los locus de TCR y/o de Ig reordenados, el primer conjunto de moldes sintéticos, las regiones de control genómico y el segundo conjunto de moldes sintéticos se amplifican y secuencian en la misma reacción de PCR multiplexada.
- En una realización, la amplificación de los locus de TCR o Ig reordenados y del primer conjunto de moldes sintéticos se realiza usando una pluralidad de cebadores oligonucleotídicos. En una realización, los cebadores oligonucleotídicos comprenden una pluralidad de cebadores oligonucleotídicos del segmento V que son cada uno independientemente capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región V de TCR o Ig, o con su complemento. En una realización, 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 codificante de una región V de TCR o Ig. En una realización, la pluralidad de cebadores del segmento V se hibrida específicamente con esencialmente todos los segmentos génicos funcionales codificantes de la región V de TCR o Ig que están presentes en la composición. En una realización, la pluralidad de cebadores incluye además una pluralidad de cebadores oligonucleotídicos del segmento J que son cada uno independientemente capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región J de TCR o Ig, o con su complemento. En una realización, 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 codificante de la región J de TCR o Ig. En una realización, la pluralidad de cebadores del segmento J se hibrida específicamente con esencialmente todos los segmentos génicos codificantes de la región J de TCR o Ig funcionales que están presentes en la composición.
- En una realización, la pluralidad de cebadores oligonucleotídicos del segmento V y dicha pluralidad de cebadores oligonucleotídicos del segmento J comprenden las secuencias expuestas en SEQ ID NO: 1-764. En una realización, la pluralidad de cebadores oligonucleotídicos del segmento V comprende secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos establecidas en SEQ ID NO: 1-120, 147-158, 167-276, 407-578, 593-740 y/o la pluralidad de cebadores oligonucleotídicos del segmento J comprenden secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos expuestas en SEQ ID NO: 121-146, 159-166, 277-406, 579-592 o 741-764.
- En una realización, la muestra es tejido recién extraído, congelado o fijo. En una realización, la muestra comprende células humanas, células de ratón o células de rata. En una realización, la muestra comprende tejido somático.

En una realización, la muestra es biopsia tumoral. En una realización, el segmento V de TCR comprende un segmento V δ de TCR, un segmento V γ de TCR, un segmento V α de TCR o un segmento V β de TCR. En una realización, el segmento J de TCR comprende un segmento J δ de TCR, un segmento J γ de TCR, un segmento J α de TCR o un segmento J β de TCR. En una realización, el segmento V de Ig comprende un segmento del gen V de IGH, un segmento del gen V de IGL o un segmento del gen V de IGK. En una realización, el segmento de la región J de Ig comprende un segmento del gen J de IGH, un segmento del gen J de IGL o un segmento del gen V de IGK.

En una realización, las secuencias de salida biológica para los locus de TCR o Ig y los moldes sintéticos contenidos en el primer conjunto de moldes sintéticos tienen una longitud de aproximadamente 100-300 nucleótidos. En otra realización, las secuencias de salida para cada región de control genómico y los moldes sintéticos contenidos en el segundo conjunto de moldes sintéticos tienen una longitud de aproximadamente 100-300 nucleótidos. Aún en otra realización, las secuencias de salida biológica para los locus de TCR o Ig, los moldes sintéticos contenidos en el primer conjunto de moldes sintéticos, las secuencias de salida para cada región de control genómico y los moldes sintéticos contenidos en el segundo conjunto de moldes sintéticos tienen una longitud de aproximadamente 100-300 nucleótidos.

En una realización, se determina un factor de amplificación para (i) una pluralidad de moléculas biológicas de ácido nucleico reordenadas codificantes de un receptor inmunitario adaptativo que comprende un receptor de linfocitos T (TCR) o inmunoglobulina (Ig) de dicha muestra biológica, cada molécula biológica de ácido nucleico reordenada que comprende un segmento génico codificante de una región variable (V) única y un segmento génico codificante de una región (J) de unión única, e (ii) una pluralidad de moléculas oligonucleotídicas molde sintéticas, cada una comprendiendo una combinación por parejas de un segmento génico de la región V única y un segmento génico de la región J única que se encuentra en una de la pluralidad de moléculas biológicas de ácido nucleico reordenadas.

En una realización adicional, la determinación del factor de amplificación incluye comparar (1) una serie de secuencias oligonucleotídicas molde sintéticas de salida obtenidas de la secuenciación de moléculas oligonucleotídicas molde sintéticas amplificadas generadas a partir de dicha PCR multiplexada con (2) una serie de moléculas oligonucleotídicas molde sintéticas de entrada añadidas a dicha PCR multiplexada, donde dichas secuencias oligonucleotídicas molde sintéticas comprenden una secuencia de fórmula I: 5'-U1-B1-V-B2-J-B3-U2-3' (I).

Dentro de la secuencia oligonucleotídica molde sintética de fórmula I: V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica codificante de la región variable (V) del receptor inmunitario adaptativo, o su complemento, y teniendo cada una de dicha pluralidad de secuencias oligonucleotídicas molde sintéticas una secuencia oligonucleotídica de la región V única, J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica codificante de la región de unión (J) al receptor inmunitario adaptativo, o su complemento, y teniendo cada una de dicha pluralidad de secuencias oligonucleotídicas molde sintéticas una secuencia oligonucleotídica de la región J única, U1 comprende una secuencia oligonucleotídica que se selecciona entre (i) una primera secuencia oligonucleotídica adaptadora universal e (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que está unida y posicionada 5' con respecto a una primera secuencia oligonucleotídica adaptadora universal, U2 comprende una secuencia oligonucleotídica que se selecciona entre (i) una segunda secuencia oligonucleotídica adaptadora universal e (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que está unida y posicionada 5' con respecto a una segunda secuencia oligonucleotídica adaptadora universal, y al menos uno de entre B1, B2 y B3 está presente, y cada uno de B1, B2 y B3 comprenden un oligonucleótido que comprende una secuencia de código de barras de 3-25 nucleótidos contiguos que identifica de forma única, como una combinación por parejas, (i) dicha secuencia oligonucleotídica de la región V única de (a) e (ii) dicha secuencia oligonucleotídica de la región J única de (b).

En una realización adicional, un número total de moléculas de ácido nucleico biológicas reordenadas de entrada se determina comparando el número de secuencias de salida de moléculas de ácido nucleico biológicas reordenadas obtenidas de la secuenciación de moléculas de ácido nucleico biológicas reordenadas amplificadas producidas a partir de dicha PCR multiplexada con dicho factor de amplificación. En otra realización más, la representación relativa de las células inmunitarias adaptativas (en una muestra biológica que comprende una mezcla de células que comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas) se determina comparando dicho número de moléculas de ácido nucleico biológicas reordenadas de entrada con dicho número de moléculas de ácido nucleico biológicas de entrada totales.

En algunas realizaciones, la determinación de dicho factor de amplificación comprende dividir (1) dicho número de secuencias oligonucleotídicas molde sintéticas de salida obtenidas de la secuenciación de moléculas oligonucleotídicas molde sintéticas amplificadas generadas a partir de la PCR multiplexada entre (2) dicho número de oligonucleótidos molde sintéticos de entrada añadidos a dicha PCR multiplexada. En otras realizaciones, la determinación de una serie de moléculas de ácido nucleico biológicas reordenadas de entrada comprende dividir (1) un número total de secuencias de salida de moléculas de ácido nucleico biológicas reordenadas obtenidas de la secuenciación de moléculas de ácido nucleico biológicas reordenadas amplificadas producidas a partir de dicha PCR

multiplexada entre (2) dicho factor de amplificación. En aún otras realizaciones, la comparación de dicho número de moléculas de ácido nucleico biológicas reordenadas de entrada con dicho número de moléculas de ácido nucleico biológicas de entrada totales comprende dividir el número de moléculas de ácido nucleico biológicas reordenadas de entrada entre dicho número de moléculas de ácido nucleico biológicas de entrada totales.

5 En una realización, dicho número de oligonucleótidos molde sintéticos de entrada añadidos en dicha PCR multiplexada se determina mediante la amplificación de un grupo de oligonucleótidos molde sintéticos sin diluir usando PCR simple para obtener una pluralidad de amplicones molde sintéticos, la secuenciación de dicha pluralidad de amplicones molde sintéticos para determinar una frecuencia de cada oligonucleótido molde sintético
10 único en el conjunto, la cuantificación de una relación basada en simulaciones *in silico* de dicha frecuencia de cada oligonucleótido molde sintético único en el conjunto, entre un número total de secuencias oligonucleotídicas molde sintéticas únicas observadas en un subconjunto del conjunto y el número de oligonucleótidos molde sintéticos totales presentes en dicho subconjunto, y la determinación de una serie de oligonucleótidos molde sintéticos totales de entrada en dicha PCR multiplexada, incluyendo dicha PCR multiplexada una dilución limitante de dicho grupo de
15 oligonucleótidos molde sintéticos, dicha determinación basada en el número de oligonucleótidos molde sintéticos únicos observados en la salida de la secuenciación de dicha PCR simple y en dicha relación cuantificada. En una realización adicional, dicho número de oligonucleótidos molde sintéticos de entrada añadidos en dicha PCR multiplexada se determina además mediante la adición de una cantidad conocida de dicho grupo de oligonucleótidos molde sintéticos diluidos a dicha PCR multiplexada para producir una serie de oligonucleótidos molde sintéticos
20 totales amplificados.

En una realización, dicha PCR multiplexada se realiza usando una pluralidad de conjuntos de cebadores oligonucleotídicos que comprenden: (a) una pluralidad de cebadores oligonucleotídicos del segmento V que son independientemente capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región V del receptor inmunitario adaptativo o con su complemento, donde cada cebador del
25 segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria a al menos un segmento génico codificante de la región V del receptor inmunitario adaptativo funcional y donde dicha pluralidad de cebadores del segmento V se hibrida específicamente con esencialmente todos los segmentos génicos codificantes de la región V del receptor inmunitario adaptativo funcionales que están presentes en la composición; y (b) una pluralidad de cebadores oligonucleotídicos del segmento J que son independientemente
30 capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región J del receptor inmunitario adaptativo o con su complemento, 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 codificante de la región J del receptor inmunitario adaptativo funcional y donde dicha pluralidad de cebadores del segmento J se hibrida específicamente con esencialmente todos los segmentos génicos codificantes de la región
35 J del receptor inmunitario adaptativo funcionales que están presentes en la composición, de modo que dicha pluralidad de cebadores oligonucleotídicos del segmento V y del segmento J son capaces de amplificar en dicha PCR multiplexada: (i) esencialmente todos los oligonucleótidos molde sintéticos para producir una pluralidad de moléculas oligonucleotídicas molde sintéticas amplificadas, e (ii) esencialmente todas las moléculas biológicas de
40 ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos en dicha muestra biológica para producir una pluralidad de moléculas biológicas amplificadas de ácido nucleico reordenadas, dicha pluralidad de moléculas de ácido nucleico biológicas reordenadas amplificadas es suficiente para cuantificar la diversidad de dichas moléculas de ácido nucleico reordenadas a partir de dicha muestra biológica. En una realización adicional, dicha pluralidad de cebadores oligonucleotídicos del segmento V y dicha pluralidad de cebadores oligonucleotídicos del segmento J comprenden las secuencias expuestas en SEQ ID NO: 1-764.
45

En otra realización, cualquiera de las dos entre: (i) dicha pluralidad de cebadores oligonucleotídicos del segmento V comprende secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos establecidas en SEQ ID NO: 1-120, 147-158, 167-276, 407-578 o 593-740, e (ii) dicha pluralidad de cebadores oligonucleotídicos del segmento J comprenden secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos establecidas en SEQ ID NO: 121-146, 159-166, 277-406, 579-592 o 741-764. En algunas realizaciones, dicha pluralidad de moléculas oligonucleotídicas molde sintéticas comprende una serie de al menos *a* o al menos *b* secuencias oligonucleotídicas únicas, lo que sea mayor, donde *a* es el número de segmentos
50 segmentos génicos codificantes de la región V del receptor inmunitario adaptativo único en el sujeto y *b* es el número de segmentos génicos codificantes de la región J del receptor inmunitario adaptativo único en el sujeto. En una realización adicional, *a* varía de 1 a varios segmentos máximos del gen V en el genoma de dicho sujeto mamífero. En una realización adicional, *b* varía de 1 a varios segmentos máximos del gen J en el genoma de dicho sujeto mamífero. En otras realizaciones, dicha pluralidad de moléculas oligonucleotídicas molde sintéticas comprende al menos una secuencia oligonucleotídica molde sintética para cada secuencia oligonucleotídica de la región V única y
60 al menos una secuencia oligonucleotídica molde sintética para cada secuencia oligonucleotídica de la región J única. En algunas realizaciones, dichas células inmunitarias adaptativas son linfocitos T o linfocitos B. En otras realizaciones, dicha muestra biológica es tejido recién extraído, tejido congelado o tejido fijo, y dicha muestra biológica comprende células humanas, células de ratón o células de rata. En realizaciones adicionales, dicha muestra biológica comprende tejido somático.
65

En una realización, dicho segmento génico codificante de la región V comprende un segmento V δ de TCR, un

- segmento V γ de TCR, un segmento V α de TCR o un segmento V β de TCR. En otra realización, dicho segmento génico codificante de la región J comprende un segmento J δ de TCR, un segmento J γ de TCR, un segmento J α de TCR o un segmento J β de TCR. En algunas realizaciones, dicho segmento génico codificante de la región V comprende un segmento génico V de IGH, un segmento del gen V de IGL o un segmento del gen V de IGK. En otras realizaciones, dicho segmento génico codificante de la región J comprende un segmento génico J de IGH, un segmento del gen J de IGL o un segmento del gen V de IGK. En algunas realizaciones, dicha pluralidad de secuencias oligonucleotídicas molde sintéticas comprende secuencias seleccionadas entre SEQ ID NO: 787-3003. En otras realizaciones, V de fórmula (I) es una secuencia oligonucleotídica que comprende al menos 30, 60, 90, 120, 150, 180 o 210, o no más de 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia génica codificante de la región V del receptor inmunitario adaptativo, o su complemento. En otras realizaciones, J de fórmula (I) es una secuencia oligonucleotídica que comprende al menos 16-30, 31-60, 61-90, 91-120 o 120-150, o no más de 500, 400, 300 o 200 nucleótidos contiguos de una secuencia génica codificante de la región J del receptor inmunitario adaptativo, o su complemento.
- 15 En algunas realizaciones, J de fórmula (I) comprende una secuencia que comprende una región constante de la secuencia génica codificante de la región J. En otras realizaciones, cada secuencia oligonucleotídica molde sintética tiene menos de 1.000, 900, 800, 700, 600, 500, 400, 300 o 200 nucleótidos de longitud.

También se divulgan en el presente documento kits que comprenden reactivos que comprenden una composición que comprende una pluralidad de oligonucleótidos molde sintéticos y un conjunto de cebadores oligonucleotídicos como se ha descrito anteriormente, e instrucciones para cuantificar una representación relativa de células inmunitarias adaptativas de una muestra biológica que comprende una mezcla de células que comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas, mediante la cuantificación de: (i) un número de productos molde sintéticos de moléculas oligonucleotídicas molde sintéticas amplificadas, e (ii) un número de productos biológicos reordenados de una serie de secuencias de salida.

Breves descripciones de los dibujos

Estas y otras características, aspectos y ventajas de la presente invención se entenderán mejor con respecto a la siguiente descripción, y dibujos adjuntos, donde:

- las Figuras 1A y 1B ilustran ejemplos de oligonucleótidos molde sintéticos, de acuerdo con una realización de la invención. Se muestran cebadores con colas de secuenciación Illumina (ILMN), específicos de las secuencias adaptadoras universales (anteriores) o las secuencias del gen V y del gen J (más adelante).
- La Figura 2 es un gráfico que muestra la relación entre el número de oligonucleótidos molde sintéticos únicos y el número total de oligonucleótidos molde sintéticos de una muestra.
- La Figura 3 es un gráfico que compara un número conocido de genomas de linfocitos B de una muestra con el número estimado de genomas de linfocitos B de la muestra usando oligonucleótidos molde sintéticos.
- La Figura 4 es un gráfico que muestra cálculos de representaciones relativas de linfocitos T del total de células de una muestra, ya sea a través de métodos ddPCR o usando oligonucleótidos molde sintéticos.
- La Figura 5 ilustra un ejemplo de un oligonucleótido molde sintético de control y su relación con un locus genómico representativo, de acuerdo con una realización de la invención, y la relación del oligonucleótido molde sintético de control con el locus genómico que representa.
- La Figura 6 ilustra que los métodos de la presente invención que utilizan regiones de control genómico son capaces de calcular con precisión el número de genomas de entrada y el número de linfocitos T en función del número de secuencias de entrada.

Descripción detallada de la invención

50 Descripción general

Se proporcionan métodos y composiciones para determinar el número absoluto de genomas de entrada de células inmunitarias adaptativas en una mezcla compleja de células. Además, la presente divulgación se refiere a métodos de determinación cuantitativa de la presencia de linfocitos en tejidos complejos, tales como tejidos sólidos. Los métodos de la invención también incluyen una cuantificación de la representación relativa de genomas de linfocitos infiltrantes tumorales (TIL) como una proporción relativa de todos los genomas celulares que están representados en una muestra, tal como una muestra de tejido sólido o de tumor sólido, o la cuantificación de los genomas de linfocitos que se han infiltrado en el tejido somático en la patogénesis de la inflamación, alergia o enfermedad autoinmunitaria o en órganos trasplantados como una proporción relativa de todos los genomas celulares que están representados en una muestra de ADN tisular.

En algunas realizaciones, el método incluye determinar la cuantificación absoluta exacta del número de células inmunitarias adaptativas de una muestra biológica. Esto implica determinar el número de genomas de entrada de las células inmunitarias adaptativas de la muestra.

Además, en el presente documento, se proporcionan composiciones y métodos que son útiles para la cuantificación

y determinación de manera fiable de las secuencias de poblaciones grandes y estructuralmente diversas de genes reordenados codificantes de receptores inmunitarios adaptativos, tales como inmunoglobulinas (Ig) y/o receptores de linfocitos T (TCR). Estos genes reordenados pueden estar presentes en una muestra biológica que contiene ADN de células linfoides de un sujeto o de una fuente biológica, incluyendo un sujeto humano y/o transcripciones de ARNm de estos genes reordenados pueden estar presentes en dicha muestra y usarse como moldes para la síntesis de ADNc mediante transcripción inversa.

Se proporcionan métodos de cuantificación de una cantidad de oligonucleótidos molde sintéticos de una muestra para determinar un número total de genomas de entrada de células inmunitarias adaptativas de una muestra biológica. En una realización, se usa una muestra de oligonucleótidos molde sintéticos para determinar una proporción del número de moléculas oligonucleotídicas molde sintéticas de entrada en comparación con el número de oligonucleótidos molde sintéticos de salida totales (amplicones). Se añade una dilución limitante de esta muestra a una muestra biológica (al comienzo de un ensayo de PCR multiplexada) y se usa para determinar el número total de genomas de entrada de células inmunitarias adaptativas de la muestra biológica. En determinadas realizaciones, los moldes sintéticos de la muestra comprenden un tramo de ácidos nucleicos aleatorios, por ejemplo, un aleatorizador de 8 nucleótidos. Por tanto, se pueden hacer diluciones limitantes de modo que cada molde sintético de la muestra esté presente solo una vez y pueda ser identificado por el aleatorizador de 8 nucleótidos contenido en la misma.

El método también incluye determinar la representación relativa de las células inmunitarias adaptativas de una muestra que contiene una mezcla de células, donde la mezcla comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas. En determinadas realizaciones, se puede cuantificar una representación relativa de ADN de células inmunitarias adaptativas (por ejemplo, linfocitos T y/o B que tienen genes de receptores inmunitarios adaptativos reordenados, incluyendo linfocitos del linaje T y B de diferentes etapas de maduración, tales como precursores, blastocitos, progenie o similar) entre el ADN total de una muestra de tipos de células mixtos. Por ejemplo, ciertas realizaciones permiten la determinación, en ADN extraído de una muestra biológica, de la representación relativa de ADN de linfocitos infiltrantes tumorales (TIL) en el ADN de la muestra biológica, donde la muestra comprende todo o una parte de un tumor que contiene células inmunitarias adaptativas y células que no son células inmunitarias adaptativas (incluyendo las células tumorales). Otras ciertas realizaciones, por ejemplo, permitir la determinación, en ADN extraído de una muestra biológica, de la representación relativa del ADN de los linfocitos infiltrantes en el ADN de la muestra biológica, donde la muestra comprende todo o una parte de un tejido somático que contiene células inmunitarias adaptativas y células que no son células inmunitarias adaptativas, tales como células de un tejido sólido. En los documentos US2013/0288237 y WO2013/059725, se divulgan métodos alternativos de cuantificación la representación relativa de las células inmunitarias adaptativas en una mezcla de células.

Puede que no todas las células de la mezcla de células sean células inmunitarias adaptativas, y se obtienen ciertas ventajas imprevistas de las realizaciones descritas en el presente documento donde las células de la mezcla de células no necesitan ser todas células inmunitarias adaptativas. Como se describe en el presente documento, se proporcionan composiciones y métodos de cuantificación de la proporción de genomas celulares en una muestra que comprende moléculas de ácido nucleico (por ejemplo, ADN) que son aportadas por células inmunitarias adaptativas en relación con el número total de genomas celulares de la muestra, a partir de una muestra de ADN que se extrajo de una mezcla de tipos celulares, tales como un tumor sólido o un tejido sólido.

En determinadas realizaciones, las moléculas de ácido nucleico del receptor inmunitario adaptativo reordenadas se amplifican en una única PCR multiplexada usando conjuntos de cebadores oligonucleotídicos específicos del receptor inmunitario adaptativo reordenados para producir secuencias de ADN específicas de las células inmunitarias adaptativas, que se usan para determinar la contribución relativa de las células inmunitarias adaptativas en comparación con el ADN total extraído de una muestra de tipos de células mixtos. En otras realizaciones, las moléculas de ARNm de células inmunitarias adaptativas reordenadas se amplifican usando rt-qPCR y conjuntos de cebadores oligonucleotídicos específicos de receptores inmunitarios adaptativos reordenados para cuantificar las señales de ADNc de receptores inmunitarios adaptativos reordenados y para determinar la contribución relativa de las células inmunitarias adaptativas al número total de genomas extraídos de una muestra de tipos de células mixtas. Los métodos de uso de la qPCR para determinar la representación relativa de células inmunitarias adaptativas en una mezcla de células se divulgan en los documentos US2013/0288237 y WO2013/059725.

Además, en otras realizaciones, donde la muestra incluye moléculas de ARNm, los métodos de la invención incluyen el uso de un ensayo cuantitativo en tiempo real de la reacción en cadena de la polimerasa (qPCR) con conjuntos de cebadores oligonucleotídicos que amplifican específicamente esencialmente todos los genes del receptor inmunitario adaptativo reordenados (por ejemplo, partes que contienen polinucleótidos codificantes de CDR3 de los genes de receptores de linfocitos T reordenados y/o genes de inmunoglobulina) que pueden estar presentes en una muestra, para generar una primera señal de ADN detectable que refleje cuantitativamente la producción de una multiplicidad de moléculas de ADN codificantes del receptor inmunitario adaptativo reordenadas amplificadas. En determinadas realizaciones, la amplificación por qPCR se puede controlar en uno o en una pluralidad de puntos de tiempo durante el curso de la reacción de qPCR, es decir, en "tiempo real". El control en tiempo real permite la determinación de la cantidad de ADN que se está generando mediante la comparación de una señal cuantificadora del ADN codificante

del receptor inmunitario adaptativo así medida con una señal cuantificadora apropiada de los moldes sintéticos (o ADN molde de control), que puede usarse como patrón de calibración. Los métodos para la cuantificación usando qPCR se describen en detalle en los documentos US2013/0288237 y WO2013/059725.

5 Más adelante, se divulgan enfoques inesperadamente ventajosos para determinar la representación relativa de células inmunitarias adaptativas en una muestra biológica usando PCR multiplexada para generar una población de moléculas de ADN amplificadas a partir de una muestra biológica que contiene genes reordenados codificantes de receptores inmunitarios adaptativos, antes de la secuenciación cuantitativa de alto rendimiento de dichos productos amplificados. Se describe la amplificación multiplexada y la secuenciación de alto rendimiento de las secuencias de
 10 ADN codificantes de TCR y BCR (Ig) reordenadas, por ejemplo, en 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 EE.UU. n.º 2012/0058902, publicación de EE.UU. n.º 2010/0330571, documento WO/2010/151416, documento WO/2011/106738, documento WO2012/027503, documento WO/2013/169957, documento WO/2013/188831 y documento US2013/253842 y documento
 15 WO2013/086450.

Además, en el presente documento, se describen composiciones y métodos para el uso de oligonucleótidos molde sintéticos que están destinados a ser incluidos directamente en las reacciones de amplificación y secuenciación de una muestra, y cuya cantidad en la reacción (el número de moléculas) puede medirse con precisión para mejorar la
 20 exactitud de la corrección del sesgo de la amplificación por PCR multiplexada y la cuantificación absoluta de los moldes de entrada. El sesgo de la amplificación se describe en más detalle en el documento WO/2013/169957, y en Carlson, C. S. *et al.* Usando moldes sintéticos para diseñar un ensayo de PCR multiplexada imparcial, *Nature Communications* 4, 2680, doi: 10.1038/ncomms3680 (2013).

25 La presente invención se dirige, en ciertas realizaciones, como se describe en el presente documento, a la cuantificación del ADN de células inmunitarias adaptativas que están presentes en tejidos sólidos, y en realizaciones particulares, a tumores sólidos, de modo que se pueda determinar la presencia relativa de células inmunitarias adaptativas como una proporción de todos los tipos de células que pueden estar presentes en el tejido (por ejemplo, tumor). Estas y otras realizaciones relacionadas son, en parte, el resultado de ciertas ventajas sorprendentes y no
 30 reconocidas hasta la fecha, que se divulgan con mayor detalle a continuación, que derivan de la exquisita sensibilidad que se brinda, para la detección de células inmunitarias adaptativas, mediante el diseño de la PCR multiplexada usando los conjuntos de cebadores de oligonucleótidos descritos en el presente documento. Estos conjuntos de cebadores oligonucleotídicos permiten la producción de moléculas de ADN reordenadas amplificadas y moléculas molde sintéticas codificantes de partes de receptores inmunitarios adaptativos. Estas realizaciones y realizaciones relacionadas presentan la selección de una pluralidad de cebadores oligonucleotídicos que se hibridan específicamente con secuencias de polinucleótidos codificantes del polipéptido de la región V y secuencias de polinucleótidos codificantes del polipéptido de la región J del receptor inmunitario adaptativo (por ejemplo, el receptor de linfocitos T, TCR; o inmunoglobulina, Ig). En el presente documento, se describen cebadores universales que son
 35 específicos de secuencias adaptadoras universales y que se unen a amplicones que comprenden secuencias adaptadoras universales. Los cebadores potencian la amplificación por PCR de moléculas de ácido nucleico, tales como ADN, que incluyen esencialmente todas las regiones de genes reordenadas codificantes de la CDR3 de TCR o codificantes de la CDR3 de Ig que pueden estar presentes en una muestra biológica de prueba, donde la muestra contiene una mezcla de células que comprende células inmunitarias adaptativas (por ejemplo, células del linaje de linfocitos T y B) y células que no son células inmunitarias adaptativas. Por ejemplo, se puede obtener una mezcla de
 40 células de un tumor sólido que comprende células tumorales y TIL.

Receptores de células inmunitarias adaptativas

El TCR nativo es una proteína de superficie celular heterodimérica de la superfamilia de inmunoglobulinas, que está asociada con proteínas invariantes del complejo CD3 que participa en la transducción de señales mediadoras. Los
 50 TCR existen en formas $\alpha\beta$ y $\gamma\delta$, que son estructuralmente similares, pero tienen ubicaciones anatómicas, y probablemente funciones, bastante distintas. Los ligandos de MHC de clase I y clase II, que se unen al TCR, también son proteínas de la superfamilia de las inmunoglobulinas, pero están especializadas para la presentación de antígenos, con un sitio de unión a péptidos altamente polimórfico que les permite presentar una amplia selección de fragmentos de péptidos cortos en la superficie celular de las APC.

Las partes 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 constante y variable incluye un enlace disulfuro intracatenario. Los dominios variables contienen los
 60 bucles altamente polimórficos análogos a las regiones determinantes de complementariedad (CDR) de los anticuerpos. La CDR3 de los TCR $\alpha\beta$ interactúan con el péptido presentado por el MHC, y las CDR 1 y 2 de los TCR $\alpha\beta$ interactúan con el péptido y el MHC. La diversidad de secuencias de TCR se genera a través de la reordenación somática de los genes unidos de la región variable (V), región de diversidad (D), región de unión (J) y región constante.

