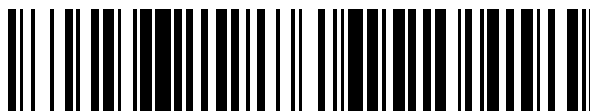


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 741 740**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12P 19/34 (2006.01)

C07H 21/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.03.2015 PCT/US2015/019029**

87 Fecha y número de publicación internacional: **11.09.2015 WO15134787**

96 Fecha de presentación y número de la solicitud europea: **05.03.2015 E 15758762 (7)**

97 Fecha y número de publicación de la concesión europea: **24.07.2019 EP 3114240**

54 Título: **Métodos que usan moléculas sintéticas que contienen segmentos de nucleótidos aleatorios**

30 Prioridad:

05.03.2014 US 201461948418 P

06.03.2014 US 201461949069 P

14.11.2014 US 201462080173 P

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

12.02.2020

73 Titular/es:

**ADAPTIVE BIOTECHNOLOGIES CORPORATION
(100.0%)**

**1551 Eastlake Avenue East, Suite 200
Seattle, Washington 98102, US**

72 Inventor/es:

SHERWOOD, ANNA M.;

EMERSON, RYAN O.;

ROBINS, HARLAN S.;

RIEDER, MARK J. y

PARSONS, JOE

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 741 740 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Métodos que usan moléculas sintéticas que contienen segmentos de nucleótidos aleatorios

5 Antecedentes

El sistema inmunitario adaptativo protege a los organismos superiores contra infecciones y otros eventos patológicos que pudieran atribuirse a sustancias extrañas, mediante el uso de los receptores de inmunidad adaptativa, las proteínas de reconocimiento con especificidad antigénica expresadas por células hematopoyéticas del linaje linfóide y que son capaces de distinguir las moléculas propias de las no propias en el huésped. Estos linfocitos pueden encontrarse en la circulación y los tejidos de un huésped, y se ha descrito su recirculación entre la sangre y la linfa, que incluye su extravasación a través de las vénulas de endotelio alto de los ganglios linfáticos, así como en sitios de infección, inflamación, lesión tisular y otros insultos clínicos. (Ver, *por ejemplo*, Stein y otros, 2005 *Immunol.* 116:1-12; DeNucci y otros, 2009 *Crit. Rev. Immunol.* 29:87-109; Marelli-Berg y otros, 2010 *Immunol.* 130:158; Ward y otros, 2009 *Biochem. J.* 418:13; Gonzalez y otros, 2011 *Ann. Rev. Immunol.* 29:215; Kehrl y otros, 2009 *Curr. Top. Microb. Immunol.* 334:107; Steinmetz y otros, 2009 *Front. Biosci. (Schol. Ed.)* 1:13.)

En consecuencia, la naturaleza dinámica del movimiento de los linfocitos en todo un organismo huésped se refleja en cambios en la distribución cualitativa (*por ejemplo*, especificidad antigénica del receptor de inmunidad adaptativa expresado clonalmente (inmunoglobulina o receptor de células T), célula T frente a célula B, célula T auxiliar (T_h) frente a célula T reguladora (T_{reg}), célula T efectora frente a célula T de memoria, etc.) y cuantitativa de los linfocitos entre los tejidos, en función de los cambios en el estado inmunitario del huésped.

Por ejemplo, numerosos estudios han descubierto una asociación entre (i) la presencia de linfocitos infiltrantes de tumor (TIL) en una variedad de tumores sólidos y (ii) el pronóstico del paciente y las tasas de supervivencia general. En algunos estudios, las células T infiltrantes de tumor que tienen un fenotipo específico (*por ejemplo*, células T CD8⁺ y CD4⁺ o células T reguladoras) son predictores positivos o negativos de la supervivencia (*por ejemplo*, Jochems y otros, 2011 *Experimental Biol. Med.* 236:567-579). En ciertos casos, sin embargo, el recuento de TIL solo, es un predictor de la supervivencia a largo plazo (*por ejemplo*, Katz y otros, 2009 *Ann. Surg. Oncol.* 16:2524-2530). Por lo tanto, la determinación cuantitativa de los recuentos de TIL tiene alto valor pronóstico en una variedad de cánceres que incluyen los cánceres colorrectales, hepatocelular, de vesícula biliar, pancreático, esofágico, ovárico endometrial, cervical, de vejiga y urotelial. Aunque se conoce más acerca de la asociación de las células T infiltrantes de tumor, se conoce que las células B también se infiltran en tumores y los estudios han demostrado una asociación de las células B infiltrantes de tumor con la ventaja de supervivencia (*por ejemplo*, Ladányi, y otros, *Cancer Immunol. Immunother.* 60(12):1729-38, julio 21, 2011 (publicado electrónicamente antes de su impresión).

La determinación cuantitativa de la presencia de células de la inmunidad adaptativa (*por ejemplo*, linfocitos T y B) en tejidos enfermos puede proporcionar por consiguiente información útil para el diagnóstico, pronóstico y otros propósitos, tales como en cáncer, infección, inflamación, lesión tisular y otras afecciones.

El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígenos de células T y B con diversidad suficiente para reconocer el universo de patógenos potenciales, los linfocitos B maduran para expresar anticuerpos (inmunoglobulinas, Ig) que aparecen como heterodímeros de un polipéptido de cadena pesada (H) y uno de cadena ligera (L), mientras que los linfocitos T expresan receptores de células T (TCR) heterodiméricos. La capacidad de las células T de reconocer el universo de antígenos asociados con varios cánceres u organismos infecciosos se confiere por su receptor de antígenos de células T (TCR), compuesto por una cadena α (alfa) y una cadena β (beta) o una cadena γ (gamma) y una δ (delta). Las proteínas que componen estas cadenas están codificadas por el ADN, que emplea un mecanismo único para la generación de la enorme diversidad de TCR. Este receptor multisubunitario de reconocimiento inmunitario se asocia con el complejo CD3 y se une a péptidos presentados por las proteínas del complejo mayor de histocompatibilidad (MHC) clase I y II en la superficie de las células presentadoras de antígeno (APC). La unión del TCR al péptido antigénico en la APC es el evento central en la activación de células T, que se produce en una sinapsis inmunológica en el punto de contacto entre la célula T y la APC.

Cada péptido de TCR contiene regiones determinantes de complementariedad (CDR) variables, así como regiones marco (FR) y una región constante. La diversidad de secuencias de células T $\alpha\beta$ está determinada en gran medida por la secuencia de aminoácidos de los lazos de la tercera región determinante de complementariedad (CDR3) de los dominios variables de las cadenas α y β , cuya diversidad es un resultado de la recombinación entre los segmentos génicos variables (V_{β}), de diversidad (D_{β}), y de unión (J_{β}) en el locus de la cadena β , y entre los segmentos génicos V_{α} y J_{α} análogos en el locus de la cadena α , respectivamente. La existencia de múltiples segmentos génicos de este tipo en los loci de las cadenas α y β del TCR permite la codificación de un gran número de secuencias de CDR3 distintas. La diversidad de secuencias de CDR3 aumenta aún más por la adición y delección independientes de nucleótidos en las uniones V_{β} - D_{β} , D_{β} - J_{β} , y V_{α} - J_{α} durante el proceso de reordenamiento de genes del TCR. En este sentido, la inmunocompetencia se refleja en la diversidad de TCR.