65 Los locus de genes de Ig y TCR contienen muchos segmentos génicos diferentes de la región variable (V), de

diversidad (D) y de unión (J), que están sujetos a procesos de reordenación durante la diferenciación linfóide temprana. Las secuencias de segmentos de genes V, D y J de TCR e Ig son conocidas en la técnica y están disponibles en bases de datos públicas tales como GENBANK. Las reordenaciones V-D-J están mediadas por un complejo de enzima recombinasa en el que las proteínas RAG1 y RAG2 desempeñan un papel clave al reconocer y cortar el ADN en las secuencias de señal de recombinación (RSS, *Recombination Signal Sequences*). Las RSS están ubicadas en dirección 3' de los segmentos del gen V, a ambos lados de los segmentos del gen D, y en dirección 5' de los segmentos del gen J. Las RSS inapropiadas reducen o incluso evitan completamente la reordenación. La RSS consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3', y nonámero, 5'-ACAAAACC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Se ha identificado una serie de posiciones de nucleótidos como importantes para la recombinación, incluyendo el dinucleótido CA en la posición uno y dos del heptámero, y también se ha mostrado que se prefiere enormemente una C en la posición tres del heptámero, así como un nucleótido A en las posiciones 5, 6, 7 del nonámero. (Ramsden *et al.* 1994 *Nucl. Ac. 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 no uniformes. El espaciador, aunque más variable, también tiene un impacto en la recombinación, y se ha demostrado que los reemplazos de un solo nucleótido tienen un impacto significativo en la eficacia de la recombinación (Fanning *et al.* 1996 *Cell. Immunol. Immunopath.* 79:1; Larijani *et al.* 1999 *Nucl. Ac. Res.* 27:2304; Nadel *et al.* 1998 *J. Immunol.* 161:6068; Nadel *et al.* 1998 *J. Exp. Med.* 187:1495). Se han descrito criterios para identificar secuencias de polinucleótidos RSS que tienen eficacias de recombinación significativamente diferentes (Ramsden *et al.* 1994 *Nucl. Ac. 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 reordenación, en general, comienza con una reordenación de D a J, seguida de una reordenación de V a D-J en el caso de los genes de cadena pesada de Ig (IgH), de TCR beta (TCRB) y de TCR delta (TCRD) o las cuestiones dirigidas a las reordenaciones de V a J en el caso de los genes de Ig kappa (IgK), Ig lambda (IgL), TCR alfa (TCRA) y TCR gamma (TCRG). Las secuencias entre los segmentos de genes reordenados, en general, se eliminan en forma de un producto de escisión circular, también denominado círculo de escisión de TCR (TREC) o círculo de escisión del receptor de linfocitos B (BREC).

Las muchas combinaciones diferentes de segmentos de genes V, D y J representan el denominado repertorio combinatorio, que se estima en $\sim 2 \times 10^6$ para las moléculas de Ig, $\sim 3 \times 10^6$ para las moléculas de TCR $\alpha\beta$ y $\sim 5 \times 10^3$ para las moléculas de TCR $\gamma\delta$. En los sitios de unión de los segmentos de genes V, D, y J, se produce la eliminación e inserción aleatoria de nucleótidos durante el proceso de reordenación, dando lugar a regiones de unión altamente diversas, que contribuyen significativamente al repertorio total de moléculas de Ig y TCR, que se estima en $> 10^{12}$.

Los linfocitos B maduros extienden aún más su repertorio de Ig tras el reconocimiento de antígenos en los centros foliculares mediante la 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 solo nucleótido y, a veces, también a inserciones o eliminaciones de nucleótidos. Los genes de Ig mutados somáticamente también se encuentran en tumores malignos maduros de linfocitos B de origen folicular o post-folicular.

Definiciones

Como se usa en el presente documento, el término "gen" se refiere a un segmento de ADN que puede expresarse como una cadena polipeptídica. La cadena polipeptídica puede ser toda o una parte de un polipéptido TCR o Ig (por ejemplo, un polipéptido que contiene CDR3). El gen puede incluir regiones que preceden y siguen a la región de codificación ("líder y remolque"), secuencias intermedias (intrones) entre segmentos de codificación individuales (exones), elementos reguladores (por ejemplo, promotores, potenciadores, sitios de unión a represores y similares), y secuencias de señal de recombinación (RSS), como se describe en el presente documento.

Los "ácidos nucleicos" o "moléculas de ácido nucleico" o "polinucleótidos" u "oligonucleótidos" pueden estar en forma de ácidos ribonucleicos (ARN) o en forma de ácidos desoxirribonucleicos (ADN). Como se cita en el presente documento, ARN incluye ARNm. ADN incluye ADNc, ADN genómico y ADN sintético. El ADN puede ser monocatenario o bicatenario, y en caso de ser monocatenario, puede ser la cadena codificante o la cadena no codificante (antisentido). Una secuencia de codificación codificante de un TCR o una inmunoglobulina, o una de sus regiones (por ejemplo, una región V, un segmento D, una región J, una región C, etc.) puede ser idéntica a la secuencia de codificación conocida en la técnica para cualquier región génica dada o dominio polipeptídico dado de TCR o inmunoglobulina (por ejemplo, dominios de región V, dominios CDR3, etc.). En otras realizaciones, la secuencia de codificación puede ser una secuencia de codificación diferente, que, como resultado de la redundancia o degeneración del código genético, codifique la misma región o el mismo polipéptido de TCR o inmunoglobulina.

El término "cebador", como se usa en el presente documento, se refiere a un oligonucleótido capaz de actuar como un punto de inicio de la síntesis de ADN en condiciones adecuadas. Dichas condiciones incluyen aquellas en las que se induce la síntesis de un producto de extensión de cebador complementario a una cadena de ácido nucleico en presencia de cuatro nucleósidos trifosfatos diferentes y un agente de extensión (por ejemplo, una ADN polimerasa o transcriptasa inversa) en un tampón apropiado y a una temperatura adecuada.

Un cebador es preferentemente un ADN monocatenario. La longitud apropiada de un cebador depende del uso previsto del cebador, pero normalmente varía de 6 a 50 nucleótidos, o en ciertas realizaciones, de 15 a 35 nucleótidos. Las moléculas de cebador cortas, en general, 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, pero debe ser lo suficientemente complementario para hibridarse 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 el presente documento.

Como se describe en el presente documento, los cebadores pueden tener incorporadas características adicionales que permitan la detección o inmovilización del cebador, pero que no alteren la propiedad básica del cebador, la de actuar como un punto de inicio de la síntesis de ADN. Por ejemplo, los cebadores pueden contener una secuencia de ácido nucleico adicional en el extremo 5' que no se hibride con el ácido nucleico diana, pero que facilite la clonación, detección o secuenciación del producto amplificado. La región del cebador que es suficientemente complementaria al molde para hibridarse se denomina en el presente documento región de hibridación.

Como se usa en el presente documento, un cebador es "específico" para una secuencia diana si, cuando se usa en una reacción de amplificación en condiciones suficientemente rigurosas, el cebador se hibrida principalmente con el ácido nucleico diana. Por lo general, un cebador es específico de una secuencia diana si la estabilidad del dúplex cebador-diana es superior a la estabilidad de un dúplex formado entre el cebador y cualquier otra secuencia encontrada en la muestra. Un experto en la materia reconocerá que diferentes factores, tales como las condiciones de salinidad, así como la composición de bases del cebador y la ubicación de los emparejamientos erróneos, afectarán a la especificidad del cebador, y esa confirmación experimental rutinaria de la especificidad del cebador será necesaria en muchos casos. Se pueden escoger condiciones de hibridación en las que el cebador pueda formar dúplex estables solo con una secuencia diana. Por lo tanto, 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.

El término "mejorar" se refiere a cualquier resultado terapéuticamente beneficioso en el tratamiento de un estado patológico, por ejemplo, una fase del cáncer, un estado de enfermedad autoinmunitaria, incluyendo la profilaxis, la disminución de la gravedad o progresión, la remisión o la cura del mismo.

La expresión "*in vivo*" se refiere a procesos que se producen en un organismo vivo.

El término "mamífero", como se usa en el presente documento, incluye tanto seres humanos como no humanos e incluye, aunque sin limitación, seres humanos, primates no humanos, cánidos, felinos, murinos, bovinos, equinos y porcinos.

La expresión porcentaje de "identidad", en el contexto de dos o más secuencias de ácido nucleico o polipeptídicas, se refiere a dos o más secuencias o subsecuencias que tienen un porcentaje específico de nucleótidos o restos de aminoácidos que son iguales, cuando se comparan y se alinean para una correspondencia máxima, medida usando uno de los algoritmos de comparación de secuencias que se describen a continuación (por ejemplo, BLASTP y BLASTN u otros algoritmos disponibles para los expertos) o mediante inspección visual. Dependiendo de la aplicación, el porcentaje de "identidad" puede existir sobre una región de la secuencia que se está comparando, por ejemplo, sobre un dominio funcional o, como alternativa, puede existir sobre la longitud total de las dos secuencias que se van a comparar.

Para comparar secuencias, normalmente, una secuencia actúa como una secuencia de referencia con la que se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, se introducen las secuencias de prueba y de referencia en un ordenador, se designan las coordenadas de la subsecuencia, en caso necesario, y se designan los parámetros del programa del algoritmo de secuencias. El algoritmo de comparación de secuencias entonces calcula el porcentaje de identidad de secuencia para la o las secuencias de prueba respecto a la secuencia de referencia, basándose en los parámetros del programa designados.

La alineación óptima de secuencias para la comparación puede realizarse, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2: 482 (1981), mediante el algoritmo de alineamiento de homología de Needleman y Wunsch, *J. Mol. Biol.* 48: 443 (1970), por la búsqueda del método de similitud de Pearson y Lipman, *Proc. Nat'l. Acad. Sci. EE.UU.* 85:2444 (1988), mediante implementaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, Wis.), o mediante inspección visual (véase, en general, Ausubel *et al.*, *infra*).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). El software para realizar los análisis BLAST está disponible públicamente a través del Centro Nacional de Información Biotecnológica (www.ncbi.nlm.nih.gov/).

Debe destacarse que, como se usa en la memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un", "una" y "el" o "la" incluyen los referentes en plural, salvo que el contexto indique claramente lo contrario.

5

Muestras (tejidos y uso)

Como se usa en el presente documento, una muestra, una muestra de prueba o una muestra biológica de prueba se refieren a tejidos biológicos (por ejemplo, un agregado de células que tienen una estructura y función similares) obtenidos de un sujeto de interés. La muestra puede incluir una mezcla compleja de células inmunitarias adaptativas (por ejemplo, células del linaje de los linfocitos T y B) y células que no son células inmunitarias adaptativas (por ejemplo, células tumorales sólidas).

10

En determinadas realizaciones, una muestra biológica de prueba de interés comprende tejido somático. El tejido somático puede comprender un tejido sólido. En algunas realizaciones, el tejido sólido puede ser un sitio para la patología de la enfermedad autoinmunitaria, tal como un tejido que esté dirigido indebidamente por el sistema inmunitario del hospedador para una respuesta inmunitaria "contra sí mismo". En otras realizaciones determinadas, el tejido somático puede comprender un tejido sólido que sea el sitio de una infección, tal como una infección bacteriana, de levadura, vírica u otra infección microbiana (por ejemplo, una infección por el virus del herpes simple (HSV)). En otras realizaciones, el tejido somático se obtiene de un órgano trasplantado (por ejemplo, un hígado, pulmón, riñón, corazón, bazo, páncreas, piel, intestino y timo).

15

20

Se pueden obtener muestras de tejidos antes de, durante y/o después del tratamiento. Las muestras se pueden usar en el diagnóstico, pronóstico, la vigilancia de la enfermedad, el control de la eficacia terapéutica y otros contextos, proporcionando así información importante, tal como la cuantificación de la representación de las células inmunitarias adaptativas en tejidos complejos que comprenden una mezcla de células. La cuantificación de células inmunitarias adaptativas (por ejemplo, la cuantificación de la representación relativa de células inmunitarias adaptativas en muestras) o la cuantificación del ADN de células inmunitarias adaptativas (por ejemplo, la cuantificación de la representación relativa del ADN de células inmunitarias adaptativas en muestras que contienen ADN de una mezcla de células) en tejidos antes y después y/o durante el curso del tratamiento de un sujeto, puede proporcionar información relevante para el diagnóstico y el pronóstico en pacientes con cáncer, inflamación y/o enfermedad autoinmunitaria, o cualquiera de una serie de otras afecciones que pueden caracterizarse por alteraciones (por ejemplo, aumentos o disminuciones estadísticamente significativos) en la presencia de células inmunitarias adaptativas en uno o más tejidos.

25

30

35

En algunas realizaciones, la muestra se obtiene de un tumor sólido en un sujeto. Se pueden obtener múltiples muestras antes de, durante y/o después de la administración de una pauta terapéutica al sujeto. Se puede obtener una muestra, por ejemplo, mediante la escisión de tejido antes o después del tratamiento de un sujeto.

40

45

50

En otras realizaciones, la muestra que comprende tejido se evalúa o se analiza de acuerdo con otros criterios aceptados en la técnica. Los indicadores del estado (por ejemplo, la evidencia de la presencia o ausencia de patología, o de eficacia de un tratamiento terapéutico administrado previamente o simultáneamente) pueden ser, por ejemplo, compuestos indicadores detectables, nanopartículas, nanoestructuras u otras composiciones que comprenden una molécula indicadora que proporcione una señal detectable indicadora del estado fisiológico de una célula o de un tejido, tal como un colorante vital (por ejemplo, azul de tripano), un indicador colorimétrico del pH, un compuesto fluorescente que puede presentar fluorescencia distinta en función de cualquiera de una serie de parámetros fisiológicos celulares (por ejemplo, pH, Ca^{2+} intracelular u otra concentración de iones fisiológicamente relevante, potencial de membrana mitocondrial, potencial de membrana plasmática, etc., véase Haugland, "The Handbook: A Guide to Fluorescent Probes and Labeling Technologies" (10ª Ed.) 2005, Invitrogen Corp., Carlsbad, CA), un sustrato enzimático, una sonda oligonucleotídica específica, un gen indicador, o similar.

Sujetos y fuente

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 con tejido modificado por ingeniería genética (incluso a través del uso de células madre). En determinadas realizaciones preferidas de la invención, se puede saber que el sujeto o la fuente biológica tienen, o se sospecha que tienen o están en riesgo de tener, un tumor sólido u otra afección maligna, o una enfermedad autoinmunitaria, o una afección inflamatoria, y en ciertas realizaciones preferidas de la invención, se puede saber que el sujeto o la fuente biológica están libres de riesgo o presencia de dicha enfermedad.

55

60

Determinadas realizaciones preferidas contemplan un sujeto o una fuente biológica que es un sujeto humano, tal como un paciente que ha sido diagnosticado de tener o de estar en riesgo de desarrollar o adquirir cáncer de acuerdo con los criterios de diagnóstico clínico aceptados en la técnica, tales como los de US National Cancer Institute (Bethesda, MD, EE.UU.) o como describen DeVita, "Cancer: Principles and Practice of Oncology" de Hellman y Rosenberg (2008, Lippincott, Williams y Wilkins, Filadelfia/ Ovid, Nueva York); Pizzo y Poplack, "Principles

65

and Practice of Pediatric Oncology" (Cuarta edición, 2001, Lippincott, Williams y Wilkins, Filadelfia/ Ovid, Nueva York); y Vogelstein y Kinzler, "The Genetic Basis of Human Cancer" (Segunda edición, 2002, McGraw Hill Professional, Nueva York); determinadas realizaciones contemplan un sujeto humano que se sabe que está libre de un riesgo de tener, desarrollar o adquirir cáncer según dichos criterios.

5 Otras determinadas realizaciones contemplan un sujeto no humano o una fuente biológica, incluyendo, pero sin limitación, un primate no humano, tal como un macaco, un chimpancé, un gorila, un mono verde, un orangután, un babuino u otro primate no humano, incluyendo los sujetos no humanos que pueden ser conocidos en la técnica como modelos preclínicos, que incluyen modelos preclínicos para tumores sólidos y/u otros cánceres. Otras
10 determinadas realizaciones contemplan un sujeto no humano que es un mamífero, por ejemplo, un ratón, una rata, un conejo, un cerdo, una oveja, un caballo, una vaca, una cabra, un jerbo, un hámster, una cobaya u otro mamífero. Muchos de dichos mamíferos pueden ser sujetos conocidos en la técnica como modelos preclínicos para ciertas enfermedades o trastornos, incluyendo tumores sólidos y/u otros cánceres (por ejemplo, Talmadge *et al.*, 2007 *Am. J. Pathol.* 170:793; Kerbel, 2003 *Canc. Biol. Therap.* 2(4 Supl 1):S134; Man *et al.*, 2007 *Canc. Met. Rev.* 26:737; Cespedes *et al.*, 2006 *Clin. Transl. Oncol.* 8:318). No se pretende que el rango de realizaciones sea tan limitado, sin embargo, también se contemplan otras realizaciones en las que el sujeto o la fuente biológica puede ser un
15 vertebrado no mamífero, por ejemplo, otro vertebrado superior, o un ave, un anfibio o especies de reptiles, u otro sujeto o fuente biológica.

20 Las muestras biológicas se pueden proporcionar mediante la obtención de una muestra de sangre, una muestra de biopsia, un explante de tejido, un cultivo de órganos, un fluido biológico o cualquier otro tejido o preparación celular de un sujeto o una fuente biológica. En determinadas realizaciones preferidas, se puede obtener una muestra biológica de prueba de un tejido sólido (por ejemplo, un tumor sólido), por ejemplo, mediante resección quirúrgica, biopsia con aguja u otros medios para obtener una muestra biológica de prueba que contenga una mezcla de
25 células.

Los tejidos sólidos son muy conocidos en las técnicas médicas y pueden incluir cualquier compartimento anatómico espacialmente diferenciado, no fluido definido, que sea esencialmente el producto de una arquitectura multicelular, intercelular, tisular y/u orgánica, tal como un compartimento definido tridimensionalmente que puede comprender o
30 derivar su integridad estructural del tejido conjuntivo asociado y puede estar separado de otras áreas del cuerpo por una membrana delgada (por ejemplo, membrana meníngea, membrana pericárdica, membrana pleural, membrana mucosa, membrana basal, omento, membrana encapsulante de órganos, o similares). Los tejidos sólidos ilustrativos no limitantes pueden incluir cerebro, hígado, pulmón, riñón, próstata, ovario, bazo, ganglio linfático (incluyendo amígdala), piel, tiroides, páncreas, corazón, músculo esquelético, intestino, laringe, esófago y estómago. Las
35 ubicaciones anatómicas, propiedades morfológicas, la caracterización histológica y el acceso invasivo y/o no invasivo a estos y otros tejidos sólidos son bien conocidos por aquellos familiarizados con las técnicas relevantes.

Se considera que los tumores sólidos de cualquier tipo son adecuados para la caracterización de TIL usando las composiciones y los métodos descritos en el presente documento. En determinadas realizaciones preferidas, el
40 tumor sólido puede ser un tumor benigno o un tumor maligno, que además puede ser un tumor primario, un tumor invasivo o un tumor metastásico. Ciertas realizaciones contemplan un tumor sólido que comprende una de entre una célula de cáncer de próstata, una célula de cáncer de mama, una célula de cáncer colorrectal, una célula de cáncer de pulmón, una célula de cáncer de cerebro, una célula de cáncer renal, una célula de cáncer de piel (tal como el carcinoma de células escamosas, carcinoma de células basales o melanoma) y una célula de cáncer de ovario, pero
45 la invención no pretende ser tan limitada, y pueden usarse otros tipos de tumores sólidos y tipos de células cancerosas. Por ejemplo, el tumor puede comprender un cáncer seleccionado entre adenoma, adenocarcinoma, carcinoma de células escamosas, carcinoma de células basales, melanoma (por ejemplo, melanoma maligno), carcinoma microcítico, carcinoma no diferenciado de células grandes, condrosarcoma y fibrosarcoma, o similares. Como también se indica en otra parte del presente documento, se han establecido criterios de diagnóstico clínico
50 aceptados en la técnica para estos y otros tipos de cáncer, tales como los promulgados por el US National Cancer Institute (Bethesda, MD, EE.UU.) o como describen De Vita, "Cancer: Principles and Practice of Oncology" de Hellman y Rosenberg (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); y Vogelstein y Kinzler, "The Genetic Basis of Human Cancer" (Segunda edición, 2002, McGraw Hill Professional, Nueva York). Se describen otros ejemplos no limitantes de tipificación y caracterización de cánceres
55 particulares, por ejemplo, en 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).

Los linfocitos B y los linfocitos T se pueden obtener de una muestra biológica, tal como de una variedad de muestras
60 de tejido y líquido biológico, incluyendo médula ósea, timo, ganglios linfáticos, nódulos linfáticos, tejidos periféricos y sangre, pero se accede más fácilmente a la sangre periférica. Se pueden tomar muestras de cualquier tejido periférico para detectar la presencia de linfocitos B y T y, por tanto, se contempla su uso en los métodos descritos en el presente documento. Los tejidos y fluidos biológicos a partir de los que se pueden obtener células inmunitarias adaptativas incluyen, pero sin limitación, piel, tejidos epiteliales, colon, bazo, una secreción mucosal, mucosa oral,
65 mucosa intestinal, mucosa vaginal o una secreción vaginal, tejido cervical, ganglios, saliva, líquido cefalorraquídeo (LCR), médula ósea, sangre de cordón umbilical, suero, fluido fluidos seroso, plasma, linfa, orina, fluido de ascitis,

fluido pleural, fluido pericardial, fluido peritoneal, fluido abdominal, medio de cultivo, medio de cultivo acondicionado o fluido de lavado. En determinadas realizaciones, las células inmunitarias adaptativas pueden aislarse de una muestra de aféresis. Se pueden obtener muestras de sangre periférica mediante flebotomía de los sujetos. Las células mononucleares de sangre periférica (PBMC, *Peripheral Blood Mononuclear Cells*) se aíslan mediante técnicas conocidas por los expertos en la técnica, por ejemplo, por separación de gradiente de densidad Ficoll-Hypaque®. En determinadas realizaciones, se utilizan PBMC enteras para el análisis.

En determinadas realizaciones relacionadas, se pueden preparar muestras que comprenden predominantemente linfocitos (por ejemplo, linfocitos T y B) o que comprenden predominantemente linfocitos T o predominantemente linfocitos B, para su uso según lo dispuesto en el presente documento, de acuerdo con metodologías aceptadas en la técnica, establecidas.

En otras realizaciones relacionadas, se pueden aislar subpoblaciones específicas de linfocitos T o B antes del análisis, usando los métodos descritos en el presente documento. En la técnica, se conocen diferentes métodos y kits disponibles en el mercado para aislar diferentes subpoblaciones de linfocitos T y B, e incluyen, pero sin limitación, la separación mediante perlas inmunomagnéticas o la clasificación de células inmunocitométricas de flujo de la selección de subconjuntos usando anticuerpos específicos de uno o más de cualquiera de una variedad de marcadores de superficie de linfocitos T y B conocidos. Los marcadores ilustrativos incluyen, pero sin limitación, 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, como lo sabe un experto en la materia, se pueden usar marcadores de superficie celular, tales como CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD45RA y CD45RO para determinar los linajes y las subpoblaciones de linfocitos T, B y monocitos, usando citometría de flujo. De forma similar, se pueden usar marcadores de dispersión de luz directa, dispersión lateral de luz y/o de superficie celular, tales como CD25, CD62L, CD54, CD137 y CD154, para determinar el estado de activación y las propiedades funcionales de las células.

Las combinaciones ilustrativas útiles en ciertos de los métodos descritos en el presente documento pueden incluir CD8⁺CD45RO⁺ (linfocitos T citotóxicos de memoria), CD4⁺CD45RO⁺ (T auxiliares de memoria), CD8⁺CD45RO⁻ (CD8⁺CD62L⁺CD45RA⁺ (linfocitos T citotóxicos sin tratamiento previo); CD4⁺CD25⁺CD62L^{alt}GITR⁺FoxP3⁺ (linfocitos T reguladores). Los anticuerpos ilustrativos para su uso en separaciones de células inmunomagnéticas o clasificación de células inmunocitométricas de flujo incluyen anticuerpos antihumanos marcados con fluorescencia, por ejemplo, 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 las células se puede realizar con la combinación apropiada de anticuerpos, seguida del lavado de las células antes del análisis. Los subconjuntos de linfocitos pueden aislarse mediante clasificación celular activada por fluorescencia (FACS, *Fluorescence Activated Cell Sorting*), por ejemplo, mediante un sistema de clasificación de células BD FACSAria™ (BD Biosciences) y mediante el análisis de los resultados con el software FlowJo™ (Treestar Inc.), y también mediante métodos conceptualmente similares que implican anticuerpos específicos inmovilizados en superficies o perlas.

Para la extracción de ácido nucleico, se puede extraer el ADN genómico total de las células usando métodos conocidos en la técnica y/o kits disponibles en el mercado, por ejemplo, usando el mini kit de sangre de ADN QIAamp® (QIAGEN®). La masa aproximada de un solo genoma haploide es de 3 picogramos (pg). En algunas realizaciones, un solo genoma diploide es de aproximadamente 6,5 picogramos. En una realización, el número absoluto de linfocitos T puede estimarse suponiendo una célula total de material de entrada por 6,5 picogramos de datos genómicos. En algunas realizaciones, se usan al menos de 100.000 a 200.000 células para el análisis, es decir, de aproximadamente 0,6 a 1,2 µg de ADN de linfocitos T o B diploides.

PCR multiplexada

Como se describe en el presente documento, se proporciona un método de cuantificación de la representación relativa de ADN de células inmunitarias adaptativas en ADN de una muestra biológica de prueba de tipos mixtos de células y, por lo tanto, para estimar el número relativo de linfocitos T o B en una mezcla compleja de células. De acuerdo con determinadas realizaciones, el método de cuantificación la representación relativa del ADN de células inmunitarias adaptativas en una mezcla compleja de células implica un método de PCR multiplexada que usa un conjunto de cebadores directos que se hibridan específicamente con los segmentos V y un conjunto de cebadores inversos que se hibridan específicamente con los segmentos J, donde la reacción de PCR multiplexada permite la amplificación de todas las combinaciones posibles de VJ (y VDJ) dentro de una población dada de linfocitos T o B. En algunas realizaciones, el método de PCR multiplexada incluye el uso del conjunto de cebadores del segmento V directos y el conjunto de cebadores del segmento J inversos para amplificar una población dada de oligonucleótidos molde sintéticos que comprenden las combinaciones VJ y VDJ. Debido a que la reacción de PCR multiplexada amplifica esencialmente todas las combinaciones posibles de segmentos V y J, es posible determinar, usando la PCR multiplexada, el número relativo de genomas de linfocitos T o linfocitos B de una muestra que comprende una población mixta de células.

Extracción de ácido nucleico

En una realización, se puede extraer el ADN genómico total de las células usando métodos convencionales conocidos en la técnica y/o kits disponibles en el mercado, por ejemplo, usando el mini kit de sangre de ADN QIAamp® (QIAGEN®). La masa aproximada de un solo genoma haploide es de 3 pg. Preferentemente, se usan de al menos 100.000 a 200.000 células para el análisis de diversidad, es decir, de aproximadamente 0,6 a 1,2 µg de ADN de linfocitos T o B diploides.

Como alternativa, el ácido nucleico total puede aislarse de células, incluyendo tanto el ADN genómico como el ARNm. Si la diversidad se va a medir a partir del ARNm en el extracto de ácido nucleico, el ARNm debe convertirse en ADNc antes de la medición. Esto se puede hacer fácilmente mediante métodos de cualquier experto habitual, por ejemplo, usando transcriptasa inversa de acuerdo con procedimientos conocidos.

En algunas realizaciones, se puede extraer ADN o ARNm de una muestra que comprende una población mixta de células. En determinadas realizaciones, la muestra puede ser una muestra de tejido neoplásico o tejido somático. Las muestras ilustrativas para su uso en los presentes métodos incluyen cualquier tipo de tumor sólido, en particular, un tumor sólido de cáncer colorrectal, hepatocelular, de vesícula biliar, pancreático, esofágico, pulmón, de mama, próstata, de cabeza y cuello, carcinoma de células renales, de ovario, endometrial, de cuello de útero, de vejiga y urotelial. Cualquier tumor sólido en el que se vayan a evaluar linfocitos infiltrantes tumorales se contempla para su uso en los presentes métodos. Los tejidos somáticos que son la diana de una reacción autoinmunitaria incluyen, pero sin limitación, tejidos articulares, piel, tejido intestinal, todas las capas de la úvea, iris, tejido vítreo, corazón, cerebro, pulmones, vasos sanguíneos, hígado, riñón, tejido nervioso, músculo, médula espinal, páncreas, glándula suprarrenal, tendón, membrana mucosa, ganglio linfático, tiroides, endometrio, tejido conjuntivo y médula ósea. En determinadas realizaciones, Se puede extraer ADN o ARN de un órgano trasplantado, tal como un hígado trasplantado, pulmón, riñón, corazón, bazo, páncreas, piel, intestino y timo.

En otras realizaciones, se pueden obtener dos o más muestras de un solo tejido (por ejemplo, un solo tejido neoplásico) y cuantificarse las representaciones relativas de células inmunitarias adaptativas de las dos o más muestras ara considerar variaciones en diferentes secciones de un tejido de prueba. En otras realizaciones determinadas, basta con la determinación de la representación relativa de las células inmunitarias adaptativas en una muestra de un tejido de prueba debido a variaciones mínimas entre las diferentes secciones del tejido de prueba.

Composiciones (cebadores para la PCR multiplexada)

Las composiciones se proporcionan para su uso en una PCR multiplexada que comprende una pluralidad de cebadores del segmento V y una pluralidad de cebadores del segmento J que son capaces de potenciar la amplificación de esencialmente todas las regiones codificantes de la CDR3 del receptor inmunitario adaptativo reordenadas productivamente en una muestra para producir una multiplicidad de moléculas de ADN reordenadas amplificadas de una población de linfocitos T (para los TCR) o linfocitos B (para las Ig) en la muestra.

Los genes de TCR e Ig pueden generar millones de proteínas distintas a través de la mutación somática. Debido a este mecanismo generador de diversidad, las regiones determinantes de la complementariedad hipervariables 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 afín de la proteína. El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígenos de linfocitos T y B con suficiente diversidad para reconocer el universo de posibles patógenos. En los linfocitos T con $\alpha\beta$ y $\gamma\delta$, que reconocen principalmente los antígenos peptídicos presentados por las moléculas MHC, la mayor parte de esta diversidad de receptores está contenida dentro de la tercera región determinante de la complementariedad (CDR3) de las cadenas α y β (o las cadenas γ y δ) del receptor de linfocitos T (TCR).