El TCR $\gamma\delta$ se diferencia del TCR $\alpha\beta$ en que codifica un receptor que interactúa estrechamente con el sistema inmunitario innato. El TCR $\gamma\delta$, se expresa temprano en el desarrollo, tiene una distribución anatómica especializada, tiene

especificidades únicas por patógeno y molécula pequeña, y tiene un amplio espectro de interacciones celulares innatas y adaptativas. Un patrón sesgado de expresión de segmentos V y J de TCR γ se establece temprano en la ontogenia a medida que los subconjuntos restringidos de células con TCR $\gamma\delta$ pueblan la boca, la piel, el intestino, la vagina, y los pulmones en la etapa prenatal. Por consiguiente, el repertorio diverso de TCR γ en tejidos adultos es el resultado de la extensa expansión periférica después de la estimulación por exposición ambiental a patógenos y moléculas tóxicas.

Las Ig expresadas por células B son proteínas que consisten en cuatro cadenas polipeptídicas, dos cadenas pesadas (cadenas H) y dos cadenas ligeras (cadenas L), que forman una estructura H₂L₂. Cada par de cadenas H y L contiene un dominio hipervariable, que consiste en una región V_L y una 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. Similar al TCR, el dominio V de las cadenas H se crea mediante la unión combinatoria de los segmentos génicos V_H, D_H, y J_H. La diversidad de secuencias de los dominios hipervariables aumenta aún más por la adición y delección independientes de nucleótidos en las uniones V_H-D_H, D_H-J_H, y V_H-J_H durante el proceso de reordenamiento de genes de Ig. En este sentido, la inmunocompetencia se refleja en la diversidad de las Ig.

La PCR múltiple y la secuenciación de moléculas de ADN presentan problemas importantes para el análisis de datos cuantitativos. El primer problema involucra la medición y corrección de la amplificación por PCR no uniforme, que puede atribuirse a los diferentes cebadores presentes en el esquema de amplificación múltiple. Los métodos actuales para abordar este problema incluyen la medición del sesgo de amplificación que puede atribuirse a cada cebador en la PCR múltiple mediante el uso de moléculas de prueba que se amplifican y secuencian por separado, y después usar la información resultante para corregir el resultado de la secuenciación en las reacciones posteriores.

El segundo problema involucra la cuantificación del número de moléculas de cada tipo único presente en la muestra inicial, a diferencia de las frecuencias relativas producidas por los datos primarios de secuenciación del ADN. Aunque las medidas estándar del número de ácidos nucleicos iniciales, como la absorbancia A₂₆₀, pueden proporcionar un estimado crudo del número de células totales, este valor no siempre puede ser confiable si las muestras se manipulan o tratan de manera incorrecta con agentes conservantes, tales como formalina. Tanto la formalina como el transcurso del tiempo pueden fragmentar el ADN, lo que hace difícil estimar el número de genomas amplificables en una muestra. Se necesitan métodos y sistemas para estimar el número total de genomas utilizables añadidos a una reacción de PCR. El estado de la técnica para abordar este problema involucra la comparación del número de lecturas de secuenciación observadas en un experimento con un estimado del número de moléculas de prueba iniciales incluidas en la reacción, lo que genera una cobertura media que puede usarse para estimar el número de plantillas iniciales que puede atribuirse a moléculas que no son de prueba.

La caracterización cuantitativa de las células de la inmunidad adaptativa en base a la presencia en tales células de genes que codifican Ig y TCR reordenados funcionalmente que dirigen la expresión productiva de receptores de inmunidad adaptativa se ha logrado mediante el uso de muestras biológicas a partir de las cuales pueden aislarse fácilmente células de la inmunidad adaptativa en cantidades significativas, tales como sangre, linfa u otros fluidos biológicos. En estas muestras, las células de la inmunidad adaptativa aparecen como partículas en suspensión líquida. Ver, *por ejemplo*, el documento US 2010/0330571; ver además, *por ejemplo*, Murphy, *Janeway's Immunobiology* (8^{va} Ed.), 2011 Garland Science, NY, Apéndice I, pp. 717-762. Los métodos anteriores incluyen la cuantificación de la representación relativa de células de la inmunidad adaptativa en una muestra mediante la amplificación a partir de la muestra de los polipéptidos de región V, los polipéptidos de región J, y un gen de control interno, y la comparación del número de células que contienen polipéptidos de región V y J con el número de células que contienen el gen de control interno. Ver, *por ejemplo*, el documento US2013/028237. Sin embargo, este método no permite la cuantificación absoluta de las células de la inmunidad adaptativa en la muestra. Aunque puede determinarse una representación relativa de las células de la inmunidad adaptativa, los métodos actuales no permiten la determinación del número absoluto de células de la inmunidad adaptativa en la muestra inicial.