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

En el presente documento, se describen composiciones que proporcionan una pluralidad de cebadores del segmento V y una pluralidad de cebadores del segmento J que son capaces de amplificar esencialmente todas las combinaciones de los segmentos V y J de un locus del receptor inmunitario reordenado. La expresión

- "esencialmente todas las combinaciones" se refiere al menos al 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de todas las combinaciones de los segmentos V y J de un locus de receptor inmunitario reordenado. En determinadas 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 de receptor inmunitario reordenado. En determinadas realizaciones, la pluralidad de cebadores del segmento V y del segmento J puede comprender o consistir en una secuencia de ácido nucleico que es igual, complementaria o esencialmente complementaria a una secuencia contigua de un segmento codificante de la región V o J diana (es decir, una parte de polinucleótido genómico codificante de un polipéptido de la región V o de la región J, o una parte de ARNm).
- En algunas realizaciones, los cebadores del segmento V y del segmento J son "completamente complementarios" a una secuencia contigua de un segmento de codificación de la región V o J diana, respectivamente. En otras realizaciones, los cebadores del segmento V y del segmento J son "esencialmente complementarios" con respecto a la secuencia contigua de un segmento de codificación de la región V o J diana. En general, no hay más de 4, 3 o 2 pares de bases emparejadas erróneamente tras la hibridación, mientras conserva la capacidad de hibridarse en las condiciones más relevantes para su aplicación final.
- En determinadas realizaciones, se diseñan dos grupos de cebadores para su uso en una reacción de PCR altamente multiplexada. El primer grupo "directo" puede incluir cebadores oligonucleotídicos que son específicos de (por ejemplo, que tienen una secuencia de nucleótidos complementaria a una región de secuencia única de) cada segmento codificante de la región V ("segmento V") en el locus respectivo del gen de TCR o Ig. En determinadas realizaciones, se usan cebadores dirigidos a una región altamente conservada, para capturar simultáneamente muchos segmentos V, reduciendo así el número de cebadores requeridos en la PCR multiplexada. De esta manera, un cebador del segmento V puede ser complementario (por ejemplo, hibridarse) con más de un segmento codificante de la región V de TCR o Ig funcional y actuar como un cebador promiscuo. En otras realizaciones, cada cebador del segmento V es específico de un segmento codificante de la región V de TCR o Ig funcional diferente.
- El cebador del grupo "inverso" pueden incluir cebadores oligonucleotídicos que son específicos de (por ejemplo, que tienen una secuencia de nucleótidos complementaria a una región de secuencia única de) cada segmento codificante de la región J ("segmento J") en el locus respectivo del gen de TCR o Ig. En algunas realizaciones, el cebador J puede hibridarse con una secuencia conservada en el segmento de unión ("J"). En determinadas realizaciones, un cebador del segmento J puede ser complementario a (por ejemplo, hibridarse con) más de un segmento J. En otras realizaciones, cada cebador del segmento V es específico de un segmento codificante de la región J de TCR o Ig funcional diferente. A modo de ilustración y no de limitación, los cebadores del segmento V se pueden usar como cebadores "directos" y los cebadores del segmento J se pueden usar como cebadores "inversos", de acuerdo con la terminología de PCR comúnmente usada, pero el experto apreciará que, en otras determinadas realizaciones, los cebadores del segmento J pueden considerarse cebadores "directos" cuando se usan con cebadores "inversos" del segmento V.
- En algunas realizaciones, el cebador del segmento V o del segmento J tiene al menos 15 nucleótidos de longitud. En otras realizaciones, el cebador del segmento V o del segmento J es de al menos 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 y tiene la misma secuencia que, o es complementaria a, una secuencia contigua del segmento codificante de la región V o J diana. En algunas realizaciones, 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 de la necesidad. Todas las longitudes intermedias de los cebadores descritos actualmente se contemplan para su uso en el presente documento. Como sería reconocido por el experto en la materia, los cebadores pueden comprender secuencias adicionales (por ejemplo, nucleótidos que pueden no ser iguales o complementarios al segmento polinucleotídico codificante de la región V o J diana), tales como sitios de reconocimiento de enzimas de restricción, secuencias adaptadoras universales para secuenciación, secuencias de códigos de barras, modificaciones químicas y similares (véase, por ejemplo, secuencias de cebadores proporcionadas en la lista de secuencias en el presente documento).
- En otras realizaciones, los cebadores del segmento V o del segmento J comprenden secuencias que comparten un alto grado de identidad de secuencia con los cebadores oligonucleotídicos para los que se presentan en el presente documento las secuencias de nucleótidos, incluyendo los establecidos en el Listado de secuencias. En determinadas realizaciones, los cebadores del segmento V o del segmento J comprenden variantes de cebador que pueden tener una identidad sustancial con las secuencias del cebador del segmento V o del segmento J del receptor inmunitario adaptativo descritas en el presente documento. Por ejemplo, dichas variantes de cebador oligonucleotídico pueden comprender al menos el 70 % de identidad de secuencia, preferentemente, al menos el 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 %, o más de identidad de secuencia en comparación con una secuencia oligonucleotídica de referencia, tales como las secuencias de cebador oligonucleotídico divulgadas en el presente documento, usando los métodos descritos en el presente documento (por ejemplo, análisis BLAST usando parámetros convencionales). Un experto en la materia reconocerá que estos valores pueden ajustarse apropiadamente para determinar la capacidad correspondiente de una variante de cebador oligonucleotídico para hibridarse con un polinucleótido codificante del segmento del receptor inmunitario adaptativo teniendo en cuenta la degeneración de los codones, el posicionamiento del marco de lectura y similares. Por lo general, las variantes de

cebador oligonucleotídico contendrán una o más sustituciones, adiciones, eliminaciones y/o inserciones, preferentemente, de manera que la capacidad de hibridación de la variante de oligonucleótido no disminuya sustancialmente en relación con la de una secuencia de cebador del segmento V o del segmento J del receptor inmunitario adaptativo que se expone específicamente en el presente documento. En otras realizaciones, los
 5 cebadores del segmento V o del segmento J están diseñados para ser capaces de amplificar una secuencia de TCR o IGH reordenada que incluye la región de codificación para CDR3.

En algunas realizaciones, como se describe en el presente documento, la pluralidad de cebadores del segmento V y del segmento J comprende cada uno secuencias adicionales en el extremo 5', tales como secuencias de
 10 adaptadores universales, secuencias de códigos de barras, secuencias oligonucleotídicas aleatorias y similares. Las secuencias pueden ser secuencias de origen no natural y/o secuencias que no aparecen de manera natural adyacentes o contiguas a un segmento codificante de la región V o J diana.

En determinadas realizaciones, la pluralidad de cebadores del segmento V y del segmento J están diseñados para producir moléculas de ADN reordenadas amplificadas que tengan menos de 600 nucleótidos de longitud, excluyendo
 15 así los productos de amplificación de locus de receptores inmunitarios adaptativos no reordenados. En algunas realizaciones, las moléculas de ADN reordenadas amplificadas son de al menos 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490,
 20 500, 510, 520, 530, 540, 550, 560, 570, 580, 590 o 600 nucleótidos de longitud. En una realización, la molécula de ADN reordenada amplificada tiene al menos 250 nucleótidos de longitud. En otra realización, la molécula de ADN reordenada amplificada tiene aproximadamente 200 nucleótidos de longitud. La molécula de ADN reordenada amplificada se puede denominar amplicón, molécula amplificada, producto de PCR o producto de amplificación, por
 ejemplo.

Un ensayo de PCR multiplexada ilustrativo usa una pluralidad de cebadores del segmento V directos y una pluralidad de cebadores del segmento J inversos para amplificar selectivamente el VDJ reordenado de cada célula. Si bien estos cebadores pueden hibridarse en los segmentos de genes V y J reordenados y en la línea germinal, la
 25 amplificación por PCR se limita a segmentos de genes reordenados, debido al sesgo de tamaño (por ejemplo, producto de PCR de 250 pb usando segmentos génicos reordenados como moldes frente al producto PCR de > 10 Kb que usa segmentos génicos de línea germinal como moldes).

En algunas realizaciones, la selección de cebadores y el diseño del conjunto de cebadores se pueden realizar de una manera que detecte preferentemente los segmentos productivos de los genes V y J, y excluya los pseudogenes
 35 de TCR o IG. Los pseudogenes pueden incluir segmentos V que contengan un codón de detención dentro del marco dentro de la secuencia de codificación del segmento V, un desplazamiento de marco entre el codón de inicio y la secuencia de codificación CDR3, una o más inserciones de elementos repetidos y eliminaciones de regiones críticas, tales como el primer exón o la RSS. En el locus de IGH humano, por ejemplo, la base de datos ImmunoGeneTics (IMGT) (M.-P. LeFranc, Université Montpellier, Montpellier, Francia; www.imgt.org) anota 165 genes del segmento V,
 40 de los que 26 son orfones en otros cromosomas y 139 están en el locus de 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 abierta, *Open-Reading Frames*) a los que les falta al menos un resto de aminoácido altamente conservado, y 81 son pseudogenes.

Para detectar reordenaciones funcionales de TCR o IG en una muestra mientras se evitan señales de amplificación potencialmente foráneas que pueden ser atribuibles a segmentos de genes V y/o J no productivos como pseudogenes y/u orfones, por tanto, se contempla de acuerdo con determinadas realizaciones, el uso de un subconjunto de cebadores oligonucleotídicos que están diseñados para incluir solo aquellos segmentos V que
 45 participan en un reordenación funcional para codificar un TCR o una Ig, sin tener que incluir cebadores de amplificación específicos para las secuencias de pseudogenes y/u orfones o similares. Así se obtienen eficacias ventajosas con respecto, entre otros aspectos, al tiempo y al gasto.

La pluralidad de cebadores del segmento V y cebadores del segmento J están diseñados para ubicarse fuera de las regiones donde se producen eliminaciones sin molde. Estas posiciones de cebadores del segmento V y de
 55 cebadores del segmento J son relativas a la secuencia de señal de recombinación del gen V (V-RSS) y la secuencia de señal de recombinación del gen J (J-RSS) del segmento génico. En algunas realizaciones, los cebadores del segmento V y los cebadores del segmento J están diseñados para proporcionar información de secuencia adecuada en el producto amplificado para identificar los genes V y J de manera única.

En algunas realizaciones, cada uno de los cebadores del segmento V comprende una primera secuencia y una segunda secuencia, donde la primera secuencia está ubicada 3' con respecto a la segunda secuencia en el cebador del segmento V. En determinadas realizaciones, la primera secuencia es complementaria a una parte de una
 60 primera región de al menos un segmento V, y la primera región del segmento V se ubica inmediatamente 5' con respecto a una segunda región del segmento V donde se producen eliminaciones sin molde durante la reordenación de genes de TCR o Ig. La segunda región del segmento V es adyacente y se encuentra en 5' con respecto a una secuencia de señal de recombinación V (V-RSS) del segmento V. La segunda región donde se producen
 65

eliminaciones sin molde en el segmento V puede tener al menos 10 pares de bases (pb) de longitud. En una realización, el extremo 3' del cebador del segmento V se puede colocar al menos 10 pb en dirección 5' de la V-RSS. En algunas realizaciones, el cebador del segmento V se coloca a más de 40 pares de bases de secuencia en dirección 5' de la V-RSS.

- 5 En otras realizaciones, cada uno de los cebadores del segmento J tiene una primera secuencia y una segunda secuencia, donde la primera secuencia está ubicada 3' con respecto a la segunda secuencia en el cebador del segmento J. La primera secuencia del cebador del segmento J es complementaria a una parte de una primera región de un segmento J, y la primera región del segmento J se ubica inmediatamente 3' con respecto a una segunda región del segmento J donde se producen eliminaciones sin molde durante la reordenación de genes de TCR o Ig. La segunda región del segmento J es adyacente a y se encuentra en 3' con respecto a una secuencia de señal de recombinación J (J-RSS) de dicho segmento J, y la segunda región del segmento J puede tener al menos 10 pares de bases de longitud. En algunas realizaciones, el extremo 3' de los cebadores del segmento J se encuentran al menos 10 pares de bases en dirección 3' de la J-RSS. En determinadas realizaciones, como en los segmentos génicos J β del TCR, la primera región del segmento J incluye un marcador único de cuatro bases en las posiciones +11 a + 14 en dirección 3' del sitio RSS. En otras realizaciones, las eliminaciones del segmento J tienen 4 pb +/- 2,5 pb de longitud, y los cebadores del segmento J se colocan al menos 4 pb en dirección 3' de la J-RSS. En algunas realizaciones, el cebador del segmento J se coloca a más de 30 pares de bases en dirección 3' de la J-RSS.
- 10
- 15
- 20 En los documentos US2010/0058902 y US2012/0058902, y Robins *et al.*, 2009 "Blood" 114, 4099, se puede encontrar una descripción adicional sobre el diseño, la colocación y el posicionamiento de los cebadores del segmento V y cebadores del segmento J, y cebadores ilustrativos.

Amplificación mediante PCR multiplexada

25 Se puede usar un sistema de PCR multiplexada para amplificar los locus de receptores de células inmunitarias adaptativas reordenados a partir de ADN genómico y de oligonucleótidos molde sintéticos, preferentemente, de una región CDR3. En determinadas realizaciones, la región CDR3 se amplifica a partir de una región CDR3 de TCR α , TCR β , TCR γ o TCR δ , o de manera similar, de un locus de Ig, tal como un locus de IgH o IgL (lambda o kappa).

30 En general, un sistema de PCR multiplexada comprende una pluralidad de cebadores directos del segmento V y una pluralidad de cebadores inversos del segmento J. La pluralidad de cebadores directos del segmento V puede comprender 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 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. Cada cebador directo se hibrida específicamente con o es complementario a una secuencia correspondiente a uno o más segmentos de la región V.

40 Por ejemplo, los cebadores del segmento V ilustrativos para la amplificación de TCRB se muestran en SEQ ID NO: 1-120. los cebadores del segmento J ilustrativos para la amplificación de TCRB se muestran en SEQ ID NO: 121-146. Los cebadores del segmento V ilustrativos de TCRG se proporcionan en SEQ ID NO: 147-158. Los cebadores del segmento J ilustrativos de TCRG se proporcionan en SEQ ID NO: 159-166. Los cebadores del segmento V ilustrativos de TCRA y TCRD se encuentran en SEQ ID NO: 167-276. Los cebadores del segmento J ilustrativos de TCRA y TCRD se encuentran en SEQ ID NO: 277-406. Los cebadores del segmento V ilustrativos de IGH se proporcionan en SEQ ID NO: 407-578. Los cebadores del segmento J ilustrativos de IGH se encuentran en SEQ ID NO: 579-592. Los cebadores del segmento V ilustrativos de IGK e IGL se encuentran en SEQ ID NO: 593-740. Los cebadores del segmento J ilustrativos de IGK e IGL se encuentran en SEQ ID NO: 741-764.

50 El sistema de PCR multiplexada puede usar al menos 3, 4, 5, 6 o 7, 8, 9, 10, 11, 12 o 13 cebadores inversos. En algunas realizaciones, cada cebador inverso se hibrida específicamente con o es complementario a una secuencia correspondiente a uno o más segmentos de la región J. En una realización, hay un cebador del segmento J diferente para cada segmento J. Ningún cebador del segmento J es un cebador universal que se una a todos los segmentos de la región J.

55 Los oligonucleótidos que son capaces de aparearse o hibridarse específicamente con una secuencia de ácido nucleico diana mediante la complementariedad de las bases de nucleótidos pueden hacerlo en condiciones de rigurosidad 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 estarían entre 25 y 80 ciclos de PCR, consistiendo cada ciclo en una etapa de desnaturalización (por ejemplo, aproximadamente 10-30 segundos (s) al menos a aproximadamente 95 °C), un etapa de hibridación (por ejemplo, aproximadamente 10-30 segundos a aproximadamente 60-68 °C), y un etapa de extensión (por ejemplo, aproximadamente 10-60 segundos a aproximadamente 60-72 °C), opcionalmente, de acuerdo con ciertas realizaciones, con los etapas de hibridación y extensión combinadas para proporcionar una PCR de dos etapas. Como sería reconocido por el experto en la materia, se pueden añadir o cambiar otros reactivos de PCR en la reacción de PCR para aumentar la especificidad de la hibridación y la amplificación del cebador, tal como alterando la concentración de magnesio, opcionalmente, añadiendo DMSO, y/o el uso de cebadores bloqueados, nucleótidos modificados, ácidos nucleicos peptídicos y

similares.

- En determinadas realizaciones, se pueden usar técnicas de hibridación de ácido nucleico para evaluar la especificidad de hibridación de los cebadores descritos en el presente documento. Las técnicas de hibridación son bien conocidas en la técnica de la biología molecular. Con fines ilustrativos, las condiciones moderadamente rigurosas adecuadas para probar la hibridación de un polinucleótido como se proporciona en el presente documento con otros polinucleótidos incluyen el lavado previo 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 una noche; seguido del lavado dos veces a 65 °C durante 20 minutos con cada una de 2 x, 0,5 x y 0,2 x SSC que contiene SDS al 0,1 %. Un experto en la materia comprenderá que es posible manipular fácilmente la rigurosidad de la hibridación, tal como 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, a excepción de que la temperatura de hibridación se aumenta, por ejemplo, hasta 60 °C-65 °C o 65 °C-70 °C.
- En determinadas realizaciones, los cebadores están diseñados para no atravesar un límite de intrón/exón. Los cebadores directos, en determinadas realizaciones, se hibridan con los segmentos V en una región de conservación de secuencia relativamente fuerte entre segmentos V para aumentar al máximo la conservación de la secuencia entre estos cebadores. Por consiguiente, esto reduce al mínimo el potencial de propiedades de hibridación diferenciales de cada cebador, y de modo que la región amplificada entre los cebadores del segmento V y J contiene suficiente información de secuencia V de TCR o Ig para identificar el segmento específico del gen V usado. En una realización, los cebadores del segmento J se hibridan con un elemento conservado del segmento J y tienen una resistencia de hibridación similar. En una realización particular, los cebadores del segmento J se hibridan con el mismo motivo de región marco conservada.
- Los oligonucleótidos (por ejemplo, cebadores) se pueden preparar 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 dietilfosforamidita de Beaucage *et al.*, 1981, *Tetrahedron Lett.* 22:1859-1862; y el método de soporte sólido de la patente de EE.UU. n.º 4.458.066. En Goodchild, 1990, *Bioconjugate Chemistry* 1(3): 165-187, se proporciona una revisión de los métodos de síntesis de conjugados de oligonucleótidos y nucleótidos modificados.

Secuenciación de alto rendimiento

Secuenciación de oligonucleótidos

- En una realización, los cebadores del segmento V y los cebadores del segmento J de la invención incluyen una segunda subsecuencia situada en sus extremos 5' que incluye una secuencia adaptadora universal complementaria y que puede hibridarse con secuencias adaptadoras de secuenciación para su uso en un secuenciador de ADN, tal como Illumina.
- En determinadas realizaciones, los segmentos génicos codificantes de la región J tienen cada uno un marcador de identificación definido por una secuencia única de 2, 3, 4, 5, 6, 7, 8, 9, 10 o aproximadamente 15, 20 o más nucleótidos, situado en una posición definida en relación con un sitio RSS. Por ejemplo, se puede usar un marcador de cuatro bases, en el segmento codificante de la región J β de las regiones codificantes de CDR3 TCR β amplificadas, en las posiciones +11 a +14 en dirección 3' del sitio RSS. Sin embargo, estas y otras realizaciones relacionadas no necesitan ser tan limitadas, y también contemplan otros marcadores de identificación definidos por secuencias de nucleótidos relativamente cortas que pueden detectarse en segmentos de genes codificantes de la región J y definidos en función de sus posiciones con respecto a un sitio RSS. Estos pueden variar entre diferentes locus codificantes de receptores inmunitarios adaptativos.
- La secuencia de señal de recombinación (RSS) consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3', y nonámero, 5'-ACAAAACC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Se ha identificado una serie de posiciones de nucleótidos como importantes para la recombinación, incluyendo el dinucleótido CA en la posición uno y dos del heptámero, y también se ha demostrado que un nucleótido C en la posición del heptámero tres es muy preferido, así como un nucleótido A en las posiciones 5, 6, 7 del nonámero. (Ramsden *et al.*, 1994; Akamatsu *et al.*, 1994; Hesse *et al.*, 1989). Las mutaciones de otros nucleótidos tienen efectos mínimos o no uniformes. El espaciador, aunque más variable, también tiene un impacto en la recombinación, y se ha demostrado que los reemplazos de un solo nucleótido tienen un impacto significativo en la eficacia de la recombinación (Fanning *et al.* 1996, Larijani *et al.*, 1999; Nadel *et al.*, 1998). Se han descrito criterios para identificar secuencias polinucleotídicas de RSS que tienen eficacias de recombinación significativamente diferentes (Ramsden *et al.* 1994; Akamatsu *et al.*, 1994; Hesse *et al.*, 1989 y Cowell *et al.*, 1994). Por consiguiente, los oligonucleótidos de secuenciación pueden hibridarse adyacentes a un marcador de cuatro bases dentro de los segmentos de genes codificantes de J amplificados en las posiciones +11 a +14 en dirección 3' del sitio RSS. Por ejemplo, se pueden diseñar oligonucleótidos de secuenciación para TCRB para que se hibriden con un motivo de nucleótido consenso observado justo en dirección 3' de este "marcador", para que las cuatro primeras bases de una secuencia de lectura identifiquen de manera única el segmento génico codificante de J. Se encuentran secuencias oligonucleotídicas de

secuenciación ilustrativas, por ejemplo, en SEQ ID NO: 765-786.

La información usada para asignar identidades a los segmentos de codificación J y V de una secuencia de lectura está completamente contenida dentro de la secuencia amplificada, y no depende de la identidad de los cebadores de PCR. En particular, los métodos descritos en el presente documento permiten la amplificación de todas las combinaciones posibles de V-J en un locus de TCR o Ig, y la secuenciación de las moléculas amplificadas individuales permite la identificación y cuantificación del ADN reordenado codificante de las regiones CDR3. La diversidad de las células inmunitarias adaptativas de una muestra dada se puede inferir de las secuencias generadas usando los métodos y algoritmos descritos en el presente documento.

Métodos de secuenciación de alto rendimiento

Los métodos de la invención comprenden además la secuenciación las moléculas de ADN codificantes del receptor inmunitario adaptativo amplificadas codificante de que se producen. La secuenciación puede realizarse en productos de amplificación producidos a partir de una muestra biológica que comprende células inmunitarias adaptativas y/o los oligonucleótidos molde sintéticos que se describen a continuación.

En una realización, la secuenciación implica el uso de un conjunto de oligonucleótidos de secuenciación (secuencias adaptadoras) que se hibridan con secuencias oligonucleotídicas de secuenciación dentro de las moléculas de ADN amplificadas o los oligonucleótidos molde sintéticos que se describen a continuación.

La secuenciación puede realizarse usando cualquiera de una variedad de máquinas y sistemas de secuenciación de moléculas individuales de alto rendimiento disponibles. Los sistemas ilustrativos de secuencias incluyen sistemas de secuencias por síntesis, tales como el analizador de genomas Illumina, Illumina MiSeq e instrumentos asociados (Illumina, Inc., San Diego, CA), Sistema de análisis genético Helicos (Helicos BioSciences Corp., Cambridge, MA), Pacific Biosciences PacBio RS (Pacific Biosciences, Menlo Park, CA) u otros sistemas que tienen capacidades similares. La secuenciación se realiza usando un conjunto de oligonucleótidos de secuenciación que se hibridan con una región definida dentro de las moléculas de ADN amplificadas. Los oligonucleótidos de secuenciación están diseñados de manera que los segmentos génicos codificantes de V y J pueden identificarse de manera única por las secuencias que se generan, basándose en la presente divulgación y en vista de las secuencias de genes del receptor inmunitario adaptativo conocidas que aparecen en bases de datos disponibles para el público en general.

En determinadas realizaciones, se secuencian al menos 30, 40, 50, 60, 70, 80, 90, 100, 101-150, 151-200, 201-300, 301-500, y no más de 1000 nucleótidos contiguos de las moléculas de ADN codificantes del receptor inmunitario adaptativo amplificadas. En algunas realizaciones, los amplicones y los oligonucleótidos molde sintéticos que se secuencian tienen menos de 600 pb de longitud. En realizaciones adicionales, las lecturas de secuenciación resultantes tienen aproximadamente 130 pb de longitud. En otras realizaciones adicionales, se producen aproximadamente 30 millones de lecturas de secuenciación por ensayo de secuenciación.

Las composiciones y los métodos para la secuenciación de secuencias de genes del receptor inmunitario adaptativo reordenadas y para la determinación del clonotipo del receptor inmunitario adaptativo se describen en mayor detalle en 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 EE.UU. n.º 2012/0058902, publicación de EE.UU. n.º 2010/0330571, documentos WO/2010/151416, WO/2011/106738 y WO2012/027503.

En determinadas realizaciones, los segmentos génicos codificantes de la región J amplificados pueden tener cada uno un marcador de identificación definido por una única secuencia de 2, 3, 4, 5, 6, 7, 8, 9, 10 o aproximadamente 15, 20 o más nucleótidos, situado en una posición definida en relación con un sitio RSS. Por ejemplo, se puede usar un marcador de cuatro bases, en el segmento codificante de la región J β de las regiones codificantes de CDR3 TCR β amplificadas, en las posiciones +11 a +14 en dirección 3' del sitio RSS. Sin embargo, estas y otras realizaciones relacionadas no necesitan ser tan limitadas, y también contemplan otros marcadores de identificación definidos por secuencias de nucleótidos relativamente cortas que pueden detectarse en segmentos de genes codificantes de la región J y definidos en función de sus posiciones con respecto a un sitio RSS. Estos pueden variar entre diferentes locus codificantes de receptores inmunitarios adaptativos.

La secuencia de señal de recombinación (RSS) consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3', y nonámero, 5'-ACAAAACC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Se ha identificado una serie de posiciones de nucleótidos como importantes para la recombinación, incluyendo el dinucleótido CA en la posición uno y dos del heptámero, y también se ha demostrado que un nucleótido C en la posición del heptámero tres es muy preferido, así como un nucleótido A en las posiciones 5, 6, 7 del nonámero. (Ramsden *et al.*, 1994; Akamatsu *et al.*, 1994; Hesse *et al.*, 1989). Las mutaciones de otros nucleótidos tienen efectos mínimos o no uniformes. El espaciador, aunque más variable, también tiene un impacto en la recombinación, y se ha demostrado que los reemplazos de un solo nucleótido tienen un impacto significativo en la eficacia de la recombinación (Fanning *et al.* 1996, Larjani *et al.*, 1999; Nadel *et al.*, 1998). Se han descrito criterios para identificar secuencias polinucleotídicas de RSS que tienen eficacias de recombinación significativamente diferentes (Ramsden

et al. 1994; Akamatsu *et al.*, 1994; Hesse *et al.*, 1989 y Cowell *et al.*, 1994). Por consiguiente, los oligonucleótidos de secuenciación pueden hibridarse adyacentes a un marcador de cuatro bases dentro de los segmentos de genes codificantes de J amplificados en las posiciones +11 a +14 en dirección 3' del sitio RSS. Por ejemplo, se pueden diseñar oligonucleótidos de secuenciación para TCRB para que se hibriden con un motivo de nucleótido consenso observado justo en dirección 3' de este "marcador", para que las cuatro primeras bases de una secuencia de lectura identifiquen de manera única el segmento génico codificante de J. Los cebadores J de TCRB ilustrativos se encuentran en SEQ ID NO: 121-146 (cebadores inversos del segmento J de TCRB (específicos del gen) y cebadores inversos del segmento J de TCRB con una secuencia adaptadora universal.

La información usada para asignar identidades a los segmentos de codificación J y V de una secuencia de lectura está completamente contenida dentro de la secuencia amplificada, y no depende de la identidad de los cebadores de PCR. En particular, los métodos descritos en el presente documento permiten la amplificación de todas las combinaciones posibles de V-J en un locus de TCR o Ig, y la secuenciación de las moléculas amplificadas individuales permite la identificación y cuantificación del ADN reordenado codificante de las regiones CDR3. La diversidad de las células inmunitarias adaptativas de una muestra dada se puede inferir de las secuencias generadas usando los métodos y algoritmos descritos en el presente documento. Una ventaja sorprendente proporcionada en ciertas realizaciones preferidas por las composiciones y los métodos de la presente divulgación fue la capacidad de amplificar con éxito todas las combinaciones posibles de V-J de un locus de receptor de células inmunitarias adaptativas en una sola reacción de PCR multiplexada.

En determinadas realizaciones, los oligonucleótidos de secuenciación descritos en el presente documento pueden seleccionarse de modo que el cebado promiscuo de una reacción de secuenciación para un segmento génico codificante de J por un oligonucleótido específico a otro segmento génico codificante de J diferente genera datos de secuencia que comienzan exactamente en el mismo nucleótido que los datos de secuencia del oligonucleótido de secuenciación correcto. De este modo, la hibridación promiscua de los oligonucleótidos de secuenciación no afecta a la calidad de los datos de secuencia generados.

La longitud media de la región de codificación de CDR3, para el TCR, definida como los nucleótidos codificantes del polipéptido TCR entre la segunda cisteína conservada del segmento V y la fenilalanina conservada del segmento J, es de 35 +/- 3 nucleótidos. Por consiguiente, y en determinadas realizaciones, la amplificación por PCR usando cebadores del segmento V y cebadores del segmento J que comienzan a partir del marcador del segmento J de una determinada región J de TCR o IgH (por ejemplo, J β de TCR, J γ de TCR y JH de IgH como se describe en el presente documento) casi siempre capturará la unión V-D-J completa en una lectura de 50 pares de bases. La longitud media 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 de TCR β , pero normalmente estará entre aproximadamente 10 y aproximadamente 70 nucleótidos. Por consiguiente, y en determinadas realizaciones, la amplificación por PCR usando cebadores del segmento V y cebadores del segmento J que comienzan a partir del marcador del segmento J de IgH capturará la unión V-D-J completa en una lectura de 100 pares de bases.