Existe la necesidad de un método que permita la cuantificación absoluta exacta de células de la inmunidad adaptativa en una muestra biológica compleja. También existe la necesidad de un método mejorado para cuantificar una representación relativa de células de la inmunidad adaptativa en una muestra biológica compleja de este tipo. Tales necesidades incluyen métodos para identificar y mejorar la exactitud de la corrección del sesgo de amplificación por PCR múltiple y métodos para determinar la cuantificación absoluta de plantillas iniciales, mientras se alivia la necesidad de datos extrínsecos para garantizar resultados exactos y cuantitativos.

Resumen de la invención

La invención proporciona un método para corregir el sesgo de amplificación en una reacción de PCR de una muestra, el método comprende:

A) amplificar por PCR múltiple, secuenciar, y cuantificar las lecturas de salida de:

(i) moléculas plantilla biológicas que comprenden secuencias oligonucleotídicas de CDR3 reordenadas de loci de receptor de células T (TCR) de células T o loci de inmunoglobulina (Ig) de células B, cada secuencia comprende un segmento V de TCR o IG y un segmento J de TCR o IG, para obtener un número total de lecturas de secuencias biológicas

de salida; y

- (ii) moléculas plantilla sintéticas que comprenden cada una un segmento V de TCR o Ig y un segmento J de TCR o IG, secuencias adaptadoras universales de cebadores directos y/o inversos, uno o más códigos de barras que identifican las moléculas plantilla como sintéticas, una secuencia oligonucleotídica de marcador interno, y una secuencia oligonucleotídica aleatoria, en donde cada secuencia oligonucleotídica aleatoria comprende una secuencia de nucleótidos única, y en donde cada molécula plantilla sintética comprende una combinación única de un segmento V y un segmento J, para obtener un número total de lecturas de secuencias sintéticas de salida;
- B) agrupar las lecturas de secuencias al:
- (i) extraer dichas lecturas de secuencias;
- (ii) identificar si una lectura de secuencia es una lectura de secuencia biológica o una lectura de secuencia sintética al:
- (a) comparar las lecturas de secuencias contra las secuencias plantilla sintéticas conocidas mediante el uso de una primera métrica para identificar secuencias plantilla sintéticas, mientras se ignora la porción de la lectura de secuencia que se espera corresponda a la secuencia oligonucleotídica aleatoria en secuencias sintéticas;
- (b) comparar las lecturas de secuencias sin coincidencia restantes contra las secuencias plantilla sintéticas conocidas mediante el uso de una segunda métrica, mientras se ignora la porción de la lectura de secuencia que se espera corresponda a la secuencia oligonucleotídica aleatoria en secuencias sintéticas;
- (iii) agrupar las lecturas de secuencias sintéticas mediante el colapso de las lecturas de secuencias que coinciden con la misma secuencia oligonucleotídica sintética esperada y comparten la misma secuencia oligonucleotídica aleatoria;
- (iv) asignar a cada grupo de lecturas de secuencias sintéticas, en base a la secuencia oligonucleotídica sintética esperada con la que coincidieron, una secuencia consenso que comprende la secuencia esperada de la secuencia oligonucleotídica sintética con la que coincidieron, que incluye un segmento V y un segmento J; y
- (v) determinar un número total de lecturas de secuencias sintéticas observadas en cada grupo;
- C) calcular uno o más factores de normalización para los segmentos V y los segmentos J en las lecturas de secuencias sintéticas al:
- (i) calcular un recuento de lecturas promedio entre todos los grupos de secuencias que coincidieron con cada secuencia oligonucleotídica sintética;
- (ii) calcular una media general de los recuentos de lecturas promedio para cada segmento V y J único, entre las secuencias oligonucleotídicas sintéticas que contienen un V dado y cualquier J o viceversa;
- (iii) calcular un sesgo de amplificación promedio mediante la división del recuento de lecturas promedio de cada segmento V y segmento J calculado en la etapa C(ii) por la media general de los recuentos de lecturas promedio de los segmentos V o de los recuentos de lecturas promedio de los segmentos J calculada en la etapa C(ii) para llegar a un factor de amplificación para cada segmento V y J; y
- (iv) producir el factor de normalización para un segmento V o J dado mediante el cálculo del recíproco del sesgo de amplificación promedio producido en la etapa C(iii); y
- D) multiplicar el recuento de lecturas de secuencias observadas de cada secuencia biológica única dada por el factor de normalización calculado en la etapa C(iv) correspondiente al segmento V presente en esa secuencia biológica única y por el factor de normalización calculado en la etapa C(iv) correspondiente al segmento J presente en esa secuencia biológica única, para corregir de este modo el sesgo de amplificación en la reacción de PCR múltiple de la muestra.