Los cebadores de PCR que se unen y apoyan la extensión de polinucleótidos en secuencias molde emparejadas erróneamente se denominan cebadores promiscuos. En determinadas realizaciones, se pueden diseñar cebadores de PCR inversos del segmento J de TCR e Ig para reducir al mínimo el solapamiento con los oligonucleótidos de secuenciación, para reducir al mínimo el cebado promiscuo en el contexto de la PCR multiplexada. En una realización, los cebadores inversos del segmento J de TCR e Ig pueden anclarse en el extremo 3' mediante hibridación con el motivo del sitio de corte y empalme de consenso, con un solapamiento mínimo de los cebadores de secuenciación. En general, los cebadores de los segmentos V y J de TCR e Ig pueden seleccionarse para operar en la PCR a temperaturas de hibridación uniformes usando el diseño de secuencia/cebador conocido y programas de análisis bajo parámetros predeterminados. Para la reacción de secuenciación, los cebadores ilustrativos de J de IGH usados para la secuenciación se encuentran en SEQ ID NO: 579-592 (que muestra los cebadores inversos del segmento J de IGH (gen específico) y los cebadores inversos del segmento J de IGH con una secuencia adaptadora universal.

Procesamiento de datos de secuencias

Como se divulga en el presente documento, también se proporcionan métodos de análisis de las secuencias del diverso conjunto de regiones codificantes de CDR3 reordenadas que se generan usando las composiciones y los métodos que se describen en el presente documento. En particular, se proporciona un algoritmo para corregir el sesgo de la PCR, los errores de secuenciación y de PCR, y para estimar la verdadera distribución de los clonotipos específicos (por ejemplo, un TCR o una Ig que tiene una secuencia CDR3 reordenada de forma única) de una muestra. En el presente documento, se describe un algoritmo preferido con mayor detalle. Como sería reconocido por el experto en la materia, los algoritmos proporcionados en el presente documento pueden modificarse adecuadamente para adaptarse a situaciones clínicas o experimentales particulares.

El uso de una etapa de PCR para amplificar las regiones CDR3 de TCR o Ig antes de la secuenciación podría introducir un sesgo sistemático en la abundancia relativa inferida de las secuencias, debido a las diferencias en la

eficiencia de la amplificación por PCR de las regiones CDR3 utilizando diferentes segmentos génicos V y J. Como se analiza con más detalle en los ejemplos, cada ciclo de amplificación por PCR puede introducir un sesgo de magnitud media $1,5^{1/15} = 1,027$. Por lo tanto, los 25 ciclos de PCR introducen un sesgo total de magnitud media $1,027^{25} = 1,95$ en la abundancia relativa inferida de distintas secuencias de la región CDR3.

5 Las lecturas secuenciadas se filtran para obtener aquellas que incluyen secuencias CDR3. El procesamiento de datos del secuenciador implica un número de etapas para eliminar los errores en la secuencia primaria de cada lectura y para comprimir los datos. Un filtro de complejidad elimina aproximadamente el 20 % de las secuencias que son lecturas erróneas del secuenciador. Después, era necesario que las secuencias tuvieran un mínimo de una
10 coincidencia de seis bases con una de las regiones J y una de las regiones V de TCR o Ig. Aplicando el filtro al carril de control que contiene la secuencia de fagos, como media, solo una secuencia de 7-8 millones pasó estas etapas. Finalmente, se usa un algoritmo de vecino más próximo para contraer los datos en secuencias únicas fusionando secuencias estrechamente relacionadas, para eliminar tanto el error de la PCR como el error de la secuenciación.

15 Analizando los datos, la proporción de secuencias del producto de PCR se obtiene trabajando hacia atrás desde los datos de secuencia previos a la estimación de la verdadera distribución de los clonotipos (por ejemplo, secuencias clónicas únicas) en la sangre. Para cada secuencia observada, un número dado de veces en los datos del presente documento, se estima la probabilidad de que se muestree esa secuencia de un grupo de PCR de un determinado tamaño. Debido a que las regiones CDR3 secuenciadas se muestrean aleatoriamente de un grupo masivo de
20 productos de PCR, el número de observaciones para cada secuencia se extrae de las distribuciones de Poisson. Los parámetros de Poisson se cuantifican de acuerdo con el número de genomas de linfocitos T que proporcionaron el molde para la PCR. Un modelo simple de mezcla de Poisson estima estos parámetros y coloca una probabilidad por pares para cada secuencia que se extrae de cada distribución. Este es un método de maximización de expectativas que reconstruye la abundancia de cada secuencia que se extrajo de la sangre.

25 Para estimar el número total de secuencias únicas de CDR3 del receptor inmunitario adaptativo que están presentes en una muestra, se puede emplear un enfoque computacional que emplee la fórmula de "especies invisibles" (Efron y Thisted, 1976 *Biometrika* 63, 435-447). Este enfoque estima el número de especies únicas (por ejemplo, secuencias únicas de receptores inmunitarios adaptativos) de una gran población compleja (por ejemplo, una población de células inmunitarias adaptativas tales como linfocitos T o linfocitos B), basándose en el número de especies únicas observadas en una muestra finita, aleatoria, de una población (Fisher *et al.*, 1943 *J. Anim. Ecol.* 12:42-58; Ionita-Laza *et al.*, 2009 *Proc. Nat. Acad. Sci. EE.UU.* 106:5008). El método emplea una expresión que predice el número de especies "nuevas" que se observaría si se analizara una segunda muestra aleatoria, finita de idéntico tamaño de la misma población. Las especies "invisibles" se refieren al número de nuevas secuencias de
30 receptores inmunitarios adaptativos que se detectarían si se repitieran un número infinito de veces las etapas de amplificación de secuencias codificantes de receptores inmunitarios adaptativos de una muestra y de determinación de la frecuencia de aparición de cada secuencia única de la muestra. A modo de teoría no limitante, se supone operativamente, a los fines de estas estimaciones, que las células inmunitarias adaptativas (por ejemplo, los linfocitos T, los linfocitos B) circulan libremente en el compartimento anatómico del sujeto que es la fuente de la muestra a partir de la que se estima la diversidad (por ejemplo, sangre, linfa, etc.).
40

Para aplicar esta fórmula, los clonotipos de receptores inmunitarios adaptativos únicos (por ejemplo, TCR β , TCR α , TCR γ , TCR δ , IgH) toman el lugar de las especies. La solución matemática proporciona que, para S, el número total de receptores inmunitarios adaptativos que tienen secuencias únicas (por ejemplo, las "especies" o los clonotipos
45 TCR β , TCR γ , IgH, que, en ciertas realizaciones, pueden ser secuencias CDR3 únicas), un experimento de secuenciación observa x_s copias de la secuencia s. Para todos los clonotipos no observados, x_s es igual a 0, y cada clonotipo de TCR o Ig se "captura" en el curso de la obtención de una muestra aleatoria (por ejemplo, una extracción de sangre) de acuerdo con un proceso de Poisson con el parámetro λ_s . El número de genomas de linfocitos T o B secuenciados en la primera medición se define como 1, y el número de genomas de linfocitos T o B secuenciados en la segunda medición se define como t.
50

Debido a que hay una gran cantidad de secuencias únicas, se usa una integral en lugar de una suma. Si $G(\lambda)$ es la función de distribución empírica de los parámetros $\lambda_1, \dots, \lambda_S$, y n_x es el número de clonotipos (por ejemplo, secuencias únicas de TCR o Ig, o secuencias únicas de CDR3) observados exactamente x veces, entonces el
55 número total de clonotipos, es decir, la medición de la diversidad E, viene dada por la siguiente fórmula (I):

$$E(n_x) = S \int_0^{\infty} \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) dG(\lambda) \quad (I)$$

Por consiguiente, La fórmula (I) se puede usar para estimar la diversidad total de especies en toda la fuente de la que se toman las muestras de tamaño idéntico. Sin desear quedar ligados a teoría alguna, el principio es que el
60 número de clonotipos muestreados en una muestra de cualquier tamaño dado contiene información suficiente para estimar la distribución subyacente de clonotipos en toda la fuente. Se puede determinar el valor para $\Delta(t)$, el número

de nuevos clonotipos observados en una segunda medición, preferentemente usando la siguiente ecuación (II):

$$\Delta(t) = \sum_x E(n_x)_{med1+med2} - \sum_x E(n_x)_{med1} = S \int_0^{\infty} e^{-\lambda} (1 - e^{-\lambda t}) dG(\lambda)$$

en la que *med1* y *med2* son el número de clonotipos de las mediciones 1 y 2, respectivamente. La expansión de Taylor de $1 - e^{-\lambda t}$ y la sustitución en la expresión de $\Delta(t)$ produce:

$$\Delta(t) = E(x_1)t - E(x_2)t^2 + E(x_3)t^3 - \dots, \quad (III)$$

que puede aproximarse reemplazando las expectativas ($E(n_x)$) por las secuencias de números reales observadas exactamente x veces en la primera medición de la muestra. La expresión para $\Delta(t)$ oscila ampliamente cuando t tiende al infinito, entonces $\Delta(t)$ se regulariza para producir un límite inferior para $\Delta(\infty)$, por ejemplo, usando la transformación de Euler (Efron *et al.*, 1976 *Biometrika* 63: 435).

De acuerdo con ciertas realizaciones divulgadas expresamente en el presente documento, también se proporcionan métodos en los que el grado de clonalidad de las células inmunitarias adaptativas que están presentes en una muestra, tal como una muestra que comprende una mezcla de células, solo algunas de las cuales son células inmunitarias adaptativas, puede determinarse ventajosamente sin la necesidad de realizar una clasificación celular o una secuenciación del ADN. Estas y otras realizaciones relacionadas superan los desafíos de la eficacia, el tiempo y el coste que, antes de la presente divulgación, han impedido la capacidad de determinar si la presencia de células inmunitarias adaptativas en una muestra (por ejemplo, TIL) es monoclonal u oligoclonal (por ejemplo, si todos los TIL son la progenie de una o un número relativamente limitado de células inmunitarias adaptativas), o si, en cambio, la presencia de células inmunitarias adaptativas en la muestra es policlonal (por ejemplo, los TIL son la progenie de un número relativamente grande de células inmunitarias adaptativas).

De acuerdo con la teoría no limitante, estas realizaciones aprovechan el conocimiento actual en la técnica (también descrito anteriormente) de que, una vez que una célula inmunitaria adaptativa (por ejemplo, un linfocito T o B) ha reordenado sus genes codificantes de receptores inmunitarios adaptativos (por ejemplo, TCR o Ig), las células de su progenie poseen la misma reordenación de genes codificantes de receptores inmunitarios adaptativos, dando lugar así a una población clónica que puede identificarse por la presencia en su interior de segmentos génicos V y J codificantes de CDR3 reordenados que pueden amplificarse mediante una combinación específica por pares de cebadores oligonucleotídicos específicos de V y J como se divulga en el presente documento.

Composiciones de oligonucleótidos molde sintéticos para su uso en la cuantificación de genomas de entrada de células inmunitarias adaptativas y en la determinación de la representación relativa de células inmunitarias adaptativas

Composiciones de moldes sintéticos útiles para la cuantificación de los números de moléculas de entrada en una muestra

Se pueden diseñar oligonucleótidos molde sintéticos para la cuantificación de una serie de moléculas de entrada en una muestra biológica. Como se usa en el presente documento, "molde sintético" significa un oligonucleótido que contiene secuencias que incluyen secuencias esencialmente idénticas a secuencias biológicas (es decir, segmentos de V, J o C de TCR o Ig, o regiones de control genómico), además de secuencias que no se producen de manera natural (es decir, códigos de barras, aleatorizadores, adaptadores, etc.). La secuencia completa de nucleótidos de moldes sintéticos, por tanto, no se da en la naturaleza, y son, en cambio, secuencias diseñadas y creadas en laboratorio. Se determina una proporción del número de moléculas oligonucleotídicas molde sintéticas de entrada en una muestra en comparación con el número de lecturas de secuenciación de salida totales de oligonucleótidos molde sintéticos (secuenciados a partir de amplicones molde sintéticos) en la muestra. En una realización, se añade una dilución limitante de oligonucleótidos molde sintéticos (que permite la determinación del número de moléculas oligonucleotídicas molde sintéticas totales presentes midiendo el número de secuencias oligonucleotídicas molde sintéticas únicas observadas) a una muestra biológica para la PCR multiplexada, y suponiendo que se mantiene la misma proporción para los moldes biológicos y sintéticos, la proporción se usa para determinar el número de moléculas de receptor de linfocitos T o B reordenadas y, por lo tanto, el número de linfocitos T o B, de la muestra biológica.

En el presente documento, se describe una composición de moldes sintéticos que comprende una pluralidad de oligonucleótidos molde de fórmula general (I):



Los oligonucleótidos molde constituyentes son diferentes con respecto a las secuencias de nucleótidos de los

oligonucleótidos molde individuales.

En una realización, U1 y U2 son cada uno nada o cada uno comprende un oligonucleótido que tiene, independientemente, una secuencia que se selecciona entre (i) una secuencia oligonucleotídica adaptadora universal e (ii) una secuencia oligonucleotídica específica de la plataforma de secuenciación que está unida a y posicionada en 5' con respecto a la secuencia oligonucleotídica adaptadora universal.

B1, B2 y B3 pueden independientemente cada uno ser nada o comprender un oligonucleótido "B" que comprende una secuencia de código de barras oligonucleotídica 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 1.000 nucleótidos contiguos (incluyendo todos los valores enteros entre ellos). En algunas realizaciones, B1, B2 y B3 pueden comprender, cada uno, una secuencia oligonucleotídica única que identifique de forma única, o identifique como una combinación de pares, (i) la secuencia oligonucleotídica V única del oligonucleótido molde e (ii) la secuencia oligonucleotídica J única del oligonucleótido molde.

El posicionamiento relativo de los oligonucleótidos de código de barras B1, B2, B3 y los adaptadores universales U1 y U2 permite ventajosamente la identificación y cuantificación rápidas de los productos de amplificación de un oligonucleótido molde único dado mediante lecturas de secuencia corta y secuenciación de extremos emparejados en secuenciadores de ADN automatizados (por ejemplo, Illumina HiSeq™ o Illumina MiSeq®, o GeneAnalyzer™ -2, Illumina Corp., San Diego, CA). En particular, estas y otras realizaciones relacionadas permiten la determinación rápida de alto rendimiento de combinaciones específicas de una secuencia del segmento V y una secuencia del segmento J que están presentes en un producto de amplificación, para caracterizar así la eficacia de la 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 ADN codificante de 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, opcionalmente, incluyendo lecturas de secuencia que se extienden a B2.

V puede ser nada o 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 de ADN. En algunas realizaciones, la secuencia de ADN es de una secuencia génica codificante de la región variable (V) del receptor inmunitario adaptativo, o su complemento, y en cada una de la pluralidad de secuencias oligonucleotídicas molde V comprende una secuencia oligonucleotídica única.

J puede ser nada o 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 de ADN. En algunas realizaciones, la secuencia de ADN es de una secuencia génica codificante de la región de unión (J) al receptor inmunitario adaptativo, o su complemento, y en cada una de la pluralidad de secuencias oligonucleotídicas molde J comprende una secuencia oligonucleotídica única.

Al construir las partes "V" y "J" de los oligonucleótidos molde sintéticos de fórmula I, se pueden usar distintas secuencias génicas de la región variable (V) y de la región de unión (J) del receptor inmunitario adaptativo. Se conoce una gran cantidad de secuencias génicas de la región V y J como secuencias de nucleótidos y/o aminoácidos, incluyendo secuencias de ADN genómico no reordenadas de locus de TCR e Ig, y secuencias de ADN reordenadas productivamente en dichos locus y sus productos codificados, y también incluyendo pseudogenes en estos locus, y también incluyendo orfones relacionados. Véanse, por ejemplo, los documentos US2012/0058902, US2010/0330571, WO2011/106738 y WO2012/027503. Además, 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 polinucleotídicas codificantes de los productos de genes de TCR y BCR expresados, reordenados, y también incluyen secuencias polinucleotídicas de pseudogenes que se han identificado en los locus de la región V. Las diferentes secuencias polinucleotídicas de V que pueden incorporarse a los moldes de fórmula general (I) divulgados en el presente documento pueden variar ampliamente en cuanto a su longitud, composición de nucleótidos (por ejemplo, contenido de GC) y secuencia polinucleotídica lineal real, y se sabe, por ejemplo, que incluyen "puntos calientes" o regiones hipervariables que presentan una determinada diversidad de secuencias. Estas y otras secuencias conocidas en la técnica pueden usarse de acuerdo con la presente divulgación para el diseño y la producción de oligonucleótidos molde que se incluirán en la composición de moldes proporcionada en el presente documento para estandarizar la eficacia de amplificación de un conjunto de cebadores oligonucleotídicos, y para el diseño y la producción del conjunto de cebadores oligonucleotídicos que es capaz de amplificar ADN reordenado codificante de cadenas polipeptídicas de TCR o Ig, cuyo ADN reordenado puede estar presente en una muestra biológica que comprende ADN de células linfoides.

La secuencia polinucleotídica completa de cada polinucleótido V de la fórmula general (I) puede, pero no es necesario, consistir exclusivamente en nucleótidos contiguos de cada gen V distinto. Por ejemplo y de acuerdo con ciertas realizaciones, en la composición de moldes descrita en el presente documento, cada polinucleótido V de fórmula (I) solo necesita tener al menos una región que comprenda una secuencia oligonucleotídica V única que se encuentre en un gen V y al que pueda hibridarse 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 parte prescrita (por ejemplo, al menos 15, 20, 30, 60, 90, 120, 150, 180 o 210 nucleótidos contiguos, o cualquier valor entero entre ellos) de una secuencia del gen V natural (incluyendo una secuencia del pseudogén V), siempre que se incluya al menos una región de secuencia oligonucleotídica V única (por ejemplo, el sitio de hibridación del cebador) que no se incluya en ningún otro polinucleótido V molde.

En algunas realizaciones, la pluralidad de polinucleótidos V que están presentes en la composición de moldes sintéticos tienen longitudes que simulan las longitudes totales de las secuencias de nucleótidos del gen V naturales, conocidas, incluso donde las secuencias de nucleótidos específicas difieren entre la región V molde y cualquier gen V natural. Las longitudes de la región V de los moldes sintéticos pueden diferir de las longitudes de las secuencias del gen V de origen natural en no más del 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 sintético descrito en el presente documento incluye un codón de parada en o cerca del extremo 3' de V de la fórmula general (I).

El polinucleótido V de la fórmula (I) puede así, en determinadas realizaciones, comprender una secuencia de nucleótidos que tenga una longitud que sea igual o similar a la longitud de un gen V típico desde su codón de inicio hasta su región codificante de CDR3 y puede, pero no es necesario, incluir una secuencia de nucleótidos codificante de la región CDR3. Las secuencias de nucleótidos codificantes de CDR3 y las longitudes de secuencia pueden variar considerablemente, y se han caracterizado por varios esquemas de numeración diferentes (por ejemplo, Lefranc, 1999 *The Immunologist* 7:132; Kabat *et al.*, 1991, en: "Sequences of Proteins of Immunological Interest", NIH Publication 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; véase también, por ejemplo, Rock *et al.*, 1994 *J. Exp. Med.* 179:323; Saada *et al.*, 2007 *Immunol. Cell Biol.* 85:323).

En resumen, la región CDR3 normalmente abarca la parte polipeptídica que se extiende desde un resto de cisteína altamente conservado (codificado por el codón trinucleotídico TGY; Y = T o C) en el segmento V hasta un resto de fenilalanina altamente conservado (codificado por TTY) en el segmento J de TCR, o hasta un triptófano altamente conservado (codificado por TGG) en IGH. Más del 90 % de las reordenaciones productivas, naturales, en el locus de TCRB tienen una longitud de codificación de CDR3 según este criterio de entre 24 y 54 nucleótidos, correspondiente a entre 9 y 17 aminoácidos codificados. Las longitudes de CDR3 de los oligonucleótidos molde sintéticos divulgados en el presente documento deberían, para cualquier locus de TCR o BCR dado, estar dentro del mismo intervalo que el 95 % de las reordenaciones naturales. Por lo tanto, por ejemplo, en una composición de moldes sintéticos descrita en el presente documento, la parte codificante de la CDR3 de la cabina del polinucleótido V tiene una longitud de 24 a 54 nucleótidos, incluyendo cada número entero intermedio. Los esquemas de numeración para las regiones codificantes de CDR3 descritas anteriormente indican 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 codificantes de estos aminoácidos conservados pueden haber sido reemplazados por un codón codificante de un aminoácido diferente. Para los pseudogenes que no usan estos aminoácidos conservados, la longitud de CDR3 puede definirse en relación con la posición correspondiente en la que se habría observado el resto conservado sin la sustitución, de acuerdo con uno de los esquemas de numeración de posiciones de la secuencia de CDR3 establecidos mencionados anteriormente.

La secuencia polinucleotídica completa de cada polinucleótido J de la fórmula general (I) puede, pero no es necesario, consistir exclusivamente en nucleótidos contiguos de cada gen J distinto. Por ejemplo y de acuerdo con ciertas realizaciones, en la composición de moldes descrita en el presente documento, cada polinucleótido J de fórmula (I) solo necesita tener al menos una región que comprenda una secuencia oligonucleotídica J única que se encuentre en un gen J y al que pueda hibridarse 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 parte prescrita (por ejemplo, al menos 15, 20, 30, 60, 90, 120, 150, 180 o 210 nucleótidos contiguos, o cualquier valor entero intermedio) de una secuencia del gen V natural (incluyendo una secuencia de pseudogén V) siempre que se incluya al menos una única la región de secuencia oligonucleotídica V (el sitio de hibridación del cebador) que no se incluya en ningún otro polinucleótido J molde.

Puede preferirse, en ciertas realizaciones, que la pluralidad de polinucleótidos J que están presentes en la composición de moldes descrita en el presente documento tengan longitudes que simulen las longitudes globales de secuencias de nucleótidos del gen J naturales, conocidas, incluso donde las secuencias de nucleótidos específicas difieren entre la región J molde y cualquier gen J natural. Las longitudes de la región J de los moldes descritos en el presente documento pueden diferir de las longitudes de las secuencias del gen J de origen natural en no más del 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 de la fórmula (I) puede así, en determinadas realizaciones, comprender una secuencia de nucleótidos que tenga una longitud que sea igual o similar a la longitud de un gen J típico natural y puede, pero no es necesario, incluir una secuencia de nucleótidos codificante de la región CDR3, como se ha descrito 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 de genes de la región J incluyen secuencias polinucleotídicas codificantes de los productos de genes de TCR y BCR reordenados expresados y no expresados. Las diferentes secuencias polinucleotídicas de J que pueden incorporarse a los
 5 moldes de fórmula general (I) divulgados en el presente documento pueden variar ampliamente en cuanto a su longitud, composición de nucleótidos (por ejemplo, contenido de GC) y secuencia polinucleotídica lineal real.

El experto en la materia puede seleccionar alternativas a las secuencias V y J descritas en el presente documento, para su uso en la construcción de los oligonucleótidos molde descritos en el presente documento y/o los cebadores
 10 oligonucleotídicos del segmento V y del segmento J, basándose en la presente divulgación usando el conocimiento en la técnica con respecto a secuencias de genes publicadas para las regiones codificantes de V y J de los genes para cada subunidad de TCR e Ig. Las entradas de referencia de Genbank para secuencias de receptores inmunitarios adaptativos humanos incluyen: TCRA: (TCRA/D): NC_000014.8 (chr14:22090057..23021075); TCRβ: (TCRB): NC_000007.13 (chr7:141998851..142510972); TCRγ: (TCRG): NC_000007.13 (chr7:38279625..38407656);
 15 cadena pesada de inmunoglobulina, IgH (IGH): NC_000014.8 (chr14: 106032614..107288051); cadena ligera de kappa de inmunoglobulina, IgLk (IGK): NC_000002.11 (chr2: 89156874..90274235); y cadena ligera de lambda de inmunoglobulina, IGLλ (IGL): NC_000022.10 (chr22: 22380474..23265085). Las entradas del Genbank de referencia para secuencias de locus del receptor inmunitario adaptativo de ratón incluyen: TCRβ: (TCRB): NC_000072.5 (chr6: 40841295..41508370) y cadena pesada de inmunoglobulina, IgH (IGH): NC_000078.5
 20 (chr12:114496979..117248165).

Se pueden realizar análisis del diseño de moldes y cebadores, y consideraciones de la selección de sitios diana, por ejemplo, usando el software de análisis de cebadores OLIGO y/o el software 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.
 25

Por consiguiente, basándose en la presente divulgación y en vista de estas secuencias de genes del receptor inmunitario adaptativo y metodologías de diseño de oligonucleótidos conocidas, para su inclusión en los presentes oligonucleótidos molde, los expertos en la materia pueden diseñar una pluralidad de secuencias polinucleotídicas específicas de la región V y específicas de la región J, cada una de las cuales contiene independientemente
 30 secuencias oligonucleotídicas que son únicas para un gen V y J dado, respectivamente. De forma similar, a partir de la presente divulgación y en vista de las secuencias de receptores inmunitarios adaptativos conocidos, los expertos en la materia también pueden diseñar un conjunto de cebadores que comprenda una pluralidad de cebadores oligonucleotídicos específicos de la región V y específicos de la región J, siendo cada uno de ellos independientemente capaz de hibridarse con una secuencia específica que es única de un gen V y J dado,
 35 respectivamente, mediante lo que la pluralidad de cebadores es capaz de amplificar esencialmente todos los genes V y esencialmente todos los genes J en un locus de codificación de receptor inmunitario adaptativo dado (por ejemplo, un locus de TCR o IgH humano). Dichos conjuntos de cebadores permiten la generación, en la PCR multiplexada (por ejemplo, usando múltiples pares de cebadores directos e inversos), de productos de amplificación que tienen un primer extremo que está codificado por un segmento génico codificante de la región V reordenado y
 40 un segundo extremo que está codificado por un segmento génico codificante de la región J.

Por lo general, y en ciertos casos, dichos productos de amplificación pueden incluir una secuencia codificante de CDR3, aunque se contemplan en el presente documento productos de amplificación que no incluyen una secuencia codificante de CDR3. Los cebadores pueden diseñarse preferentemente para producir productos de amplificación que tengan suficientes partes de secuencias V y J y/o secuencias de código de barras (B) V-J como se describe en el presente documento, de modo que al secuenciar los productos (amplicones), es posible identificar según las secuencias que son únicas para cada segmento génico (i) el gen V particular, e (ii) el gen J particular en las proximidades de las cuales el gen V se reorganizó para producir un gen codificante de un receptor inmunitario adaptativo funcional. Por lo general, y en realizaciones preferidas, los productos de amplificación por PCR no tendrán más de 600 pares de bases de tamaño, que, de acuerdo con la teoría no limitante, excluirá los productos de amplificación de genes de receptores inmunitarios adaptativos no reordenados. En otras ciertas realizaciones preferidas, los productos de amplificación no tendrán más de 500, 400, 300, 250, 200, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30 o 20 pares de bases de tamaño, de modo que pueden proporcionar ventajosamente la cuantificación de alto rendimiento, rápida, de amplicones distintos de secuencia mediante lecturas de secuencia corta.
 55

En una realización de fórmula I, V es una secuencia polinucleotídica codificante de al menos 10-70 aminoácidos contiguos de una región V del receptor inmunitario adaptativo, o su complemento; J es una secuencia polinucleotídica codificante de al menos 5-30 aminoácidos contiguos de una región J del receptor inmunitario adaptativo, o su complemento; U1 y U2 no son nada o comprenden un oligonucleótido que comprende una
 60 secuencia de nucleótidos que se selecciona entre (i) una secuencia oligonucleotídica adaptadora universal e (ii) una secuencia oligonucleotídica específica de la plataforma de secuenciación que está unida y posicionada en 5' con respecto a la secuencia oligonucleotídica adaptadora universal; B1, B2 y B3 cada uno independientemente son nada o comprenden un oligonucleótido B que comprende una secuencia de código de barras oligonucleotídica de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos contiguos, donde, en cada una de la pluralidad de secuencias oligonucleotídicas, B comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación por parejas, (i) la secuencia oligonucleotídica V única y (ii) la secuencia oligonucleotídica J
 65

única.

En otra realización de fórmula (I), V es una secuencia polinucleotídica 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, 340, 350, 360, 370, 380, 390, 400 o 450 y no más de 1.000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia del gen de la región variable (V) del receptor inmunitario adaptativo (por ejemplo, TCR o BCR), o su complemento, y en cada una de la pluralidad de secuencias oligonucleotídicas V comprende una secuencia oligonucleotídica única.

Se puede encontrar una descripción adicional sobre oligonucleótidos molde sintéticos en el documento WO2013/169957.

La Figura 1A ilustra un ejemplo de un oligonucleótido molde sintético, de acuerdo con una realización de la invención. En una realización, un oligonucleótido molde sintético comprende las siguientes regiones (de izquierda a derecha, como se muestra en la Figura 1): una secuencia de cebador universal (UA) (102), un código de barras (CB) específico del molde (104), una secuencia que comprende una parte o la totalidad de una secuencia génica codificante de la región variable (V) única del receptor inmunitario adaptativo (gen V) (106), un marcador interno (MI) del molde sintético (108), una repetición del código de barras (CB) (104), una repetición del marcador interno (MI) (108), una secuencia que comprende una parte o la totalidad de una secuencia génica codificante de la región variable (J) única del receptor inmunitario adaptativo (gen J) (110), una tercera repetición del código de barras (CB) (104) y una secuencia de cebador universal inverso (UB) (112). Cada oligonucleótido molde sintético incluye una secuencia génica única codificante de la región variable (V) del receptor inmunitario adaptativo y una secuencia génica única codificante de la región de unión (J) al receptor inmunitario adaptativo. La combinación de secuencias V y J en los oligonucleótidos molde sintéticos es la misma que la que se encuentra en las moléculas biológicas que comprenden combinaciones únicas de secuencias V y J reordenadas en la muestra.