Los métodos de la invención abordan los problemas indicados anteriormente mediante el uso de moléculas sintéticas destinadas a la inclusión directa en reacciones de amplificación y secuenciación de una muestra, y cuya cantidad en la reacción (el número exacto de moléculas) puede medirse con precisión para mejorar la exactitud de la corrección del sesgo de amplificación por PCR múltiple y la cuantificación absoluta de plantillas iniciales, mientras se alivia la necesidad de cualquier dato extrínseco para garantizar resultados exactos y cuantitativos. El sesgo de amplificación se describe en más detalle en el documento WO2013/169957, presentado el 8 de mayo de 2013.

En una modalidad, se proporciona un método para determinar y corregir el sesgo de amplificación en una reacción de PCR de una muestra. En un caso, el método proporciona la amplificación por PCR múltiple y la secuenciación de loci de receptor de células T (TCR) de células T o loci de inmunoglobulina (Ig) de células V reordenados en una muestra para obtener un número total de secuencias biológicas de salida. En un caso adicional, se proporcionan métodos para amplificar por PCR múltiple y secuenciar un conjunto de plantillas sintéticas que comprenden cada una un segmento V de TCR o Ig y un segmento J o C de TCR de Ig y un código de barras único que identifica dicha plantilla sintética como sintética. En una modalidad, cada plantilla sintética comprende una combinación única de segmentos V y J o C, secuencias adaptadoras universales de cebadores directos y/o inversos, uno o más códigos de barras que identifican las moléculas plantilla como sintéticas, una secuencia oligonucleotídica de marcador interno, y una cadena de oligonucleótidos aleatorios. En una modalidad adicional, la cadena de oligonucleótidos aleatorios comprende una secuencia de nucleótidos única. En un caso adicional, cada plantilla sintética comprende una combinación única de segmentos V y segmentos J.

En un caso adicional el método comprende agrupar e identificar las lecturas de secuenciación resultantes a través de la extracción de las lecturas y la comparación de las lecturas contra las secuencias plantilla sintéticas agrupadas para hacer coincidir las secuencias leídas con las secuencias plantilla sintéticas agrupadas. Las lecturas de secuenciación que se identifican como secuencias plantilla sintéticas se colapsan entre sí si comparten la misma secuencia oligonucleotídica aleatoria. En un caso adicional, se identifica el número de lecturas de cada plantilla sintética única y los segmentos V y J, y se calcula un recuento de lecturas promedio para cada segmento V único y cada segmento J de referencia asociado con dicho segmento V y se compila una lista de recuentos de lecturas promedio para cada segmento V particular. En un