En un ejemplo, el oligonucleótido molde sintético puede ser una secuencia de 495 pb que comprende una secuencia de cebador universal (UA) (102), un código de barras (CB) específico del molde de 16 pb (104), una secuencia génica codificante de la región variable (V) del receptor inmunitario adaptativo de 300 pb (gen V) (106), un marcador interno (MI) del molde sintético de 9 pb (108), una repetición del código de barras (CB) (104), una repetición del marcador interno (MI) (108), una secuencia génica codificante de la región variable (J) del receptor inmunitario adaptativo (gen J) de 100 pb (110), una tercera repetición del código de barras (CB) (104) y una secuencia de cebador universal inverso (UB) (112). Se pueden usar diferentes longitudes de las secuencias y orden de las regiones en el diseño de los oligonucleótidos molde sintéticos, como es sabido por un experto en la materia.

Los oligonucleótidos molde sintéticos de Fórmula I también pueden incluir secuencias adaptadoras. Las secuencias adaptadoras pueden añadirse a los oligonucleótidos molde sintéticos mediante el diseño de cebadores que incluyan secuencias adaptadoras en sus extremos 5' y que se hibriden específicamente con las regiones UA y UB adaptadoras en los oligonucleótidos molde sintéticos (véase la Figura 1(A)). Un ejemplo de una secuencia adaptadora es una secuencia adaptadora Illumina, como la descrita en el apartado "Adaptadores" que figura más adelante.

En una realización, los amplicones de oligonucleótidos sintéticos resultantes tienen la estructura de fórmula general I y pueden incluir una secuencia adaptadora o secuencias adaptadoras (secuencia Illumina), de modo que la secuencia del oligonucleótido molde sintético comprende lo siguiente: una secuencia adaptadora, una secuencia de cebador universal (UA) (102), un código de barras (CB) específico del molde (104), una secuencia génica codificante de la región variable (V) del receptor inmunitario adaptativo (gen V) (106), un marcador interno (MI) del molde sintético (108), una repetición del código de barras (CB) (104), una repetición del marcador interno (MI) (108), una secuencia génica codificante de la región variable (J) del receptor inmunitario adaptativo (gen J) (110), una tercera repetición del código de barras (CB) (104), una secuencia de cebador universal inverso (UB) (112) y una segunda secuencia adaptadora.

Número de oligonucleótidos molde sintéticos en la muestra

En determinadas realizaciones, la composición de molde sintético comprende una pluralidad de oligonucleótidos molde sintéticos distintos y únicos. En una realización, la pluralidad de oligonucleótidos molde sintéticos comprende al menos a o al menos b secuencias oligonucleotídicas únicas, lo que sea mayor, donde a es el número de segmentos génicos codificantes de la región V única del receptor inmunitario adaptativo en el sujeto, y b es el número de segmentos génicos codificantes de la región J única del receptor inmunitario adaptativo 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.

En otra realización, la pluralidad de oligonucleótidos molde comprende al menos ($a \times b$) secuencias oligonucleotídicas únicas, donde a es el número de segmentos génicos codificantes de la región V única del receptor inmunitario adaptativo en el sujeto y b es el número de segmentos génicos codificantes de la región J única del receptor inmunitario adaptativo en el sujeto, y la composición comprende al menos un oligonucleótido molde por

cada posible combinación de un segmento génico codificante de la región V y un segmento génico codificante de la región J.

5 Por consiguiente, la composición puede albergar al menos una aparición de cada secuencia de polinucleótido V único y al menos una aparición de cada secuencia de polinucleótido J único, donde, en algunos casos, la al menos una aparición de un determinado polinucleótido V único estará presente en el mismo oligonucleótido molde donde se puede encontrar al menos una aparición de un determinado polinucleótido J único. Por lo tanto, por ejemplo, "al menos un oligonucleótido molde para cada polinucleótido V único y al menos un oligonucleótido molde para cada polinucleótido J único", en ciertos casos, puede referirse a un solo oligonucleótido molde donde están presentes un polinucleótido V único y un polinucleótido J único.

15 En una realización, *a* es de 1 a un número máximo de segmentos génicos V en el genoma de mamífero del sujeto. En otra realización, *b* es de 1 a un número máximo de segmentos génicos J en el genoma de mamífero del sujeto. En otras realizaciones, *a* es 1. En otras realizaciones, *b* es 1.

20 En algunas realizaciones, *a* puede variar de 1 segmento génico V a 54 segmentos génicos V para TCRA, 1-76 segmentos génicos V para TCRB, 1-15 segmentos génicos V para TCRG, 1-7 segmentos génicos V para TCRD, 1-165 segmentos génicos V para IGH, 1-111 para IGK, o 1-79 segmentos génicos V para IGL. En otras realizaciones, *b* puede variar de 1 segmento génico J a 61 segmentos génicos J para TCRA, 1-14 segmentos génicos J para TCRB, 1-5 segmentos génicos J para TCRG, 1-4 segmentos génicos J para TCRD, 1-9 segmentos génicos J para IGH, 1-5 segmentos génicos J para IGK, o 1-11 segmentos génicos J para IGL. En determinadas realizaciones, un conjunto de oligonucleótidos molde sintéticos que comprende cualquier combinación posible de un segmento génico codificante de la región V y un segmento génico codificante de la región J comprende 248 tipos de moldes sintéticos únicos para TCRA/D, 858 tipos sintéticos únicos para TCRB, 70 tipos de moldes sintéticos únicos para TCRG, 1116 tipos de moldes sintéticos únicos para IGH y 370 tipos de moldes sintéticos únicos para IGK/I.

25 La siguiente tabla enumera el número de segmentos génicos V (*a*) y segmentos génicos J (*b*) para cada locus de receptor inmunitario adaptativo humano, incluyendo segmentos V y J funcionales.

30 **Tabla 1: Número de segmentos génicos V (*a*) y segmentos génicos J (*b*)**

	Segmentos V*	Segmentos V funcionales**	Segmentos J*	Segmentos J funcionales**
TCRA	54	45	61	50
TCRB	76	48	14	13
TCRG	15	6	5	5
TCRD	7	7	4	4
IGH	165	51	9	6
IGK	111	44	5	5
IGL	79	33	11	7
*Genes de segmentos de región variable y de unión totales				
**Genes de segmentos de región variable y de unión con al menos un alelo funcional				

35 En algunas realizaciones, el polinucleótido J del oligonucleótido molde sintético 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 región constante J del receptor inmunitario adaptativo, o su complemento.

40 Sin embargo, la invención que se contempla en el presente documento no pretende ser tan limitada, de modo que, en determinadas realizaciones, se puede usar ventajosamente un número esencialmente menor de oligonucleótidos molde. En estas y otras realizaciones relacionadas, donde *a* es el número de segmentos génicos codificantes de la región V única del receptor inmunitario adaptativo en un sujeto y *b* es el número de segmentos génicos codificantes de la región J única del receptor inmunitario adaptativo en el sujeto, el número mínimo de secuencias oligonucleotídicas únicas de las cuales está compuesta la pluralidad de oligonucleótidos molde sintéticos puede determinarse por el mayor de entre *a* y *b*, siempre que cada secuencia de polinucleótido V única y cada secuencia de polinucleótido J única esté presente en al menos un oligonucleótido molde sintético en la composición de moldes. Por lo tanto, de acuerdo con determinadas realizaciones relacionadas, la composición de moldes puede comprender al menos un oligonucleótido molde sintético para cada polinucleótido V único, por ejemplo, que incluye uno solo de cada polinucleótido V único de acuerdo con la fórmula general (I), y al menos un oligonucleótido molde sintético para cada polinucleótido J único, por ejemplo, que incluye uno solo de cada polinucleótido J único de acuerdo con la fórmula general (I).

50 En otras realizaciones determinadas, la composición de moldes comprende al menos un oligonucleótido molde sintético al que se puede hibridar cada cebador de amplificación de oligonucleótido de un conjunto de cebadores de amplificación.

Es decir, en determinadas realizaciones, la composición de moldes comprende al menos un oligonucleótido molde

sintético que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento V puede hibridarse específicamente, y al menos un oligonucleótido molde sintético que tiene una secuencia oligonucleotídica de fórmula general (I) con la que cada cebador oligonucleotídico del segmento J puede hibridarse específicamente.

5 De acuerdo con dichas realizaciones, el conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado codificante de 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. La pluralidad a' de los cebadores oligonucleotídicos del segmento V son cada uno independientemente capaces de hibridarse o hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región V del receptor inmunitario adaptativo o con su complemento, 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 codificante de la región V del receptor inmunitario adaptativo. La pluralidad b' de cebadores oligonucleotídicos del segmento J son cada uno independientemente capaces de hibridarse o hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región J del receptor inmunitario adaptativo o con su complemento, 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 codificante de la región J del receptor inmunitario adaptativo.

20 En algunas realizaciones, a' es lo mismo que a (descrito anteriormente para oligonucleótidos molde sintéticos). En otras realizaciones, b' es lo mismo que b (descrito anteriormente para los oligonucleótidos molde sintéticos).

Por lo tanto, en ciertas realizaciones y como también se analiza en otra parte del presente documento, la presente composición de moldes sintéticos se puede usar en reacciones de amplificación con cebadores de amplificación que están diseñados para amplificar todas las secuencias de genes codificantes de receptores inmunitarios adaptativos reordenados, incluyendo los que no se expresan. En otras realizaciones determinadas, la composición de moldes y los cebadores de amplificación pueden diseñarse de manera que no produzcan productos de amplificación de genes reordenados que no se expresen (por ejemplo, pseudogenes, orfones). Por lo tanto, se apreciará que, en determinadas realizaciones, solo se puede amplificar deseablemente un subconjunto de genes codificantes de receptores inmunitarios adaptativos reordenados, de modo que los subconjuntos de cebadores de amplificación adecuados pueden diseñarse y emplearse para amplificar solo aquellas secuencias V-J reordenadas que son de interés. En estas y otras realizaciones relacionadas, de forma correspondiente, puede usarse una composición de moldes sintéticos que solo comprenda un subconjunto de interés de secuencias V-J reordenadas, siempre y cuando la composición de moldes sintéticos comprenda al menos un oligonucleótido molde sintético con el que se pueda hibridar cada cebador de amplificación de oligonucleótidos de un conjunto de cebadores de amplificación. El número real de oligonucleótidos molde sintéticos de la composición de moldes puede variar considerablemente entre las realizaciones contempladas, en función del conjunto de cebadores de amplificación que se use.

Por ejemplo, en determinadas realizaciones relacionadas, en la composición de moldes, la pluralidad de oligonucleótidos molde sintéticos comprende secuencias encontradas en SEQ ID NO: 787-1644.

Cebadores para su uso con oligonucleótidos molde sintéticos

45 El polinucleótido V de la fórmula general (I) (o su complemento) incluye secuencias con las que se pueden hibridar específicamente los miembros de conjuntos de cebadores oligonucleotídicos específicos para los genes TCR o BCR. Se describen conjuntos de cebadores que son capaces de amplificar el ADN reordenado codificante de una pluralidad de TCR o BCR, por ejemplo, en los documentos US2012/0058902, US2010/0330571, WO2011/106738 y WO2012/027503; o similares; o que, como se describe en el presente documento, pueden diseñarse para incluir secuencias oligonucleotídicas que pueden hibridarse específicamente con cada gen V único y con cada gen J único en un locus de gen de TCR o BCR particular (por ejemplo, TCR α , β , γ o δ , o IgH μ , γ , δ , α o ϵ , o IgL k o λ).

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 ADN reordenado codificante de uno o una pluralidad de TCR o BCR puede incluir normalmente 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 hibridarse 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 el presente documento. En determinadas realizaciones, los cebadores pueden comprender al menos 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 nucleótidos, y en determinadas realizaciones, 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. También se contemplan expresamente cebadores y sitios de hibridación de cebadores de otras longitudes, como se divulga en el presente documento.

65 El polinucleótido J de la fórmula general (I) (o su complemento) incluye secuencias con las que se pueden hibridar específicamente los miembros de conjuntos de cebadores oligonucleotídicos específicos de los genes de TCR o BCR. Se describen conjuntos de cebadores que son capaces de amplificar el ADN reordenado codificante de una

pluralidad de TCR o BCR, por ejemplo, en los documentos US2012/0058902, US2010/0330571, WO2011/106738 y WO2012/027503; o similares; o, como se describe en el presente documento, que se pueden diseñar para incluir secuencias oligonucleotídicas que pueden hibridarse específicamente con cada gen V único y con cada gen J único en un locus de gen de TCR o BCR particular (por ejemplo, α , β , γ o δ de TCR, o μ , γ , δ , α o ϵ de IgH, o k o λ de IgL).

Estos cebadores oligonucleotídicos del segmento V y del segmento J pueden comprender secuencias adaptadoras universales en sus extremos 5' para secuenciar los amplicones resultantes, como se ha descrito anteriormente y en los documentos US2012/0058902, US2010/0330571, WO2011/106738 y WO2012/027503. La Figura 1B ilustra cebadores que se hibridan con regiones específicas de las secuencias del segmento V y del segmento J, y también incluyen secuencias adaptadoras universales.

En determinadas realizaciones, se pueden proporcionar conjuntos de cebadores oligonucleotídicos para amplificación en cantidades esencialmente equimolares. Como también se describe en el presente documento, de acuerdo con otras determinadas realizaciones, la concentración de uno o más cebadores de un conjunto de cebadores puede ajustarse deliberadamente para que determinados cebadores no estén presentes en cantidades equimolares o en cantidades esencialmente equimolares.

ADAPTADORES

Los oligonucleótidos molde descritos en el presente documento de fórmula general (I) también pueden comprender, en determinadas realizaciones, la primera secuencia oligonucleotídica adaptadora universal (U1) (102) y la segunda secuencia oligonucleotídica adaptadora universal (U2) (112), o pueden carecer de cualquiera de o ambas U1 (102) y U2 (112). Por lo tanto, U1 (102) puede comprender nada o un oligonucleótido que tenga una secuencia seleccionada entre (i) una primera secuencia oligonucleotídica adaptadora universal e (ii) una primera secuencia oligonucleotídica específica de la plataforma de secuenciación que esté unida y posicionada en 5' con respecto a una primera secuencia oligonucleotídica adaptadora universal; y U2 (112) puede comprender nada o un oligonucleótido que tenga una secuencia seleccionada entre (i) una segunda secuencia oligonucleotídica adaptadora universal e (ii) una segunda secuencia oligonucleotídica específica de la plataforma de secuenciación que esté unida y posicionada en 5' con respecto a una segunda secuencia oligonucleotídica adaptadora universal.

U1 (102) y/o U2 (112) pueden, por ejemplo, comprender secuencias oligonucleotídicas adaptadoras universales y/o secuencias oligonucleotídicas específicas de la plataforma de secuenciación que son específicas de una tecnología de secuenciación de una sola molécula que se esté empleando, por ejemplo, los sistemas HiSeq™ o CeneAnalyzer™-2 (GA-2) (Illumina, Inc., San Diego, CA) u otro conjunto adecuado de instrumentos de secuenciación, reactivos y programas informáticos. La inclusión de dichas secuencias adaptadoras específicas de la plataforma permite la secuenciación cuantitativa directa de la composición de moldes descrita en el presente documento, que comprende una pluralidad de oligonucleótidos molde diferentes de fórmula general (I), usando 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 de moldes.

En particular, la capacidad de secuenciar todos los componentes de la composición de moldes permite verificar directamente que cada oligonucleótido molde de la pluralidad de oligonucleótidos molde está presente en una cantidad esencialmente equimolar. Por ejemplo, se puede generar un conjunto de oligonucleótidos molde descritos en el presente documento que tengan secuencias adaptadoras universales en ambos extremos, de modo que las secuencias adaptadoras puedan usarse para incorporar además oligonucleótidos específicos de la plataforma de secuenciación en cada extremo de cada molde.

Sin desear quedar ligados a teoría alguna, los oligonucleótidos específicos de la plataforma se pueden añadir a los extremos de dichos moldes modificados usando oligonucleótidos 5' (secuencia 3' del adaptador universal 1 de la secuencia de la plataforma 5') y 3' (secuencia 3' del adaptador universal 2 de la secuencia de la plataforma 5') en tan solo dos ciclos de desnaturalización, hibridación y extensión, de modo que no se modifique cuantitativamente la representación relativa en la composición de moldes de cada uno de los oligonucleótidos molde constituyentes. Las secuencias de identificación únicas (por ejemplo, las secuencias de códigos de barras B que comprenden secuencias oligonucleotídicas V y B únicas que están asociadas y, por lo tanto, identifican, respectivamente, regiones V y J individuales, como se describe en el presente documento) se colocan adyacentes a las secuencias adaptadoras, permitiendo así la secuenciación cuantitativa en lecturas de secuencias cortas, para caracterizar la población de moldes según el criterio de la cantidad relativa de cada secuencia molde única que está presente.

Cuando dicha secuenciación cuantitativa directa indique que uno o más oligonucleótidos particulares pueden estar expresados excesiva o deficientemente en una preparación de la composición de moldes, se puede realizar el ajuste de la composición de moldes consecuentemente para obtener una composición de moldes en la que todos los oligonucleótidos estén presentes en cantidades esencialmente equimolares. La composición de moldes en la que todos los oligonucleótidos están presentes en cantidades esencialmente equimolares se puede usar como un patrón de calibración para conjuntos de cebadores de amplificación, tal como en los métodos divulgados en el presente documento para determinar y corregir el potencial de amplificación no uniforme entre los miembros de un conjunto de cebadores.

Quando los cebadores tienen cola con los adaptadores universales + Illumina y se secuencian con los adaptadores Illumina (véase la Figura 1), estos moldes se comportan de la misma manera que los moldes sintéticos típicos. Cuando se amplifican usando cebadores de PCR multiplexada VF y JR, y se secuencian con cebadores JR, estas moléculas producen una lectura de secuenciación con la siguiente estructura (5' a 3'): (1) secuencia del gen J (aproximadamente 15 pares de bases), (2) un marcador interno (MI) de molde sintético de 9 pares de bases, (3) un código de barras (CB) de V-J de 16 pares de bases, (4) un segundo marcador interno (MI) de molde sintético de 9 pares de bases y (5) un gen V (aproximadamente 15 pares de bases).

Además de las secuencias adaptadoras descritas en SEQ ID NO: 765-786, otras secuencias oligonucleotídicas que pueden usarse como secuencias adaptadoras universales serán conocidas por aquellos familiarizados con la técnica en vista de la presente divulgación, incluyendo la selección de secuencias oligonucleotídicas adaptadoras que sean distintas de las secuencias encontradas en otras partes de los moldes descritos en el presente documento.

15 **CÓDIGOS DE BARRAS**

Como se describe en el presente documento, determinadas realizaciones contemplan el diseño de las secuencias oligonucleotídicas molde para que contengan secuencias de identificación cortas que permitan la identificación inequívoca de la secuencia molde y, por lo tanto, de al menos un cebador responsable de la amplificación de ese molde, sin tener que secuenciar todo el producto de amplificación. En los oligonucleótidos molde sintéticos descritos en el presente documento de fórmula general (I), B1, B2, B3 y B4 pueden, cada uno independientemente, ser nada o comprender un oligonucleótido B que comprenda una secuencia de código de barras oligonucleotídica 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 1.000 nucleótidos contiguos (incluyendo todos los valores enteros intermedios), donde, en cada una de la pluralidad de secuencias oligonucleotídicas molde B, comprende una secuencia oligonucleotídica única que identifica de forma única, como una combinación por parejas, (i) la secuencia oligonucleotídica V única del oligonucleótido molde e (ii) la secuencia oligonucleotídica J única del oligonucleótido molde.

Por lo tanto, por ejemplo, los oligonucleótidos molde sintéticos que tienen secuencias de identificación de código de barras pueden permitir la realización de lecturas de secuencias de productos de amplificación relativamente cortas, tales como lecturas de secuencias de código de barras de no más de 1.000, 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, tras lo que se hace coincidir esta información de secuencias de 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. Mediante este enfoque, se puede secuenciar simultáneamente una gran cantidad de productos de amplificación de forma simultánea mediante secuenciación paralela de alto rendimiento, para identificar cebadores que sean responsables del sesgo de amplificación en un conjunto complejo de cebadores.

Los códigos de barras ilustrativos pueden comprender un primer oligonucleótido de código de barras de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 nucleótidos que identifique de forma única cada polinucleótido V del molde y un segundo oligonucleótido de código de barras de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 nucleótidos que identifique de forma única cada polinucleótido J del 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 estas y otras realizaciones relacionadas no pretenden ser tan limitadas. Los oligonucleótidos de código de barras pueden comprender secuencias oligonucleotídicas de cualquier longitud, siempre que se obtenga una longitud mínima de código de barras que impida la aparición de una secuencia de código de barras dada en dos o más oligonucleótidos molde que tengan secuencias, por lo demás, distintas (por ejemplo, secuencias V y J).

Por lo tanto, la longitud mínima del código de barras, para evitar dicha redundancia entre los códigos de barras que se usan para identificar de forma exclusiva diferentes pares de secuencias V-J, es X nucleótidos, donde 4^X es superior al número de especies de moldes distintas que deben diferenciarse basándose en si poseen secuencias no idénticas. Por ejemplo, para el conjunto de 858 oligonucleótidos molde establecidos en SEQ ID NO: 1888-3003, la longitud mínima del código de barras sería de cinco nucleótidos, lo que permitiría un total teórico de 1.024 (es decir, superior a 871) secuencias pentanucleotídicas posibles diferentes. En la práctica, las longitudes de las lecturas de secuencias oligonucleotídicas de código de barras solo pueden estar limitadas por los límites de la longitud de las lecturas de secuencias del instrumento de secuenciación de nucleótidos que se vaya a emplear. Para ciertas realizaciones, diferentes oligonucleótidos de código de barras que distinguirán especies individuales de oligonucleótidos molde deberían tener al menos dos emparejamientos erróneos de nucleótidos (por ejemplo, una distancia de Hamming mínima de 2) cuando se alinean para aumentar al máximo el número de nucleótidos que coinciden en determinadas posiciones en las secuencias oligonucleotídicas de código de barras.

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

El experto en la materia estará familiarizado con el diseño, la síntesis y la incorporación a una construcción de oligonucleótidos o polinucleótidos de mayor tamaño, de secuencias de código de barras oligonucleotídicas 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 intermedios. Para ejemplos no limitantes del diseño y la implementación de estrategias de identificación de secuencias de código de barras oligonucleotídicas, véase, por ejemplo, de Career *et al.*, 2011 *Adv. Env. Microbiol.* 77:6310; Parameswaran *et al.*, 2007 *Nucl. Ac. Res.* 35(19):330; Roh *et al.*, 2010 *Trends Biotechnol.* 28:291.

Por lo general, los códigos de barras se colocan en moldes en ubicaciones donde no se encuentran de manera natural, es decir, los códigos de barras comprenden secuencias de nucleótidos que son distintas de cualquier secuencia oligonucleotídica de origen natural que se pueda encontrar en las proximidades de las secuencias adyacentes a las que se encuentran los códigos de barras (por ejemplo, secuencias V y/o J). Dichas secuencias de código de barras pueden incluirse, de acuerdo con determinadas realizaciones descritas en el presente documento, como los elementos B1, B2 y/o B3 del oligonucleótido molde divulgado en el presente documento de fórmula general (I). Por consiguiente, algunos de los oligonucleótidos molde descritos en el presente documento 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 otras determinadas realizaciones, algunos o todos estos códigos de barras pueden estar ausentes. En determinadas realizaciones, todas las secuencias de código de barras tendrán un contenido de GC idéntico o similar (por ejemplo, diferirán en el contenido de GC en no más del 20 %, o en no más del 19, 18, 17, 16, 15, 14, 13, 12, 11 o 10 %).

En las composiciones de moldes de acuerdo con determinadas realizaciones divulgadas en el presente documento, el elemento B que contiene código de barras (por ejemplo, B1, B2, B3 y/o B4) comprende la secuencia oligonucleotídica que identifica de forma única una combinación de V-J por parejas. Opcionalmente, y en determinadas realizaciones, el elemento B que contiene el 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, situado en dirección 5' y/o en dirección 3' de la secuencia de código de barras específica que identifica de forma única cada combinación V-J por parejas específica. Cuando está presente tanto en dirección 5' como en dirección 3' de la secuencia de código de barras específica, el nucleótido aleatorio o la secuencia polinucleotídica aleatoria son independientes entre sí, es decir, pueden pero no necesitan comprender el mismo nucleótido o la misma secuencia polinucleotídica.

ALEATORIZADORES

En algunas realizaciones, el oligonucleótido molde sintético comprende una secuencia oligonucleotídica generada aleatoriamente, o una secuencia "aleatorizadora". La secuencia aleatoria, en general, se sitúa entre las secuencias V y J, pero puede ubicarse en otro lugar a lo largo del oligonucleótido molde sintético. En una realización, la secuencia aleatorizadora solo se da una vez en el molde sintético. N comprende una secuencia oligonucleotídica aleatoria de 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 o más de 30 nucleótidos contiguos.

El número de posibles secuencias de nucleótidos de longitud X es 4^x , así pues, un segmento de nucleótidos aleatorio de incluso una longitud corta puede codificar muchas posibles secuencias de nucleótidos únicas. Por ejemplo, una secuencia aleatorizadora de 12 pares de bases podría codificar una cualquiera de 16.777.216 secuencias de nucleótidos únicas. La secuencia aleatorizadora garantiza que cualquiera de los dos oligonucleótidos molde sintéticos tenga una probabilidad de aproximadamente 1 de 17 millones de contener la misma secuencia aleatorizadora. Por lo tanto, se pueden incluir decenas o cientos de miles de oligonucleótidos molde sintéticos en la reacción de PCR con superposición mínima o nula en secuencias aleatorizadoras entre dos oligonucleótidos molde sintéticos distintos.

Las secuencias aleatorizadoras permiten la cuantificación exacta de cada oligonucleótido molde sintético. Tras la amplificación de un grupo de oligonucleótidos molde sintéticos, cada secuencia de nucleótidos aleatoria única observada en el resultado de la secuenciación representa una sola molécula de material de entrada. Por lo tanto, el número de entrada de oligonucleótidos molde sintéticos añadidos a la reacción de amplificación se puede determinar contando el número de secuencias de nucleótidos aleatorias únicas. Además, el número de entrada de oligonucleótidos molde sintéticos asociados con un determinado código de barras (y, por lo tanto, asociado con una determinada combinación por parejas de una secuencia oligonucleotídica V y una secuencia oligonucleotídica J) puede determinarse contando el número de secuencias de nucleótidos aleatorias únicas asociadas con un determinado código de barras. Se pueden encontrar ejemplos de moldes sintéticos que comprendan aleatorizadores, por ejemplo, en las SEQ ID NO: 3004-3159.

SITIOS DE ENZIMAS DE RESTRICCIÓN

De acuerdo con determinadas realizaciones divulgadas en el presente documento, el oligonucleótido molde también puede comprender un sitio de reconocimiento de endonucleasa de restricción (ER) que está situado entre las secuencias V y J, y que no ocurre en ninguna otra parte de la secuencia oligonucleotídica molde. El sitio de reconocimiento de ER puede estar opcionalmente adyacente a un sitio de código de barras que identifique la

secuencia de la región V. El sitio ER puede incluirse para cualquiera de una serie de fines, incluyendo, pero sin limitación, una característica estructural que puede aprovecharse para destruir moldes selectivamente mediante la puesta en contacto con la enzima de restricción apropiada. Puede ser deseable degradar selectivamente los presentes oligonucleótidos molde poniéndolos en contacto con una ER adecuada, por ejemplo, para eliminar oligonucleótidos molde de otras composiciones en las que puedan haberse introducido deliberada o accidentalmente. Como alternativa, el sitio ER puede aprovecharse útilmente en el curso de la secuenciación de oligonucleótidos molde de la composición de moldes, 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 ER ilustrativo es el motivo oligonucleotídico GTCGAC, que es reconocido por la enzima de restricción Sal I. En la técnica, se conoce una gran cantidad de enzimas de restricción adicionales y sus respectivas secuencias del sitio de reconocimiento de ER y están disponibles en el mercado (por ejemplo, New England Biolabs, Beverly, MA). Estas incluyen, por ejemplo, EcoRI (GAATTC) y SphI (GCATGC). Aquellos familiarizados con la técnica apreciarán que se puede incorporar cualquiera de una variedad de dichos sitios de reconocimiento de ER a determinadas realizaciones de los oligonucleótidos molde divulgados en el presente documento.

Composiciones de moldes sintéticos de control útiles para la cuantificación de una representación relativa de células inmunitarias adaptativas en una muestra biológica

Se pueden diseñar oligonucleótidos molde sintéticos de control para la cuantificación de una serie de moléculas de entrada en una muestra biológica. Estos oligonucleótidos molde sintéticos de control son similares a los oligonucleótidos molde sintéticos descritos anteriormente, pero no contienen una secuencia oligonucleotídica V o una secuencia oligonucleotídica J. Cuando se hace referencia a moldes sintéticos, a menudo, los oligonucleótidos que contienen la región V y J se denominan "primer" conjunto de moldes sintéticos, mientras que los moldes sintéticos de control se suelen denominar "segundo" conjunto de moldes sintéticos. En cambio, una composición de moldes sintéticos de control comprende una pluralidad de oligonucleótidos molde de fórmula general (II):



Los segmentos U1, B1, B2, N, B3 y U2 son los mismos que los descritos anteriormente. En una realización, X1 y X2 son nada o cada uno comprende un polinucleótido que comprende al menos 10, 20, 30 o 40, y no más de 1.000, 900 u 800 nucleótidos contiguos de una secuencia de ADN. En algunas realizaciones, la secuencia de ADN es de un gen de control genómico (también denominado "gen de control interno"), o el complemento del mismo. Como se usa en el presente documento, "gen de control genómico" o "gen de control interno" es cualquier gen que se encuentra en todas las células (incluyendo las células inmunitarias adaptativas y las células que no son células inmunitarias adaptativas), tal como un gen constitutivo como RNasa P, PSMB2, RAB7A, UBC, VCP, REEF5 o EMC7.

Los oligonucleótidos molde sintéticos de fórmula (I) se usan para determinar un número total de moléculas de receptor inmunitario adaptativo de entrada (y, por lo tanto, células inmunitarias adaptativas) de una muestra biológica. Como se explica a continuación, los oligonucleótidos molde sintéticos de control de fórmula (II) se pueden usar para determinar el número total de todos los genomas de entrada de una muestra biológica, incluyendo la muestra biológica células inmunitarias adaptativas y células que no son células inmunitarias adaptativas.

En algunas realizaciones, una composición de moldes sintéticos de control comprende una de las secuencias encontradas en SEQ ID NO: 3160-3166 y 3241-3152. Las SEQ ID NO: 3167-3194 muestran cebadores de secuenciación ilustrativos para composiciones de moldes sintéticos de control que contienen distintos segmentos de genes de control, las SEQ ID NO: 3195-3222 son secuencias de cebadores ilustrativos para secuencias adaptadoras de las composiciones de moldes sintéticos de control, y las SEQ ID NO: 3223-3236 son secuencias de cebadores ilustrativos específicas de las composiciones de moldes sintéticos de control. Las Figuras 1A y 1B ilustran un ejemplo de un oligonucleótido molde sintético de control, de acuerdo con una realización de la invención.

En determinadas realizaciones, es ventajoso que los moldes de control sintéticos de control tengan una longitud similar a la de los moldes sintéticos que contienen segmentos V y J o C de TCR y/o Ig. Además, también es ventajoso, en muchas realizaciones, que los moldes sintéticos (tanto los moldes de control como los que contienen secuencias biológicas de TCR o Ig) tengan una longitud similar al producto de amplificación de los locus de TCR/Ig y la región de control genómico de la muestra de entrada. En algunas realizaciones, la longitud de los moldes sintéticos y los amplicones correspondientes de material biológico están entre aproximadamente 100 y aproximadamente 300 nucleótidos (por ejemplo, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 o 300 nucleótidos).

MÉTODOS DE USO

I. MÉTODOS DE CUANTIFICACIÓN DE LOS LINFOCITOS INFILTRANTES TUMORALES SIN ALEATORIZADORES

a. DETERMINACIÓN DEL NÚMERO DE OLIGONUCLEOTIDOS MOLDE SINTÉTICOS DE ENTRADA EN UNA MUESTRA Y DE UN FACTOR DE AMPLIFICACIÓN USANDO UNA DILUCIÓN LIMITANTE DE

OLIGONUCLEOTIDOS MOLDE SINTÉTICOS

En algunas realizaciones, los métodos de la invención comprenden determinar el número de oligonucleótidos molde sintéticos de entrada de una muestra y determinar el factor de amplificación para cada molde sintético único.

5 Los métodos incluyen etapas para determinar una serie de oligonucleótidos molde sintéticos añadidos a una muestra inicial para su uso en PCR. El número de oligonucleótidos molde sintéticos de entrada puede estimarse usando una dilución limitante de oligonucleótidos molde sintéticos en un ensayo de PCR multiplexada. Este número de oligonucleótidos molde sintéticos de entrada de un ensayo de PCR y el número de lecturas de secuenciación de salida producidas a partir del ensayo de PCR se pueden usar para calcular una proporción de amplificación.

15 Se logra una dilución limitante cuando la cantidad de ADN de una muestra se diluye hasta el punto en que solo está presente en la dilución un subconjunto muy pequeño de oligonucleótidos molde sintéticos. Por ejemplo, en un grupo de 1.000 oligonucleótidos molde sintéticos únicos, la dilución limitante puede incluir solo 100 de los 1.000 oligonucleótidos molde sintéticos únicos. La mayoría de los moldes sintéticos únicos estarían ausentes en la dilución limitante. Por ejemplo, la dilución limitante puede incluir solo 100 tipos únicos de oligonucleótido molde sintético y solo 1 copia de cada oligonucleótido molde sintético único. Por lo tanto, una parte de los oligonucleótidos molde sintéticos únicos se añade como una sola copia o solo un pequeño número de copias, y el resto de los oligonucleótidos molde sintéticos del conjunto se añade a cero copias (es decir, están ausentes). En determinadas realizaciones, la dilución limitante de los oligonucleótidos molde sintéticos únicos incluye una molécula de cada oligonucleótido molde sintético único detectable. En otras realizaciones, la dilución limitante puede incluir dos moléculas de uno o más de los oligonucleótidos molde sintéticos únicos detectables. Por lo tanto, la dilución limitante incluye una concentración muy baja de oligonucleótidos molde sintéticos únicos.

25 La dilución limitante de oligonucleótidos molde sintéticos se amplifica como parte de una PCR multiplexada, y se calcula el número de tipos únicos de amplicones de oligonucleótidos molde sintéticos (que tienen una secuencia de código de barras única, por ejemplo).

30 La PCR simple permite la amplificación de cada oligonucleótido molde sintético único usando un par de cebadores de PCR para todos los moldes sintéticos del conjunto completo de oligonucleótidos molde sintéticos. La PCR simple se puede realizar en los oligonucleótidos molde sintéticos mediante el uso de cebadores universales que incluyen las secuencias adaptadoras e hibridarse con las secuencias de cebadores universales (UA (102) y UB (112), como se muestra en la Figura 1B). Después, el banco resultante de amplicones de oligonucleótidos molde sintéticos puede secuenciarse individualmente usando las secuencias adaptadoras en cada amplicón en un secuenciador, tal como un secuenciador Illumina. Este proceso permite la medición directa de la frecuencia de cada oligonucleótido molde sintético en el conjunto complejo.

40 En determinadas realizaciones, se usa una simulación *in silico* para analizar la relación entre el número de amplicones de oligonucleótidos molde sintéticos únicos secuenciados a partir de la dilución limitante usada en una reacción de PCR multiplexada y el número de entrada total estimado de oligonucleótidos molde sintéticos añadidos a dicha reacción de PCR multiplexada. La Figura 2 proporciona una simulación *in silico* de la relación que hay entre el número de tipos únicos de oligonucleótidos molde sintéticos observados (por ejemplo, secuenciados a partir de la muestra) y el número de moléculas molde sintéticas muestreadas (por ejemplo, número de oligonucleótidos molde sintéticos en la muestra inicial). Por ejemplo, si se secuencian y se observan 400 tipos únicos de oligonucleótidos molde sintéticos a partir de la muestra, se puede determinar que la muestra inicial incluía aproximadamente 500 moléculas oligonucleotídicas molde sintéticas. Por consiguiente, el número total de oligonucleótidos molde sintéticos de entrada puede determinarse a partir del número de oligonucleótidos molde sintéticos únicos observados.

50 Luego, se puede añadir una parte de este conjunto de oligonucleótidos molde sintéticos a una reacción de PCR multiplexada que comprende moléculas de ácido nucleico de TCR o Ig reordenadas biológicas obtenidas de los linfocitos en una muestra dada. El número determinado de oligonucleótidos molde sintéticos agregados ("añadidos") y la proporción de la amplificación calculada se pueden usar para determinar un número total de linfocitos de la muestra.

55 Como se describe en detalle en el presente documento, tras la caracterización del conjunto de oligonucleótidos molde sintéticos, se puede añadir una dilución limitante de este grupo a una muestra biológica para determinar el número de linfocitos B o T presentes en dicha muestra biológica. Se determina un factor de amplificación basado en el número de oligonucleótidos molde sintéticos de una muestra inicial de oligonucleótidos molde sintéticos que se ha añadido a una muestra biológica.

60 El factor de amplificación se calcula comparando el número de lecturas de secuenciación total para los oligonucleótidos molde sintéticos observados en una muestra con el número total de oligonucleótidos molde sintéticos de entrada en la muestra. El factor de amplificación se puede usar para determinar el número total de linfocitos (linfocitos T o linfocitos B) de una muestra biológica. Se puede suponer que este factor de amplificación se aplica a los moldes biológicos (por ejemplo, moléculas de ácido nucleico de TCR o Ig reordenadas) que se han

amplificado con los mismos cebadores específicos del segmento V y del segmento J usados para amplificar las moléculas oligonucleotídicas molde sintéticas.

5 En una realización, el factor de amplificación (la proporción) del número de lecturas de secuenciación de amplicones de oligonucleótidos molde sintéticos con respecto al número de moléculas oligonucleotídicas molde sintéticas de entrada total se compara con el número de lecturas de secuenciación totales de amplicones de moléculas biológicas para calcular el número inicial de moléculas biológicas de entrada. Por lo tanto, el cálculo del número de moléculas oligonucleotídicas molde sintéticas al comienzo del ensayo de PCR se puede usar en los cálculos de la representación relativa de las células inmunitarias adaptativas en la muestra, como se describe en detalle más adelante.

b. MÉTODOS DE DETERMINACIÓN DE LA REPRESENTACIÓN ABSOLUTA DE CÉLULAS INMUNITARIAS ADAPTATIVAS EN UNA MUESTRA

15 Se proporcionan métodos para determinar la representación absoluta de secuencias codificantes del receptor inmunitario adaptativo reordenadas en una muestra.

20 Los métodos incluyen la extracción de moléculas biológicas de ácido nucleico (por ejemplo, moléculas de ADN de TCR o Ig reordenadas) de una muestra biológica que comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas. Las moléculas biológicas de ácido nucleico de la muestra se "añaden" con una cantidad conocida de oligonucleótidos molde sintéticos (por ejemplo, como se ha descrito anteriormente en el Apartado I. a., y que se determina mediante dilución limitante). Los oligonucleótidos molde sintéticos comprenden las mismas combinaciones de secuencias oligonucleotídicas del segmento V y del segmento J que los moldes biológicos de moléculas de ácido nucleico.

25 En determinadas realizaciones, el método de cuantificación del número absoluto de moléculas de ADN reordenadas codificantes de una pluralidad de receptores inmunitarios adaptativos en una muestra biológica de un sujeto comprende las siguientes etapas:

30 I. Amplificar, en un ensayo de PCR multiplexada, un subconjunto de moléculas oligonucleotídicas molde sintéticas obtenidas de un conjunto de oligonucleótidos molde sintéticos, estando el subconjunto de moléculas oligonucleotídicas molde sintéticas diluido de manera que solo esté presente una única copia o un pequeño número de copias de una parte de oligonucleótidos molde sintéticos únicos. Los oligonucleótidos molde sintéticos amplificados se secuencian, y se determina el número de oligonucleótidos molde sintéticos únicos basado en secuencias de códigos de barras únicas. El número de lecturas de secuenciación totales de los oligonucleótidos molde sintéticos también se determina a partir del resultado de la secuenciación. A continuación, se hace referencia a los resultados de una simulación *in silico* basada en la caracterización previa del conjunto de oligonucleótidos molde sintéticos (por PCR simple) para la determinación a partir del número de secuencias únicas de oligonucleótidos molde sintéticos, el número de entrada total de moléculas oligonucleotídicas molde sintéticas (por ejemplo, en función de la relación que se muestra en la Figura 2). Se determina un factor de amplificación a partir de la proporción del resultado total de las lecturas de secuenciación de la muestra y del número total estimado de oligonucleótidos molde sintéticos de entrada. Este factor de amplificación se puede usar para estimar el número total de moléculas biológicas reordenadas y, por tanto, el número total de células linfoides, de una muestra dada. Esto se puede hacer mediante la agregación ("adición") de una pequeña parte del conjunto de oligonucleótidos molde sintéticos diluidos a la PCR multiplexada.

45 II. Amplificar las moléculas de ácido nucleico obtenidas de una muestra dada, en una PCR multiplexada, usando un conjunto de cebadores de amplificación de oligonucleótidos que comprende cebadores del segmento V y del segmento J como se describe en el presente documento que son capaces de amplificar esencialmente todas las combinaciones del segmento V y segmento J de receptores inmunitarios adaptativos reordenados, comprendiendo la muestra i) moléculas de ácido nucleico del receptor inmunitario adaptativo de TCR o Ig biológicas reordenadas, cada una de las cuales comprende una región V y una región J, e ii) una parte de oligonucleótidos molde sintéticos "añadidos" como se ha descrito anteriormente que tiene una cantidad de entrada conocida, generando así amplicones que comprenden una pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados de forma única y una pluralidad de amplicones molde sintéticos.

50 III. Secuenciar cuantitativamente la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados de manera única y una pluralidad de amplicones de moldes sintéticos generados en (I) para determinar el número total de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados observados por secuenciación (denominado en el presente documento A_i) y el número total de amplicones de moldes sintéticos observados por secuenciación (denominado en el presente documento A_{ii}). La información de secuenciación incluye el número de productos de secuenciación de salida de la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados (A_i) y el número de productos de secuenciación de salida de los amplicones de moldes sintéticos (A_{ii}).

65 IV. Determinar una representación absoluta de células inmunitarias adaptativas en la muestra en función de la información de secuenciación cuantitativa determinada a partir de la etapa II.

Para determinar la representación absoluta de las células inmunitarias adaptativas, primero se calcula un factor de amplificación. El factor de amplificación es la proporción del número de productos de secuenciación de salida de los amplicones de moldes sintéticos (A_{ij}) con el número conocido de oligonucleótidos molde sintéticos de entrada (denominado en el presente documento como A_{iii}). El número de oligonucleótidos molde sintéticos de entrada se puede determinar basándose en la simulación *in silico* realizada en (I) para determinar la relación entre el número de amplicones de oligonucleótidos molde sintéticos únicos y el número de entrada total de oligonucleótidos molde sintéticos. Se supone que el factor de amplificación de un determinado conjunto de cebadores para un oligonucleótido molde sintético es el mismo factor de amplificación para el molde biológico.

Factor de amplificación = A_{ij}/A_{iii} = número de productos de secuenciación de salida de los amplicones de moldes sintéticos/número conocido de oligonucleótidos molde sintéticos de entrada.

Al calcular este factor de amplificación, se supone que la proporción del número de lecturas de secuenciación de salida por molécula de entrada es la misma para una molécula de oligonucleótido molde sintético y una molécula de ácido nucleico de receptor inmunitario adaptativo de TCR o Ig reordenada biológica.

Después de calcular el factor de amplificación, se puede determinar el número total de moléculas de receptor inmunitario adaptativo de TCR o Ig reordenadas en la muestra y, por consiguiente, el número total de linfocitos.

En una realización, el número de moléculas biológicas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos está determinado por lo siguiente:

Número de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos = $A_i / (A_{ij}/A_{iii})$ = (Número de productos de secuenciación de salida determinados a partir de la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados)/(Factor de amplificación)

El número total de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos es igual al número total de células inmunitarias adaptativas (por ejemplo, linfocitos T o linfocitos B) en la muestra. Por consiguiente, se puede determinar el número total de células inmunitarias adaptativas de la muestra.

C. DETERMINACIÓN DE LA REPRESENTACIÓN RELATIVA DE CÉLULAS INMUNITARIAS ADAPTATIVAS EN UNA MEZCLA COMPLEJA DE CÉLULAS

Los métodos de la invención incluyen determinar una representación relativa de células inmunitarias adaptativas de una mezcla compleja de células que incluyen células inmunitarias adaptativas y células que no son células inmunitarias adaptativas. En algunas realizaciones, el número total de células inmunitarias adaptativas se determina como se describe en el Apartado I. b. y luego se usa para calcular la representación relativa de las células inmunitarias adaptativas en la muestra total de células.

El número total de moléculas de ácido nucleico reordenadas codificantes de los receptores inmunitarios adaptativos (o el número total de células inmunitarias adaptativas) se usa para determinar la representación relativa de las células inmunitarias adaptativas en la mezcla compleja. En una realización, la masa total de ADN de la muestra se usa para cuantificar el número total de genomas (células inmunitarias adaptativas y células inmunitarias no adaptativas) de la mezcla compleja. Suponiendo que cada célula tiene aproximadamente 6,5 picogramos de ADN y dada una masa total conocida de ADN de entrada en el ensayo de PCR, el número total de células inmunitarias adaptativas totales y células inmunitarias no adaptativas de la muestra se cuantifica dividiendo la masa total conocida de ADN de entrada entre 6,5 picogramos. Esto produce la representación relativa de las células inmunitarias adaptativas en la mezcla compleja de células que incluyen células inmunitarias adaptativas y células que no son células inmunitarias adaptativas.

En otras palabras, la representación relativa de las células inmunitarias adaptativas = número total de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos/(masa total de ADN que representa las células inmunitarias adaptativas y las células inmunitarias no adaptativas).

Se pueden usar otros cálculos distintos conocidos por los expertos en la materia para determinar la representación relativa de las células inmunitarias adaptativas en una mezcla compleja.

II. MÉTODOS DE CUANTIFICACIÓN DE LOS LINFOCITOS INFILTRANTES TUMORALES EN UNA MUESTRA USANDO ALEATORIZADORES Y GENES DE CONTROL

a. DETERMINACIÓN DEL NÚMERO DE OLIGONUCLEÓTIDOS MOLDE SINTÉTICOS DE ENTRADA EN UNA MUESTRA Y DE UN FACTOR DE AMPLIFICACIÓN

En algunas realizaciones, los métodos de la invención incluyen determinar una serie de oligonucleótidos molde sintéticos añadidos a una muestra de entrada para su uso en PCR. El número de oligonucleótidos molde sintéticos de entrada puede estimarse amplificando un conjunto de oligonucleótidos molde sintéticos que incluye

aleatorizadores en un ensayo de PCR multiplexada y contando el número de aleatorizadores únicos observados en la salida de secuenciación de nucleótidos. Dado que cada oligonucleótido molde sintético de la muestra de entrada tiene una secuencia aleatorizadora única, cada secuencia aleatorizadora única representa una sola molécula, y la cantidad de aleatorizadores únicos del resultado de la secuenciación de nucleótidos representa la cantidad de oligonucleótidos molde sintéticos de entrada.

En una realización, el número de entrada de oligonucleótidos molde sintéticos asociados con un determinado código de barras (y, por lo tanto, asociado con una determinada combinación por parejas de una secuencia oligonucleotídica V y una secuencia oligonucleotídica J) también puede determinarse contando el número de secuencias de ADN aleatorias únicas asociadas con un determinado código de barras. El número cuantificado de oligonucleótidos molde sintéticos de entrada y el número total de lecturas de secuenciación de salida producidas a partir del ensayo de PCR se pueden usar para calcular un factor de amplificación.

Se determina el factor de amplificación basado en el número de oligonucleótidos molde sintéticos de una muestra inicial de oligonucleótidos molde sintéticos que se ha añadido a una muestra biológica. El factor de amplificación se calcula comparando el número de lecturas de secuenciación totales para los oligonucleótidos molde sintéticos observados a partir de una muestra con el número total de oligonucleótidos molde sintéticos de entrada de la muestra, y se puede usar para determinar el número de linfocitos T totales (linfocitos T o B) de una muestra biológica. Se puede suponer que este factor de amplificación se aplica a los moldes biológicos (por ejemplo, moléculas de ácido nucleico de TCR o Ig reordenadas) que se han amplificado con los mismos cebadores específicos del segmento V y del segmento J usados para amplificar las moléculas oligonucleotídicas molde sintéticas. En una realización, el factor de amplificación también se puede definir como aquel basado en el número de oligonucleótidos molde sintéticos asociados con una determinada combinación por parejas de una secuencia oligonucleotídica V y una secuencia oligonucleotídica J en una muestra biológica. Por lo tanto, el factor de amplificación se calcula comparando el número de lecturas de secuenciación totales para los oligonucleótidos molde sintéticos asociados con la combinación por parejas en particular, con el número total de oligonucleótidos molde sintéticos de entrada asociados con la combinación por parejas particular en la muestra. Este factor de amplificación se puede usar para determinar el número de linfocitos totales (linfocitos T o linfocitos B) que portan la combinación por parejas en particular de una muestra biológica.

En una realización, el factor de amplificación (la proporción) del número de lecturas de secuenciación de amplicones de oligonucleótidos molde sintéticos con respecto al número de moléculas oligonucleotídicas molde sintéticas de entrada total se compara con el número de lecturas de secuenciación totales de amplicones de moléculas biológicas para calcular el número inicial de moléculas biológicas de entrada. El número de moléculas oligonucleotídicas molde sintéticas al comienzo del ensayo de PCR se puede usar en los cálculos de la representación relativa de las células inmunitarias adaptativas en la muestra, como se describe en detalle más adelante.

b. MÉTODOS DE DETERMINACIÓN DE LA REPRESENTACIÓN ABSOLUTA DE CÉLULAS INMUNITARIAS ADAPTATIVAS EN UNA MUESTRA

Se proporcionan métodos para determinar la representación absoluta de secuencias codificantes del receptor inmunitario adaptativo reordenadas en una muestra.

Los métodos de la invención incluyen la extracción de moléculas biológicas de ácido nucleico (por ejemplo, moléculas de ADN de TCR o IG reordenadas) de una muestra biológica que comprende células inmunitarias adaptativas y células que no son células inmunitarias adaptativas. Las moléculas biológicas de ácido nucleico de la muestra reciben la "adición" de oligonucleótidos molde sintéticos. Los oligonucleótidos molde sintéticos comprenden las mismas combinaciones de secuencias oligonucleotídicas del segmento V y del segmento J que los moldes biológicos de moléculas de ácido nucleico.

En determinadas realizaciones, el método de cuantificación del número absoluto de moléculas de ADN reordenadas codificantes de una pluralidad de receptores inmunitarios adaptativos en una muestra biológica de un sujeto comprende las siguientes etapas:

I. Amplificar las moléculas de ácido nucleico obtenidas de una muestra dada, en una PCR multiplexada, usando un conjunto de cebadores de amplificación de oligonucleótidos que comprende cebadores del segmento V y del segmento J como se describe en el presente documento que son capaces de amplificar esencialmente todas las combinaciones del segmento V y segmento J de receptores inmunitarios adaptativos reordenados, comprendiendo la muestra i) moléculas de ácido nucleico de receptor inmunitario adaptativo de TCR o Ig reordenadas, cada una comprendiendo una región V y una región J, e ii) oligonucleótidos molde sintéticos como se ha descrito anteriormente, generando así amplicones que comprenden una pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados de forma única y una pluralidad de amplicones de moldes sintéticos.

II. Secuenciar cuantitativamente la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados de manera única y una pluralidad de amplicones de moldes sintéticos generados en (I) para determinar el número total de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados

observados por secuenciación (denominado en el presente documento A_i) y el número total de amplicones de moldes sintéticos observados por secuenciación (denominado en el presente documento A_{ii}). La información de secuenciación incluye el número de productos de secuenciación de salida de la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados (A_i) y el número de productos de secuenciación de salida de los amplicones de moldes sintéticos (A_{ii}).

III. Determinar una representación absoluta de células inmunitarias adaptativas en la muestra en función de la información de secuenciación cuantitativa determinada a partir de la etapa II.

Para determinar la representación absoluta de las células inmunitarias adaptativas, primero se calcula un factor de amplificación. El factor de amplificación es la proporción del número de productos de secuenciación de salida de los amplicones de moldes sintéticos (A_{ii}) con el número de oligonucleótidos molde sintéticos de entrada (denominado en el presente documento A_{iii}), calculado contando el número de oligonucleótidos molde sintéticos de salida que tienen secuencias aleatorizadoras únicas. Se supone que el factor de amplificación de un determinado conjunto de cebadores para un oligonucleótido molde sintético es el mismo factor de amplificación para el molde biológico.

Factor de amplificación = A_{ii}/A_{iii} = número de productos de secuenciación de salida de los amplicones de moldes sintéticos/número conocido de oligonucleótidos molde sintéticos de entrada.

Al calcular este factor de amplificación, se supone que la proporción del número de lecturas de secuenciación de salida por molécula de entrada es la misma para una molécula de oligonucleótido molde sintético y una molécula de ácido nucleico de receptor inmunitario adaptativo de TCR o Ig reordenada biológica.

Después de calcular el factor de amplificación, se puede determinar el número total de moléculas de receptor inmunitario adaptativo de TCR o Ig reordenadas en la muestra y, por consiguiente, el número total de linfocitos.

En una realización, el número de moléculas biológicas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos está determinado por lo siguiente:

Número de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos = $A_i / (A_{ii}/A_{iii})$ = (Número de productos de secuenciación de salida determinados a partir de la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados)/(Factor de amplificación).

El número total de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos es igual al número total de células inmunitarias adaptativas (por ejemplo, linfocitos T o linfocitos B) en la muestra. Por consiguiente, se puede determinar el número total de células inmunitarias adaptativas de la muestra.

c. DETERMINACIÓN DE LA REPRESENTACIÓN RELATIVA DE CÉLULAS INMUNITARIAS ADAPTATIVAS EN UNA MEZCLA COMPLEJA DE CÉLULAS

Los métodos de la invención incluyen determinar una representación relativa de células inmunitarias adaptativas de una mezcla compleja de células que incluyen células inmunitarias adaptativas y células que no son células inmunitarias adaptativas. En algunas realizaciones, el número total de células inmunitarias adaptativas se determina como se describe en el apartado anterior, y luego se usa para calcular la representación relativa de las células inmunitarias adaptativas en la muestra total de células.

En determinadas realizaciones, el método de cuantificación de la representación relativa de las células inmunitarias adaptativas en una mezcla compleja de células es similar al método descrito anteriormente para cuantificar el número absoluto de moléculas de ADN reordenadas. Sin embargo, el método de cuantificación de la representación relativa incluye el uso de oligonucleótidos molde sintéticos de control y cebadores de segmentos génicos de control para determinar el número total de genomas de entrada en una muestra biológica. El método comprende las siguientes etapas:

I. Amplificar las moléculas de ácido nucleico obtenidas de una muestra dada, en una PCR multiplexada, usando un conjunto de cebadores de amplificación de oligonucleótidos que comprende cebadores del segmento V y del segmento J como se describe en el presente documento que son capaces de amplificar esencialmente todas las combinaciones del segmento V y segmento J de receptores inmunitarios adaptativos reordenados, y un conjunto de cebadores del segmento génico de control que comprende una pluralidad de cebadores capaces de amplificar al menos una parte de un gen de control que se encuentra en las células inmunitarias adaptativas y las células que no son células inmunitarias adaptativas. La muestra dada comprende i) moléculas de ácido nucleico de receptor inmunitario adaptativo de TCR o Ig reordenadas, cada una comprendiendo una región V y una región J, ii) oligonucleótidos molde sintéticos como se ha descrito anteriormente, iii) moléculas del segmento génico de control e iv) oligonucleótidos molde sintéticos de control como se ha descrito anteriormente, generando así amplicones que comprenden una pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados de forma única, una pluralidad de amplicones de segmentos génicos de control, una pluralidad de amplicones de moldes sintéticos y una pluralidad de amplicones de moldes sintéticos de control.

II. Secuenciar cuantitativamente pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig

reordenados de forma única, la pluralidad de amplicones de segmentos génicos de control, la pluralidad de amplicones de moldes sintéticos y la pluralidad de amplicones de moldes sintéticos de control generados en (I) para determinar el número total de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados observados por secuenciación (denominado en el presente documento A_i), el número total de amplicones de moldes sintéticos observados por secuenciación (denominado en el presente documento A_{ii}), el número total de amplicones de segmentos génicos de control observados por secuenciación (denominado en el presente documento B_i) y el número total de amplicones de moldes sintéticos de control observados por secuenciación (denominados en el presente documento B_{ii}). La información de secuenciación incluye el número de productos de secuenciación de salida de la pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados (A_i), el número de productos de secuenciación de salida de los amplicones de moldes sintéticos (A_{ii}), el número de productos de secuenciación de salida de los amplicones de segmentos génicos de control (B_i) y el número de productos de secuenciación de salida de los amplicones de moldes sintéticos de control (B_{ii}).

III. Determinar una representación absoluta de células inmunitarias adaptativas en la muestra en función de la información de secuenciación cuantitativa determinada a partir de la etapa II. Esta determinación de la representación absoluta de las células inmunitarias adaptativas se describe en detalle anteriormente.

IV. Determinar una representación absoluta de las células totales (células inmunitarias adaptativas y células que no son células inmunitarias adaptativas) en la muestra basándose en la información de secuenciación cuantitativa determinada a partir de la etapa II.

Para determinar la representación absoluta de las células totales, primero se calcula un factor de amplificación de control. El factor de amplificación de control es la proporción del número de productos de secuenciación de salida de los amplicones de moldes sintéticos de control (B_{ii}) con el número de oligonucleótidos molde sintéticos de control de entrada (denominado en el presente documento B_{iii}), calculado contando el número de oligonucleótidos molde sintéticos de control de salida que tienen secuencias aleatorizadoras únicas. Se supone que el factor de amplificación de control para el oligonucleótido molde sintético de control es el mismo factor de amplificación para el molde biológico.

Factor de amplificación de control = B_{ii}/B_{iii} = número de productos de secuenciación de salida de los amplicones de moldes sintéticos/número de oligonucleótidos molde sintéticos de control de entrada.

Al calcular este factor de amplificación de control, se supone que la proporción del número de lecturas de secuenciación de salida por molécula de entrada es la misma para una molécula de oligonucleótido molde sintético de control y una molécula biológica de segmentos génicos de control.

Después de calcular el factor de amplificación de control, se puede determinar el número total de genomas (o de células) de la muestra.

En una realización, el número de genomas de entrada totales está determinado por lo siguiente:

Número de genomas de entrada totales = $B_i/(B_{ii}/B_{iii})$ = (Número de productos de secuenciación de salida determinados a partir de los amplicones de segmentos génicos de control)/(Factor de amplificación de control)

El número total de genomas de entrada es el número total de células (células inmunitarias adaptativas y células que no son células inmunitarias adaptativas) en la muestra. Por consiguiente, se puede determinar el número total de células de la muestra.

V. Determinar la representación relativa de las células inmunitarias adaptativas en la mezcla compleja de células basándose en la representación absoluta de las células inmunitarias adaptativas en la muestra y la representación absoluta de las células totales en la muestra. El número total de células inmunitarias adaptativas se compara con el número total de células. En una realización, la representación relativa de las células inmunitarias adaptativas se determina dividiendo el número total de células inmunitarias adaptativas de la muestra entre el número total de células de la muestra.

En otra realización, la masa total de ADN de la muestra se usa para cuantificar el número total de células inmunitarias adaptativas y células inmunitarias no adaptativas de la mezcla compleja. Suponiendo que cada célula tiene aproximadamente 6,5 picogramos de ADN y dada una masa total conocida de ADN de entrada en el ensayo de PCR, el número total de células inmunitarias adaptativas totales y células inmunitarias no adaptativas de la muestra se cuantifica dividiendo la masa total conocida de ADN de entrada entre 6,5 picogramos. Esto produce la representación relativa de las células inmunitarias adaptativas en la mezcla compleja de células que incluyen células inmunitarias adaptativas y células que no son células inmunitarias adaptativas.

En otras palabras, la representación relativa de las células inmunitarias adaptativas = número total de moléculas de ácido nucleico reordenadas codificantes de receptores inmunitarios adaptativos/(masa total de ADN que representa las células inmunitarias adaptativas y las células inmunitarias no adaptativas).

Se pueden usar otros cálculos distintos conocidos por los expertos en la materia para determinar la representación relativa de las células inmunitarias adaptativas en una mezcla compleja.

d. Métodos de determinación de una proporción de linfocitos T o linfocitos B en una muestra en relación con el número total de genomas de entrada usando moldes sintéticos y uno o más genes de control (regiones de control genómico)

Los métodos de la invención incluyen etapas para determinar una proporción de linfocitos T o linfocitos B de una muestra con respecto al número total de genomas de entrada contenidos en dicha muestra, que comprenden:

A) amplificar por PCR multiplexada y secuenciar:

- 5 i) secuencias oligonucleotídicas de CDR3 reordenadas de locus de receptores de linfocitos T (TCR) de linfocitos T o locus de inmunoglobulinas (Ig) de linfocitos B en dicha muestra para obtener un número total de secuencias biológicas de salida, comprendiendo cada secuencia oligonucleotídica un segmento V y un segmento J;
- 10 ii) un primer conjunto de moldes sintéticos que representa esencialmente todas las combinaciones posibles de segmento V y segmento J, y cada una de ellas comprende un segmento V de TCR o Ig y un segmento J o C de TCR o Ig y un código de barras único que identifica dicho molde sintético como sintético, y donde cada molde sintético comprende una combinación única de un segmento V y un segmento J o C;

15 B) determinar un factor de amplificación para cada molde sintético que comprende una combinación única de un segmento V y un segmento J o C, donde dicho factor de amplificación está representado por un número total de primeros moldes sintéticos observados en la etapa A (ii) dividido entre un número total de entrada de primeros moldes sintéticos únicos de entrada en la etapa A (ii);

C) determinar el número total de linfocitos T o linfocitos B de la muestra dividiendo el número total de secuencias biológicas de salida observadas en la etapa A (i) entre el factor de amplificación de la etapa B;

20 D) amplificar por PCR multiplexada y secuenciar:

- i) una o más regiones de control genómico del ADN obtenido de dicha muestra para obtener un número total de secuencias biológicas de salida para cada región de control genómico; e
- 25 ii) un segundo conjunto de moldes sintéticos que comprende la secuencia de una o más de dichas secuencias de control genómico, un código de barras único y un tramo de ácidos nucleicos aleatorios, donde cada molde sintético se representa solo una vez;

E) determinar un factor de amplificación para cada una de dichas regiones de control genómico dividiendo el número total de segundos moldes sintéticos amplificados y secuenciados en la etapa D (ii) entre el número total de entrada de los segundos moldes sintéticos únicos amplificados y secuenciados en la etapa D(ii);

30 F) determinar el número total de genomas de entrada dividiendo el número total de secuencias biológicas de salida para cada región de control genómico de la etapa D (i) entre el factor de amplificación correspondiente para esa región de control genómico de la etapa E; y

35 G) determinar la proporción de linfocitos T o linfocitos B contenidos en la muestra con respecto al número de genomas totales de la muestra dividiendo el número total de linfocitos T o linfocitos B obtenidos en la etapa C entre el número total de genomas de entrada obtenidos en la etapa F.

40 En algunas realizaciones, el método comprende amplificar mediante PCR multiplexada y secuenciar dos, tres, cuatro, cinco o más regiones de control genómico en la etapa D (i). En determinadas realizaciones, cinco regiones de control genómico se amplifican y se secuencian en la etapa D(i).

45 En otras realizaciones, el número total de genomas de entrada se calcula en la etapa F tomando una media usando cada uno de los cinco factores de amplificación determinados para cada una de dichas cinco regiones de control genómico. En otra realización, se descarta el mayor y menor número calculado de genomas de entrada antes de tomar dicha media.

Los ejemplos de regiones de control genómico son PPIA, PSMB2, RAB7A, UBC, VCP, REEP5 y EMC7, o cualquier otro gen que tenga un número de copias conocido o predecible (véase la SEQ ID NO: 3160-3166 y 3241-3252 para una lista más amplia, aunque no exhaustiva de regiones de control genómico).

50 En alguna realización, el número total de moldes sintéticos en dicho primer conjunto de moldes sintéticos sujeto a amplificación en la etapa A (ii) se determina usando una dilución limitante de dichos moldes sintéticos, comprendiendo cada uno de ellos una región V y J o C de TCR o Ig única de manera que cada molde sintético único se encuentra en una sola copia.

55 En otras realizaciones, el número total de moldes sintéticos de dicho primer conjunto de moldes sintéticos sujeto a amplificación en la etapa A(ii) se determina contando el número de moldes sintéticos únicos basados en los nucleótidos aleatorios únicos contenidos en cada molde sintético.

60 **I. MÉTODOS PARA EL DIAGNÓSTICO, LA PREVENCIÓN O EL TRATAMIENTO DE ENFERMEDADES EN PACIENTES BASÁNDOSE EN LA DETERMINACIÓN DE LA REPRESENTACIÓN RELATIVA DE CÉLULAS INMUNITARIAS ADAPTATIVAS EN LA MUESTRA DE UN PACIENTE**

65 Se proporcionan métodos en el presente documento para determinar un curso de tratamiento para un paciente que lo necesita, que comprenden cuantificar la representación relativa de linfocitos infiltrantes tumorales o linfocitos infiltrantes en tejido somático que es la diana de una reacción autoinmunitaria, usando los métodos descritos en el presente documento. A este respecto, el paciente que lo necesita puede ser un paciente con cáncer o un paciente

que tenga una enfermedad autoinmunitaria. En determinados casos, un paciente puede tener un cáncer que incluye, pero sin limitación, cáncer colorrectal, hepatocelular, de vesícula biliar, pancreático, esofágico, pulmón, de mama, de próstata, de piel (por ejemplo, melanoma), de cabeza y cuello, carcinoma de células renales, de ovario, endometrial, de cuello de útero, de vejiga y urotelio. En otros determinados casos, un paciente puede tener un trasplante de

5 órgano, tal como un trasplante de hígado, un trasplante de pulmón, un trasplante de riñón, un trasplante de corazón, un trasplante de bazo, un trasplante de páncreas, un trasplante de piel/injerto, un trasplante de intestino y un trasplante de timo.

Las enfermedades autoinmunitarias incluyen, pero sin limitación, artritis (incluyendo artritis reumatoide, artritis reactiva), lupus eritematoso sistémico (LES), psoriasis, enfermedad inflamatoria intestinal (EII) (incluyendo colitis ulcerosa y enfermedad de Crohn), encefalomielitis, uveítis, miastenia grave, esclerosis múltiple, diabetes insulino dependiente, enfermedad de Addison, enfermedad celíaca, síndrome de fatiga crónica, hepatitis autoinmunitaria, alopecia autoinmunitaria, espondilitis anquilosante, fibromialgia, pénfigo vulgar, síndrome de Sjögren, enfermedad de Kawasaki, hipertiroidismo/enfermedad de Graves, hipotiroidismo/enfermedad de Hashimoto, endometriosis, esclerodermia, anemia perniciosa, síndrome de Goodpasture, síndrome de Guillain-Barre, enfermedad de Wegener, glomerulonefritis, anemia aplásica (incluyendo pacientes con anemia aplásica transfundida de forma múltiple), hemoglobinuria paroxística nocturna, púrpura trombocitopénica idiopática, anemia hemolítica autoinmunitaria, síndrome de Evan, síndrome del inhibidor del factor VIII, vasculitis sistémica, dermatomiositis, polimiositis y fiebre reumática, síndrome linfoproliferativo autoinmunitario (SLPS), penfigoide ampolloso autoinmunitario, enfermedad de Parkinson, sarcoidosis, vitíligo, cirrosis biliar primaria y miocarditis autoinmunitaria.

10
15
20

Los métodos descritos en el presente documento pueden usarse para enumerar la presencia relativa de linfocitos infiltrantes tumorales o de linfocitos infiltrantes en tejido somático que es la diana de una reacción autoinmunitaria, basados en la cuantificación de la representación relativa del ADN de dichas células inmunitarias adaptativas en el ADN extraído de una muestra biológica, que comprende una mezcla de tipos celulares, que se ha obtenido de dicho tumor o tejido. Dichos métodos son útiles para determinar un pronóstico y diagnóstico de cáncer o enfermedad autoinmunitaria, para evaluar los efectos de un tratamiento terapéutico (por ejemplo, evaluar la eficacia del fármaco y/o las relaciones dosis-respuesta), y para identificar cursos terapéuticos para el tratamiento del cáncer, para el tratamiento de enfermedades autoinmunitarias, o para el tratamiento del rechazo de trasplantes, y puede encontrar otros usos relacionados.

25
30

Para evaluar un tratamiento terapéutico, por ejemplo, ciertas realizaciones contemplan un método donde se evalúa un efecto del tratamiento terapéutico sobre la representación relativa de células inmunitarias adaptativas en al menos un tejido en un sujeto a quien se ha administrado el tratamiento. A modo de ilustración y no de limitación, de acuerdo con determinadas realizaciones de este tipo, un tratamiento que modifica (por ejemplo, aumenta o disminuye de manera estadísticamente significativa) la representación relativa de las células inmunitarias adaptativas en un tejido o tejidos puede conferir determinados beneficios al sujeto. Por ejemplo, se ha diseñado determinadas inmunoterapias contra el cáncer para aumentar la cantidad de linfocitos infiltrantes tumorales (TIL). Se ha demostrado que la presencia de TIL CD3⁺ en tumores de ovario está estrechamente correlacionada con el resultado del paciente (véase, por ejemplo, Hwang *et al.*, 2011 *Gynecol. Oncol.*, 124(2):192). Otros datos aclararon que, además de la presencia de TIL, las características de las poblaciones de TIL también fueron significativas: los TIL CD8⁺ y los TIL clónicos se asociaron con una supervivencia libre de enfermedad (SLE) más prolongada, y los linfocitos T reguladores infiltrantes se asociaron con una SLE más corta (véase, Stumpf *et al.*, 2009 *Br. J. Cancer* 101:1513-21). Estos estudios indicaron que los TIL pueden ser un factor de pronóstico independiente (véase, Clarke *et al.*, 2009 *Mod. Pathol.* 22:393-402). Por lo tanto, la cuantificación de la representación relativa del ADN de células inmunitarias adaptativas como se describe en el presente documento, con el fin de detectar posibles aumentos en los TIL en muestras de tejido tumoral obtenidas en uno o varios puntos de tiempo antes del tratamiento, durante el curso del tratamiento y/o después del tratamiento puede proporcionar información muy útil con respecto a la determinación de la eficacia del tratamiento y, a partir de ahí, desarrollar un pronóstico para el sujeto.

35
40
45
50

Como otro ejemplo, determinadas inmunoterapias dirigidas a enfermedades autoinmunitarias están diseñadas para reducir la cantidad de linfocitos infiltrantes de tejidos en uno o más tejidos afectados, tales como tejidos u órganos que pueden ser la diana de un ataque autoinmunitario clínicamente inapropiado, de modo que la cuantificación de la representación relativa del ADN de células inmunitarias adaptativas como se describe en el presente documento, con el fin de detectar posibles disminuciones en las células inmunitarias adaptativas en muestras de tejido obtenidas en uno o varios puntos de tiempo antes del tratamiento, durante el curso del tratamiento y/o después del tratamiento puede proporcionar información muy útil con respecto a la determinación de la eficacia del tratamiento y, a partir de ahí, desarrollar un pronóstico para el sujeto.

55

Como un ejemplo adicional, determinadas inmunoterapias dirigidas al rechazo de trasplantes están diseñadas para reducir la cantidad de linfocitos infiltrantes de tejidos en los órganos trasplantados, de modo que la cuantificación de la representación relativa del ADN de células inmunitarias adaptativas como se describe en el presente documento, con el fin de detectar posibles disminuciones en las células inmunitarias adaptativas en muestras de tejido de órganos trasplantados obtenidos en uno o una pluralidad de puntos de tiempo antes del tratamiento, durante el curso del tratamiento y/o después del tratamiento puede proporcionar información muy útil con respecto a la determinación de la eficacia del tratamiento y, a partir de ahí, desarrollar un pronóstico para el sujeto.

60
65

En estas y otras realizaciones relacionadas, los métodos descritos en el presente documento de cuantificación de la representación relativa del ADN de células inmunitarias adaptativas se pueden poner en práctica usando muestras biológicas de prueba obtenidas de un sujeto en uno o una pluralidad de puntos de tiempo antes de administrar el tratamiento terapéutico al sujeto, y en uno o una pluralidad de puntos de tiempo después de administrar el tratamiento terapéutico al sujeto. Las muestras se pueden obtener del mismo tejido o de diferentes tejidos, que puede variar en función del estado particular del sujeto. Por ejemplo, a modo de ilustración y no de limitación, en el caso de un tumor inoperable, las muestras biológicas de prueba que se obtienen del sujeto antes y después del tratamiento pueden ser del mismo tejido, mientras que, en el caso de un tumor que se extirpa parcialmente quirúrgicamente, o que se produce en múltiples sitios en el sujeto, las muestras biológicas de prueba se pueden obtener de diferentes tejidos o de diferentes sitios de tejidos antes y después de administrar el tratamiento terapéutico.

También se contemplan en el presente documento realizaciones en las que cualquiera de los métodos descritos en el presente documento puede comprender además la determinación de la diversidad estructural relativa de los receptores inmunitarios adaptativos (por ejemplo, la diversidad de secuencia entre productos de genes de TCR y/o de inmunoglobulina reordenados productivamente) en el componente de células inmunitarias adaptativas de la mezcla de células que está presente en la muestra biológica de prueba. En determinadas realizaciones de este tipo, las actuales metodologías de qPCR que usan el receptor inmunitario adaptativo reordenado descrito en el presente documento codificante de conjuntos de cebadores oligonucleotídicos específicos permiten la identificación fácil de las combinaciones de cebadores particulares que generan la producción de moléculas de ADN reordenadas amplificadas. Por consiguiente, por ejemplo, estas realizaciones permiten la determinación del grado relativo de clonalidad de una población de células inmunitarias adaptativas que está presente como parte de una población de células mixtas en una muestra biológica de prueba, que puede tener valor de pronóstico.

Por ejemplo, en una muestra de tumor sólido donde se detectan TIL cuantificando la representación relativa del ADN de células inmunitarias adaptativas en el ADN extraído de la muestra como se describe en el presente documento, los presentes métodos contemplan la determinación de si solo una o unas cuantas (por ejemplo, no más de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10) combinaciones de un cebador oligonucleotídico del segmento V particular y un cebador oligonucleotídico del segmento J particular son predominantemente (por ejemplo, generan al menos el 80, 85, 90, 95, 97 o 99 por ciento de los productos de amplificación) responsables de la producción por PCR de moléculas de ADN de células inmunitarias adaptativas reordenadas amplificadas. Dicha observación de uno o unos cuantos productos de amplificación codificantes de genes del receptor inmunitario adaptativo predominante, de acuerdo con la teoría no limitante, indican un bajo grado de heterogeneidad de los TIL. Por el contrario, la determinación de un alto grado de heterogeneidad en la diversidad estructural de receptores inmunitarios adaptativos mediante la caracterización del ADN de TIL indicaría que no está presente un clon de TIL predominante.

Por consiguiente, en el presente documento, se describen métodos para medir el número de células inmunitarias adaptativas (por ejemplo, linfocitos T) en una mezcla compleja de células. Los presentes métodos tienen una utilidad particular en la cuantificación de linfocitos infiltrantes tumorales o linfocitos infiltrantes en tejido somático que es la diana de una respuesta autoinmunitaria. Los métodos existentes para la cuantificación de linfocitos T y B dependen de la separación física de dichas células de la mezcla. Sin embargo, en muchos casos, los linfocitos T y B no pueden separarse de la muestra inicial, tal como muestras de tejido congeladas o fijadas con formalina. Además, los métodos anteriores para la cuantificación de las células inmunitarias adaptativas (por ejemplo, inmunocitofluorimetría de flujo, análisis de selección de células activadas por fluorescencia (FACS), inmunohistoquímica (IHC, ImmunoHistoChemistry) se basa en la expresión de proteínas específicas de linfocitos T o linfocitos B, tales como los receptores de la superficie celular. Dado que las células inmunitarias expresan cantidades variables de estos receptores específicos del linaje, la cuantificación del número de células de una medida tan altamente variable requiere una estandarización costosa, un equipo especializado y un personal altamente capacitado. Los métodos divulgados en el presente documento son, por el contrario, independientes de la plataforma y se pueden realizar en cualquier instrumento de PCR e instrumento de secuenciación de alto rendimiento, y los reactivos se pueden sintetizar y proporcionar en forma de kit. Los métodos divulgados en el presente documento también son muy sensibles y pueden aplicarse en configuraciones de alto rendimiento que no se podían obtener previamente. Como se describe en el presente documento, la cuantificación de las células inmunitarias adaptativas puede lograrse mediante una simple preparación de ADN a partir de una mezcla compleja de células, en concierto con la cuantificación de la proporción relativa de células inmunitarias adaptativas presentes mediante la amplificación de los genes codificantes de CDR3 de células inmunitarias adaptativas reordenadas. _

Ciertos casos incluyen métodos para comparar cantidades de ADN de células inmunitarias adaptativas con ADN celular total (por ejemplo, de células inmunitarias adaptativas más células inmunitarias no adaptativas en la mezcla celular). Los métodos también incluyen opcionalmente comparar otros parámetros relevantes antes, durante o después de la administración a un sujeto de control de composiciones de control que pueden ser, por ejemplo, controles negativos que previamente se ha demostrado que no han sufrido alteraciones estadísticamente significativas del estado fisiológico, tal como la inyección simulada, solución salina, DMSO, u otro vehículo o control de tamponamiento, enantiómeros inactivos, péptidos o nucleótidos desordenados, etc., y/o antes, durante o después de la administración de controles positivos que previamente han demostrado causar una alteración estadísticamente

significativa del estado fisiológico, tal como un compuesto terapéutico aprobado por la FDA.

La práctica de determinadas 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 forman parte de las habilidades de la técnica, y se hace referencia a varias de ellas a continuación con fines ilustrativos. Dichas técnicas se explican por completo en la bibliografía. Véase, 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 and Wiley-Interscience; Glover, "DNA Cloning: A Practical Approach", vol. I y II (IRL Press, Oxford Univ. Press, EE.UU., 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 and Applications", Editado por Julie Logan, Kirstin Edwards y Nick Saunders, 2009, Caister Academic Press, Norfolk, RU; Anand, "Techniques for the Analysis of Complex Genomes", (Academic Press, Nueva York, 1992); Guthrie y Fink, "Guide to Yeast Genetics and Molecular Biology" (Academic Press, Nueva York, 1991); "Oligonucleotide Synthesis" (N. Gait, Ed., 1984); "Nucleic Acid Hybridization" (B. Hames y S. Higgins, Eds., 1985); "Transcription and Translation" (B. Hames y 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); "the treatise, 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, Londres, 1987); "Handbook Of Experimental Immunology", Volúmenes I-IV (D. M. Weir y C. C. Blackwell, eds., 1986); Riott, "Essential Immunology", 6ª edición, (Blackwell Scientific Publications, Oxford, 1988); "Embryonic Stem Cells: Methods and Protocols" (*Methods in Molecular Biology*) (Kurstad Turksen, Ed., 2002); "Embryonic Stem Cell Protocols: Volume I: Isolation and 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*) (Kurstad Turksen Ed., 2006); "Mesenchymal Stem Cells: Methods and 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 and Protocols" (*Methods in Molecular Biology*) (Leslie P. Weiner Ed., 2008).

A menos que se proporcionen definiciones específicas, la nomenclatura utilizada vinculada con, y los procedimientos de laboratorio y las técnicas de, biología molecular, la química analítica, la química orgánica sintética, y la química médica y farmacéutica descritas en el presente documento son aquellas bien conocidas y comúnmente usadas en la técnica. Se pueden usar técnicas convencionales de tecnología recombinante, biología molecular, microbiología, síntesis química, análisis químicos, preparaciones farmacéuticas, formulaciones y administración, y tratamiento de pacientes.

A menos que el contexto requiera otra cosa, a lo largo de la presente memoria descriptiva y de las reivindicaciones, el término "comprender" y las variaciones del mismo, tales como, "comprende" y "que comprende" deben interpretarse en un sentido abierto e inclusivo, es decir, como "que incluye/n, pero sin limitación". Por "que consiste en" se entiende que incluye, y normalmente se limita a, lo que siga a la expresión "que consiste en". Por "consistir esencialmente en" se entiende que incluye cualquier elemento enumerado detrás de la expresión y limitado a otros elementos que no interfieren con o contribuyen a la actividad o acción especificada en la divulgación de los elementos enumerados. Por lo tanto, la expresión "que consiste esencialmente en" indica que los elementos enumerados son necesarios u obligatorios, pero que se requieren otros elementos y pueden estar presentes o no, dependiendo de si afectan a la actividad o acción de los elementos enumerados.

En la presente memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un", "uno/a" y "el/la" incluyen las referencias en plural salvo que el contenido indique claramente otra cosa. Como se usa en el presente documento, en realizaciones particulares, el término "aproximadamente" cuando precede a un valor numérico indica el valor más o menos un intervalo del 5 %, 6 %, 7 %, 8 % o 9 %. En otras realizaciones, el término "aproximadamente" cuando precede a un valor numérico indica el valor más o menos un intervalo del 10 %, 11 %, 12 %, 13 % o 14 %. En otras realizaciones, el término "aproximadamente" cuando precede a un valor numérico indica el valor más o menos un intervalo del 15 %, 16 %, 17 %, 18 %, 19 % o 20 %.

La referencia a lo largo de la presente memoria descriptiva a "una realización" o "un aspecto" significa que un determinado rasgo distintivo, estructura o característica descritos junto con la realización está incluida en al menos una realización de la presente invención. Por lo tanto, las apariciones de la expresión "en una realización" en diferentes lugares a lo largo de la presente memoria descriptiva no se refieren todas necesariamente a la misma realización. Además, los rasgos distintivos, las estructuras o las características particulares pueden combinarse de

cualquier manera adecuada en una o más realizaciones.

Ejemplos

- 5 Los siguientes ejemplos se ofrecen con fines meramente ilustrativos y no se pretende limitar con ellos el alcance de la invención de modo alguno. Se han realizado esfuerzos para garantizar la precisión con respecto a los números usados (por ejemplo, cantidades, temperaturas, etc.), pero, como es evidente, se debería permitir algún error experimental y desviación.

10 **EJEMPLO 1: Construcción de moldes sintéticos y estimación del número de moldes sintéticos de entrada**

Los moldes sintéticos se diseñaron y se construyeron como se muestra en las Figuras 1A y 1B. Cada oligonucleótido molde sintético incluía una secuencia de 495 pb que comprendía una secuencia de cebador universal (UA) (102), un código de barras (CB) específico del molde de 16 pb (104), una secuencia génica codificante de la región variable (V) del receptor inmunitario adaptativo de 300 pb (gen V) (106), un marcador interno (MI) del molde sintético de 9 pb (108), una repetición del código de barras (CB) (104), una repetición del marcador interno (MI) (108), una secuencia génica codificante de la región variable (J) del receptor inmunitario adaptativo (gen J) (110) de 100 pb, una tercera repetición del código de barras (CB) (104) y una secuencia de cebador universal inverso (UB) (112). Los moldes sintéticos también se pueden diseñar como se describe en Carlson, C.S. *et al.* "Using synthetic templates to design an unbiased multiplex PCR assay". *Nature Communications* 4, 2680, doi: 10.1038/ncomms3680 (2013). Se construyeron moldes sintéticos que incluían esencialmente todas las combinaciones del segmento V y segmento J de receptores inmunitarios adaptativos reordenados. Las secuencias de los moldes sintéticos se encuentran en la Tabla A.

25 El grupo inicial de moldes sintéticos se cuantificó usando PCR simple. Se usaron cebadores universales específicos para las regiones cebadoras universales (102) y (112), y se combinaron con regiones de adaptador Illumina para amplificar y secuenciar los moldes sintéticos. Para estimar el número de lecturas de secuencia de los amplicones de moldes sintéticos generados por molde sintético de entrada, se estimó el número de moldes sintéticos de entrada añadiendo a la PCR usando una dilución limitante. Se logró una dilución limitante diluyendo la cantidad de ADN de cada muestra hasta el punto en que solo una cantidad muy pequeña de oligonucleótidos molde sintéticos mezclados estaba presente en la dilución (por ejemplo, un pequeño subconjunto de oligonucleótidos molde sintéticos únicos en una copia baja por molde). Cuando se añadió una dilución limitante de oligonucleótidos molde sintéticos al ensayo de PCR, la mayoría de los oligonucleótidos molde sintéticos únicos no se detectó mediante PCR en la muestra, y los oligonucleótidos molde sintéticos que se detectaron probablemente se añadieron como una sola copia. Por lo tanto, los resultados de la PCR simple caracterizaron con precisión la concentración de cada molde sintético único en la muestra.

La Figura 2 es un gráfico que muestra la relación entre el número de moldes sintéticos únicos de la muestra y el número total de moldes sintéticos de la muestra. Esta relación se analizó usando una simulación *in silico* de un muestreo aleatorio del conjunto de moldes sintéticos. La desviación media y típica del número de moldes sintéticos únicos observados (eje y) se presenta en función del número total de moléculas muestreadas del conjunto de oligonucleótidos molde sintéticos (eje x). El gráfico demuestra que el número detectado de oligonucleótidos molde sintéticos únicos proporciona información sobre el número total de moléculas en el conjunto. Se pueden añadir partes alícuotas de este conjunto de moldes sintéticos a PCR multiplexadas posteriores (en los siguientes ejemplos, por ejemplo) y usarse para determinar una serie de moléculas de entrada en una determinada muestra.

45 **EJEMPLO 2: Método de validación para estimar el número de moldes sintéticos de entrada**

Para validar el método descrito en el Ejemplo 1 de estimación del número de moldes sintéticos de entrada añadidos a la PCR, se usó ADN_g de fibroblastos en masa y ADN_g de linfocitos B clasificados (CD19+) para construir mezclas de 300 ng de ADN_g total, que contenían diferentes combinaciones de ADN de linfocitos B (de 0,07 ng a 300 ng) y ADN de fibroblastos. Como cada muestra contenía una proporción conocida de ADN de linfocitos B de entrada, también se conoció el número de locus de IgH reordenados de cada muestra. Este número se comparó con el número de moldes sintéticos de entrada de IgH reordenados estimado usando el método descrito en el Ejemplo 1. Se encuentran ejemplos de secuencias de moldes sintéticos de entrada de IgH reordenados en las SEQ ID NO: 1645-3003.

Este experimento de validación se repitió cuatro veces usando cuatro PCR y reacciones de secuenciación separadas. El gráfico de la Figura 3 muestra que las cuatro PCR presentaron resultados replicables, y que el método del Ejemplo 1 para estimar el número de moldes sintéticos de entrada fue un éxito, ya que los números estimados estaban de acuerdo con los números conocidos de locus de IgH reordenados.

60 **EJEMPLO 3: Determinación de una representación relativa de linfocitos T en una mezcla de células**

65 Para determinar la representación relativa de los linfocitos infiltrantes tumorales (TIL) en cuatro muestras de tejido tomadas de carcinomas de ovario seroso de alto grado, se realizó una ddPCR de TCRB (como se describe en el

documento US2013/0288237; y Robins, H.S. *et al.* "Digital genomic quantification of tumor-infiltrating lymphocytes". *Science Translational Medicine* 5, 214ra169, doi:10.1126/scitranslmed.3007247 (2013)). Además, se usaron moldes sintéticos para determinar por separado la representación relativa de los TIL. En primer lugar, se preparó una muestra que incluía tanto ADNg extraído del tejido tumoral como moldes sintéticos como se muestra en la Figura 1, que incluye esencialmente todas las combinaciones del segmento V y segmento J de receptores inmunitarios adaptativos reordenados (las secuencias de moldes sintéticos se encuentran en SEQ ID NO: 787-1644). A continuación, se realizó una PCR multiplexada en la muestra usando cebadores del segmento V y del segmento J para amplificar moléculas de ácido nucleico que comprendían i) moléculas de ácido nucleico del receptor inmunitario adaptativo de TCR o Ig reordenadas, cada una comprendiendo una región V y una región J, e ii) oligonucleótidos molde sintéticos que tenían una cantidad de entrada conocida, generando así amplicones que comprendían una pluralidad de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados y una pluralidad de amplicones de moldes sintéticos.

Los amplicones se secuenciaron para determinar el número de amplicones de receptores inmunitarios adaptativos de TCR o Ig reordenados y el número de amplicones de moldes sintéticos únicos. El factor de amplificación se determinó calculando la proporción del número de productos de secuenciación de salida de los amplicones de moldes sintéticos con el número conocido de oligonucleótidos molde sintéticos de entrada, determinado usando el método de estimación del Ejemplo 1 y descrito en el presente documento.

A continuación, el número absoluto de moléculas de receptor inmunitario adaptativo de TCR o Ig reordenadas de entrada de la muestra se determinó dividiendo el número de productos de secuenciación de salida determinado a partir de los amplicones de receptor inmunitario adaptativo adaptado de TCR o Ig reordenados entre el factor de amplificación. El número absoluto de moléculas de receptor de TCR o Ig reordenadas de entrada equivale al número total de células inmunitarias adaptativas de la muestra.

A partir de ahí, el número de células inmunitarias adaptativas totales y de células inmunitarias no adaptativas de la muestra se determinó dividiendo una masa total conocida de ADN de entrada entre 6,5 picogramos (suponiendo que cada célula tiene aproximadamente 6,5 picogramos de ADN y dada una masa total conocida de ADN de entrada para el ensayo de PCR). La representación relativa de las células inmunitarias adaptativas en la muestra se calculó dividiendo el número de células inmunitarias adaptativas únicas entre el número total de células inmunitarias adaptativas y células inmunitarias no adaptativas de la muestra.

La Figura 4 es un gráfico que muestra (1) la representación relativa de los linfocitos T en la muestra determinada usando métodos de ddPCR, como se describe en los documentos US2013/0288237 y WO2013/059725, en comparación con (2) la representación relativa de los linfocitos T de la muestra usando moldes sintéticos y los métodos descritos anteriormente. La Figura 4 muestra que se calculan números similares usando ambos métodos. Por tanto, los moldes sintéticos se pueden usar para calcular con precisión la representación relativa de las células inmunitarias adaptativas de una muestra que contiene células inmunitarias adaptativas y células inmunitarias no adaptativas.

EJEMPLO 4: Determinación de una representación relativa de linfocitos T en una mezcla de células usando regiones de control genómico

En este ejemplo, se usaron moldes sintéticos y genes de control genómico para calcular con precisión la representación relativa de las células inmunitarias adaptativas en una muestra que contenía células inmunitarias adaptativas y células inmunitarias no adaptativas.

Fuente de la muestra:

Se aislaron linfocitos T de la sangre completa usando técnicas convencionales de biología celular. Se extrajo el ADN de la población de linfocitos T purificados. Se normalizó el ADN, suponiendo 6,4 pg de ADN/genoma humano bicatenario, de modo que se añadieron aproximadamente 5 genomas, 250 genomas, 1.250 genomas o 6.250 genomas de ADN de linfocitos T a una reacción de PCR de TCRB convencional.

Reacción de PCR multiplexada:

Ensayo de TCRB: Se amplificaron genes de TCRB reordenados usando una PCR multiplexada. Se diseñaron cebadores del segmento V y del segmento J para amplificar ~110 pb de fragmentos reordenados. Se añadieron moldes sintéticos a cada reacción de PCR y se amplificaron con los mismos cebadores, incluyendo los moldes sintéticos un código de barras para diferenciarlos de los moldes biológicos. Se añadió el volumen de ADN necesario para añadir 5, 250, 1.250 y 6.250 genomas a cada reacción de PCR.

Se realizó una segunda reacción de colas de PCR usando cebadores de colas que comprendían códigos de barras muy específicos y adaptadores de secuenciación Illumina. La reacción de colas de PCR añadió códigos de barras muy específicos y adaptadores de secuenciación Illumina a cada producto de PCR.

Ensayo de control genómico:

Además del ensayo de TCRB, se amplificaron cinco locus autosómicos de copia única usando un ensayo de PCR multiplexada. Cada locus autosómico de una sola copia está presente en cada célula y sirve como control genómico.

5 Los controles genómicos se usaron para contar el número de genomas presentes en la muestra. Los cebadores se diseñaron para amplificar fragmentos de 110 pb de cada locus, que eran del mismo tamaño que los cebadores de TCRB.

10 La reacción de PCR multiplexada incluyó la amplificación conjunta de moldes sintéticos que incluían secuencias oligonucleotídicas de cada uno de los cinco genes autosómicos. Los moldes sintéticos incluían códigos de barras únicos que identificaban las moléculas como moldes sintéticos y una secuencia aleatoria de 6 pb. Se usó la misma concentración de ADN para los controles genómicos que los genes de TCRB, pero a un octavo del volumen, de modo que se añadieron menos de 1, 31, 156, 781 genomas bicatenarios a cada reacción de PCR. Se añadieron
15 códigos de barras específicos y adaptadores de secuenciación illumina a cada producto de PCR en un segundo ensayo de PCR de colas, como se ha descrito anteriormente.

Secuenciación:

20 Las muestras se agruparon, se normalizaron y se cargaron en un Illumina MiSEQ. Se procesaron los datos de las secuencias de salida, y se usaron las lecturas de secuencias de los moldes sintéticos para medir la cobertura de la secuenciación. La cobertura de la secuenciación es una estimación del número de grupos de secuenciación derivados de una sola molécula añadida a la reacción de PCR.

Análisis:

25 Se estimó el número de moléculas de TCRB de la muestra usando los métodos descritos anteriormente. Se determinó el número de genomas añadidos al ensayo de TCRB estimando el número de genomas en el ensayo de control genómico como se ha descrito previamente (Apartado III). Se aumentó a escala el número calculado de genomas del ensayo de control genómico por 4 para justificar 1) que había 2 locus/genoma y 2) la reducción de ocho
30 veces de la entrada (Figura 6).

REIVINDICACIONES

1. Un método de determinación de una proporción de linfocitos T o linfocitos B en una muestra en relación con el número total de genomas de entrada contenidos en dicha muestra que comprende:

- 5 A) amplificar por PCR multiplexada y secuenciar:
 - 10 i) secuencias oligonucleotídicas de CDR3 reordenadas de locus de receptores de linfocitos T (TCR) de linfocitos T o locus de inmunoglobulinas (Ig) de linfocitos B en dicha muestra para obtener un número total de secuencias biológicas de salida, comprendiendo cada secuencia oligonucleotídica un segmento V y un segmento J;
 - 15 ii) un primer conjunto de moldes sintéticos que representan esencialmente todas las combinaciones posibles de segmentos V y segmentos J, donde cada molde sintético del primer conjunto de moldes sintéticos comprende:
 - (a) una combinación única de un segmento V y un segmento J o C, donde el segmento V es un segmento V de TCR o Ig, y donde el segmento J o C es un segmento J o C de TCR o Ig; y
 - (b) un código de barras único que identifica dicho molde sintético como sintético;
- 20 B) determinar un factor de amplificación para cada molde sintético que comprende una combinación única de un segmento V y un segmento J o C, donde dicho factor de amplificación está representado por el número total de primeros moldes sintéticos amplificados y secuenciados en la etapa A (ii) dividido entre el número total de entrada de los primeros moldes sintéticos únicos introducidos en la etapa A (ii);
- 25 C) determinar el número total de linfocitos T o linfocitos B de la muestra dividiendo el número total de secuencias biológicas de salida observadas en la etapa A (i) entre el factor de amplificación de la etapa B;
- D) amplificar por PCR multiplexada y secuenciar:
 - 30 i) una o más regiones de control genómico del ADN obtenido de dicha muestra para obtener un número total de secuencias biológicas de salida para cada región de control genómico; y
 - 35 ii) un segundo conjunto de moldes sintéticos, donde cada molde sintético del segundo conjunto de moldes sintéticos comprende la secuencia de una o más de dichas secuencias de control genómico, un código de barras único y un tramo de ácidos nucleicos aleatorios, y donde cada molde sintético del segundo conjunto de moldes sintéticos se representa solo una vez;
- 40 E) determinar un factor de amplificación para cada una de dichas regiones de control genómico dividiendo el número total de segundos moldes sintéticos amplificados y secuenciados en la etapa D (ii) entre el número total de entrada de los segundos moldes sintéticos únicos introducidos en la etapa D (ii);
- F) determinar el número total de genomas de entrada dividiendo el número total de secuencias biológicas de salida para cada región de control genómico de la etapa D (i) entre el factor de amplificación correspondiente para esa región de control genómico de la etapa E; y
- 45 G) determinar la proporción de linfocitos T o linfocitos B contenidos en la muestra con respecto al número de genomas totales de la muestra dividiendo el número total de linfocitos T o linfocitos B obtenidos en la etapa C entre el número total de genomas de entrada obtenidos en la etapa F.

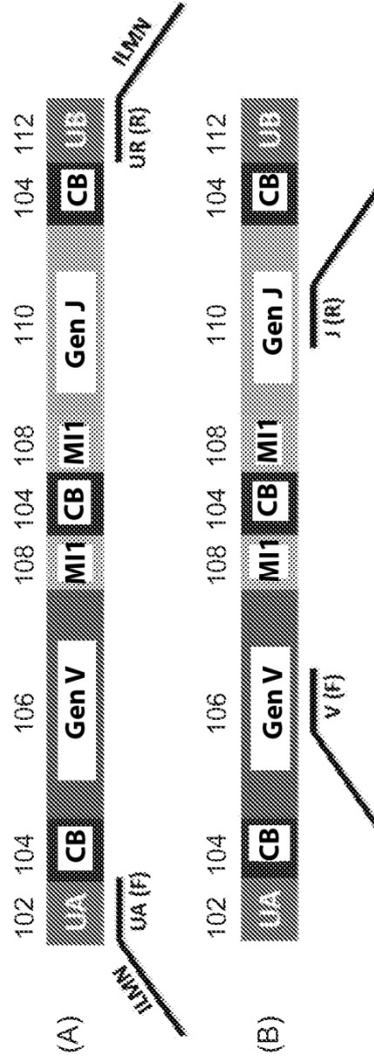
2. El método de la reivindicación 1, donde el primer conjunto de moldes sintéticos comprende la secuencia de fórmula I: 5'-U1-B1-V-B2-J-B3-U2-3', donde

- 50 A) V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica codificante de la región variable (V) de TCR o Ig, o su complemento, y teniendo cada molde de dicho primer conjunto de moldes sintéticos una secuencia oligonucleotídica de región V única;
- B) J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica codificante de la región de unión (J) de TCR o Ig, o su complemento, y comprendiendo cada molde de dicho primer conjunto de moldes sintéticos una secuencia oligonucleotídica de región J única;
- 55 C) U1 comprende una secuencia oligonucleotídica que se selecciona entre (i) una primera secuencia oligonucleotídica adaptadora universal; e (ii) una primera secuencia oligonucleotídica de plataforma de secuenciación que está unida y posicionada en 5' con respecto a una primera secuencia oligonucleotídica adaptadora universal;
- 60 D) U2 comprende una secuencia oligonucleotídica que se selecciona entre (i) una segunda secuencia oligonucleotídica adaptadora universal; e (ii) una segunda secuencia oligonucleotídica de plataforma de secuenciación que está unida y posicionada en 5' con respecto a una segunda secuencia oligonucleotídica adaptadora universal;
- E) B1, B2 y B3 comprenden cada uno independientemente nada o una secuencia de código de barras oligonucleotídica de 3-25 ácidos nucleicos que identifica de forma única, como una combinación por parejas (i) dicha secuencia oligonucleotídica de región V única; y dicho oligonucleótido de región J único, donde al menos uno de B1, B2 y B3 está presente en cada molde sintético contenido en dicho primer conjunto de oligonucleótidos.

3. El método de la reivindicación 2, donde cada uno de los moldes sintéticos contenidos en dicho primer conjunto de moldes sintéticos comprende además un tramo de nucleótidos aleatorios únicos.
- 5 4. El método de la reivindicación 3, donde el tramo aleatorio de nucleótidos comprende de 4 a 50 nucleótidos, o donde el tramo aleatorio de nucleótidos comprende 8 nucleótidos.
- 10 5. El método de una cualquiera de las reivindicaciones 1-4, donde el número total de moldes sintéticos de dicho primer conjunto de moldes sintéticos sometido a amplificación en la etapa A(ii) se determina usando una dilución limitante de dichos moldes sintéticos, comprendiendo cada uno una región única V y J o C de TCR o Ig, de modo que el número de moldes sintéticos únicos observados permite la inferencia del número total de moléculas molde sintéticas de entrada.
- 15 6. El método de una cualquiera de las reivindicaciones 1-5, donde el número total de moldes sintéticos de dicho primer conjunto de moldes sintéticos sujeto a amplificación en la etapa A(ii) se determina contando el número de moldes sintéticos únicos basados en los nucleótidos aleatorios únicos contenidos en cada molde sintético.
- 20 7. El método de una cualquiera de las reivindicaciones 1-6, que comprende amplificar mediante PCR multiplexada y secuenciar dos o más, o tres o más, o cuatro o más, o cinco o más regiones de control genómico en la etapa D(i).
8. El método de la reivindicación 1, que comprende amplificar mediante PCR multiplexada y secuenciar cinco regiones de control genómico en la etapa D(i), y donde cada una de las cinco regiones de control genómico tiene un número de copias predecible.
- 25 9. El método de la reivindicación 8, donde los factores de amplificación se determinan para cada una de dichas cinco regiones de control genómico en la etapa E.
- 30 10. El método de la reivindicación 9, donde el número total de genomas de entrada se calcula en la etapa F tomando una media usando cada uno de los cinco factores de amplificación determinados para cada una de dichas cinco regiones de control genómico.
11. El método de la reivindicación 1, donde la amplificación mediante PCR multiplexada en la etapa A o D se realiza usando una pluralidad de conjuntos de cebadores oligonucleotídicos que comprenden:
- 35 A) una pluralidad de cebadores oligonucleotídicos del segmento V que son cada uno independientemente capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región V de TCR de Ig o con su complemento, 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 codificante de la región V de TCR o Ig, y donde dicha pluralidad de cebadores del segmento V se
- 40 hibrida específicamente con esencialmente todos los segmentos génicos funcionales codificantes de la región V de TCR o Ig que están presentes en la composición, y
- 45 B) una pluralidad de cebadores oligonucleotídicos del segmento J que son cada uno independientemente capaces de hibridarse específicamente con al menos un polinucleótido codificante de un polipéptido de la región J de TCR o Ig, o con su complemento, 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 codificante de la región J de TCR o Ig funcional, y donde dicha pluralidad de cebadores del segmento J se hibrida específicamente con esencialmente todos los segmentos génicos codificantes de la región J de TCR o Ig funcionales que están presentes en la composición.
- 50 12. El método de la reivindicación 1, donde una o ambas de entre:
- (i) dicha pluralidad de cebadores oligonucleotídicos del segmento V comprende secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos establecidas en SEQ ID NO: 1-120, 147-158, 167-276, 407-578, 593-740, y
- 55 (ii) dicha pluralidad de cebadores oligonucleotídicos del segmento J comprende secuencias que tienen al menos un 90 % de identidad de secuencia con las secuencias de nucleótidos establecidas en SEQ ID NO: 121-146, 159-166, 277-406, 579-592, 741-764.
- 60 13. El método de una cualquiera de las reivindicaciones 1-12, donde dicha muestra es tejido recién extraído, tejido congelado o tejido fijo.
14. El método de la reivindicación 1, donde dichas secuencias de salida obtenidas en la etapa A(i), dichos moldes sintéticos contenidos en dicho primer conjunto de moldes sintéticos en la etapa A (ii), dichas secuencias de salida para cada región de control genómico en la etapa D(i) y dichos moldes sintéticos contenidos en dicho segundo conjunto de moldes sintéticos en la etapa D(ii) son cada uno de aproximadamente 100-300 nucleótidos de longitud.
- 65

15. El método de una cualquiera de las reivindicaciones 1-14, donde la una o más regiones de control genómico se seleccionan del grupo que consiste en ACTB, B2M, C1 orf34, CHMP2A, GPI, GUSB, HMBS, HPRT1, PSMB4, RPL13A, RPLP0, SDHA, SNRPD3, UBC, VCP, VPS29, PPIA, PSMB2, RAB7A, UBC, VCP, REEP5 y EMC7; o donde la una o más regiones de control genómico son PSMB2, RAB7A, PPIA, REEP5 y EMC7.

5



Figuras 1A y 1B

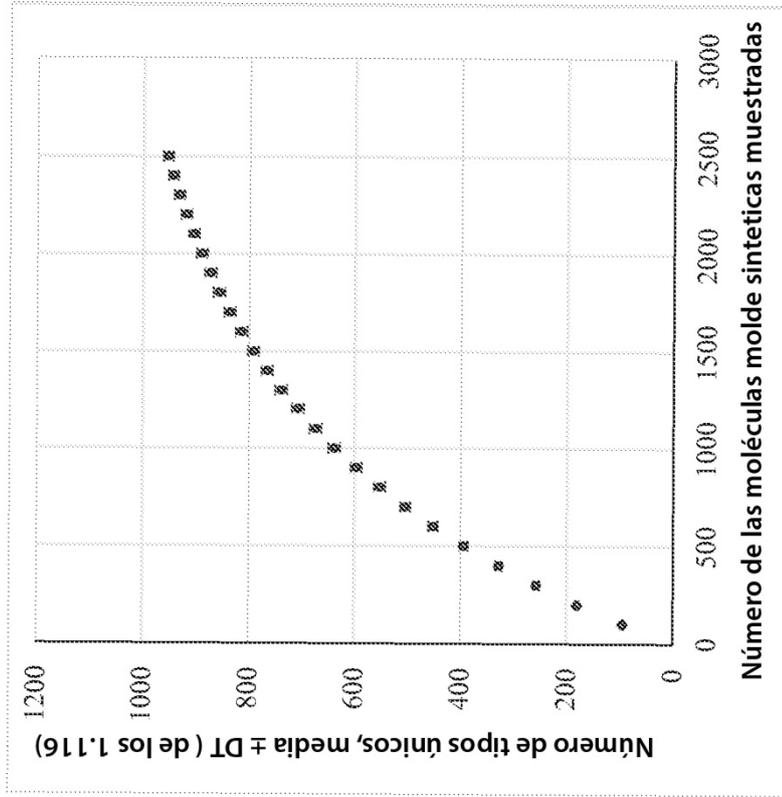


Figura 2

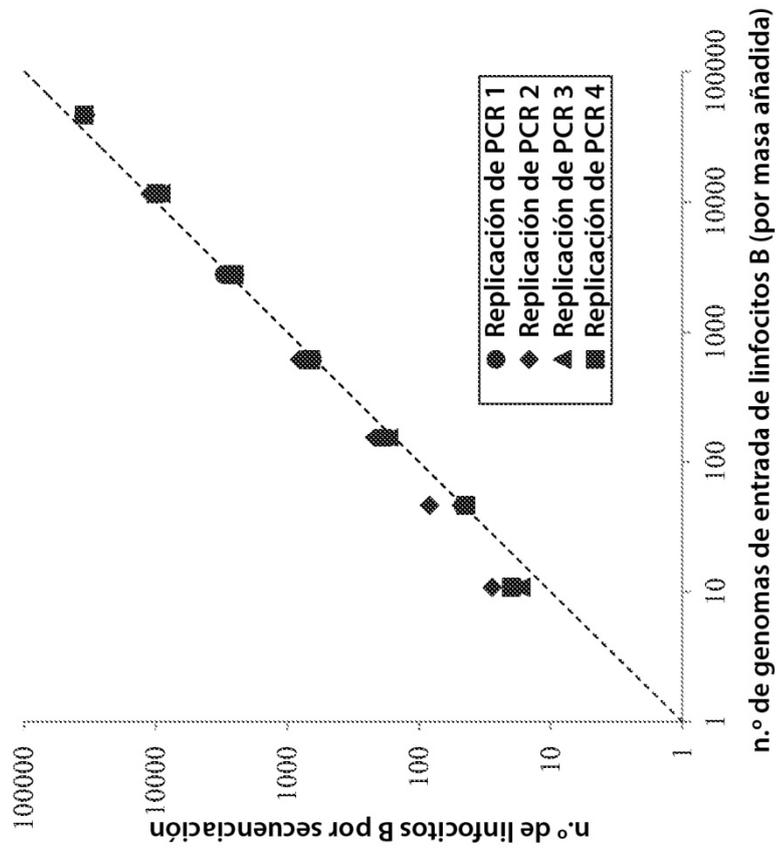


Figura 3

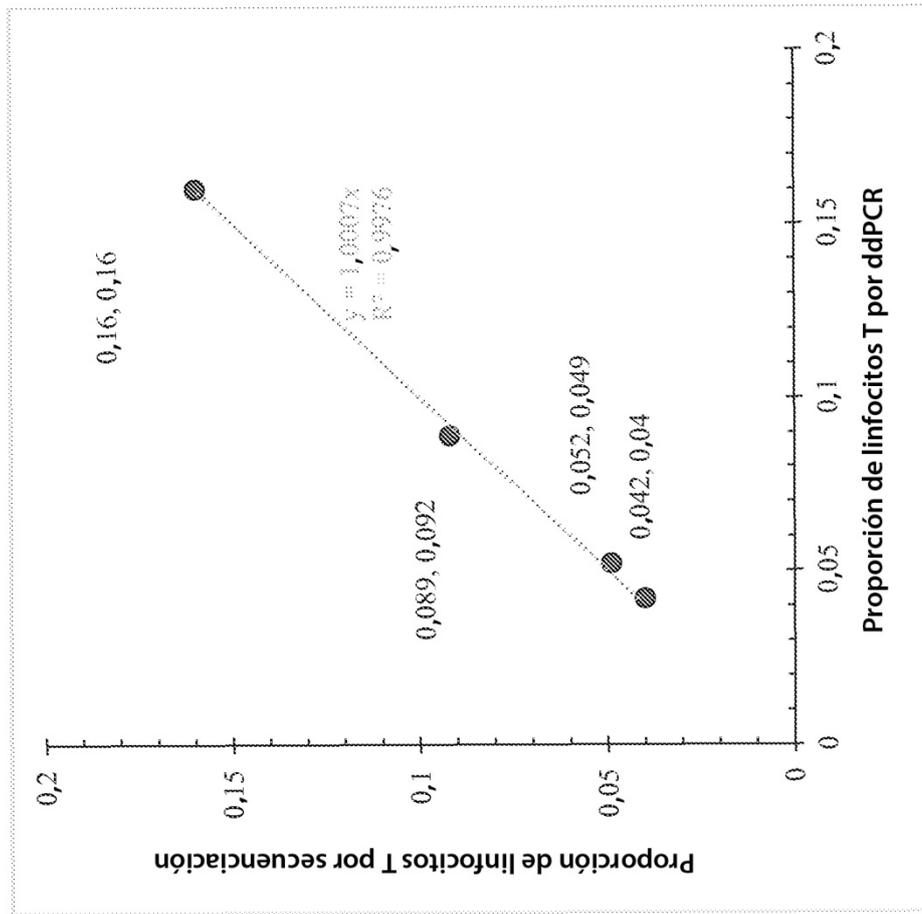


Figura 4

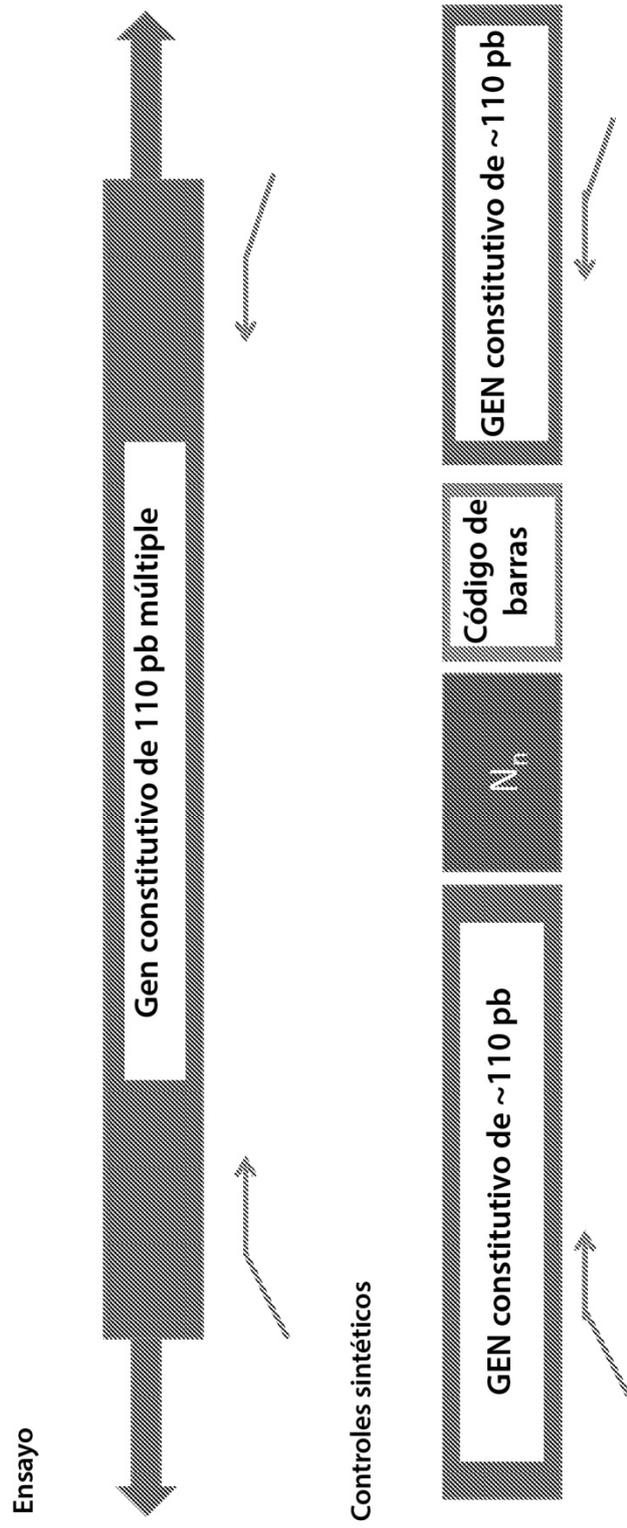
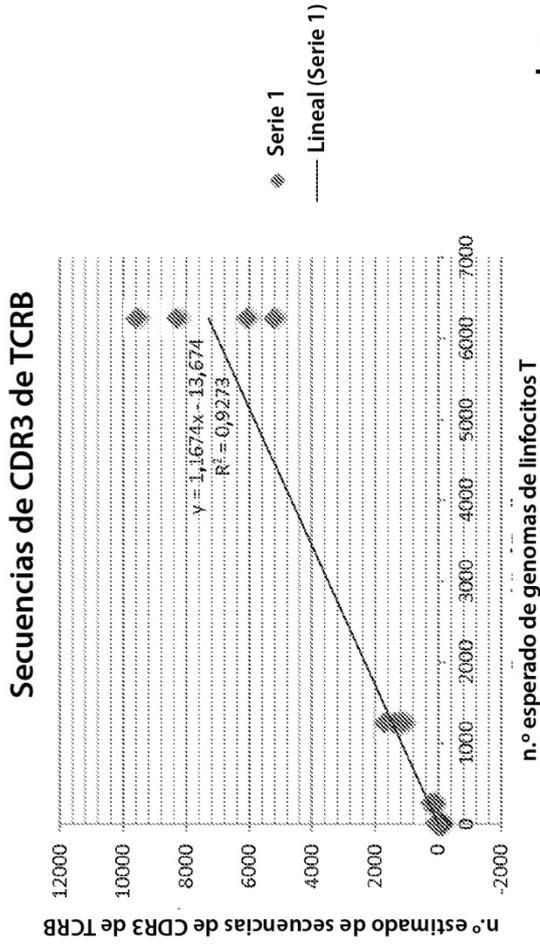
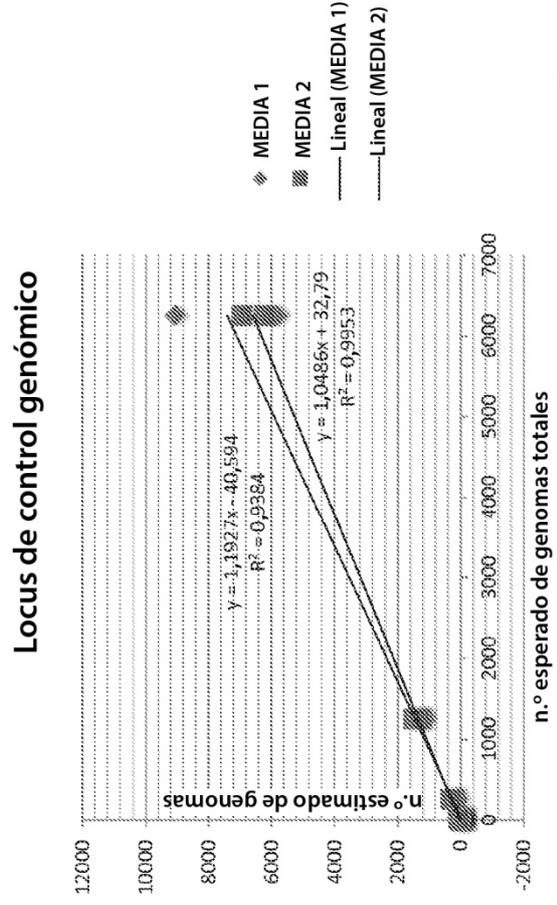


Figura 5

Figura 6



n.º esperado de genomas de linfocitos T



n.º esperado de genomas totales