caso adicional, se calcula una media general de los recuentos de lecturas promedio de todos los segmentos V únicos y los segmentos J de referencia y el recuento de lecturas promedio para cada combinación de segmentos V/J se divide por la media general de los recuentos de lecturas promedio para llegar a un factor de amplificación para el segmento V y el segmento J de referencia correspondiente. En un caso adicional, el factor de normalización para un segmento V dado se produce mediante el cálculo del recíproco de la media de los factores de amplificación para cada segmento V con diferentes genes J de referencia. En un caso adicional, el factor de normalización para los segmentos J se calcula para cada segmento J y el segmento V de referencia correspondiente como se describió anteriormente para los segmentos V. El factor de normalización calculado para cada segmento V y segmento J de aplica entonces al número de secuencias biológicas de salida para cada segmento V y segmento J.

En una modalidad adicional, la etapa de comparar las lecturas de secuenciación contra las secuencias plantilla sintéticas agrupadas se realiza con la métrica de Hamming. En una modalidad adicional, la etapa de comparar las lecturas de secuencias sin coincidencia restantes contra las secuencias plantilla sintéticas agrupadas se realiza con la métrica de Levenshtein.

En modalidades adicionales, la muestra puede obtenerse a partir de un sujeto mamífero. En una modalidad adicional, la muestra puede comprender una mezcla de células T y/o células B, así como células que no son células T o células B. En una modalidad adicional, la muestra puede comprender tejido somático o comprende una biopsia de tumor. En una modalidad adicional, puede ser tejido fresco, tejido congelado, o tejido fijado. En una modalidad adicional, la muestra puede comprender células de seres humanos, ratas o ratones.

En una modalidad, el método incluye plantillas sintéticas que comprenden la secuencia, 5'-U1-B1-V-I-B2-N-J-B3-U2-3'. En una modalidad, V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) de TCR o Ig o el complemento de esta. En una modalidad, cada plantilla sintética comprende una secuencia oligonucleotídica de región V única. En una modalidad, J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica que codifica la región de unión (J) de TCR o Ig o el complemento de esta. En una modalidad, U1 comprende una secuencia oligonucleotídica que es una primera secuencia adaptadora universal o una primera secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 5' a una primera secuencia oligonucleotídica adaptadora universal. En una modalidad adicional, U2 comprende una secuencia oligonucleotídica que es una secuencia adaptadora universal o una segunda secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 5' a una segunda secuencia oligonucleotídica adaptadora universal. En una modalidad, I es una secuencia oligonucleotídica de marcador interno que comprende al menos 2 y no más de 100 nucleótidos. En una modalidad, N es una secuencia oligonucleotídica aleatoria que comprende al menos 2 y no más de 100 nucleótidos. En una modalidad, B1, B2 y B3 son cada uno independientemente, nada o una secuencia oligonucleotídica de código de barras de al menos 2 y no más de 100 ácidos nucleicos que identifica de manera única como una combinación pareada una secuencia oligonucleotídica de región V única y un oligonucleótido de región J único. En una modalidad al menos uno de B1, B2 y B3 está presente en cada plantilla sintética. En una modalidad al menos dos de B1, B2 y B3 están presentes en cada plantilla sintética. En una modalidad los tres de B1, B2 y B3 están presentes en cada plantilla sintética. En una modalidad, las plantillas sintéticas comprenden una cadena de oligonucleótidos aleatorios que comprende al menos 4 y no más de 15 nucleótidos. En una modalidad, la cadena de oligonucleótidos aleatorios comprende al menos 4 y no más de 50 nucleótidos. En una modalidad, el segmento aleatorio de oligonucleótidos comprende alrededor de 8 oligonucleótidos. En una modalidad, los oligonucleótidos aleatorios comprenden alrededor de 12 oligonucleótidos.

En una modalidad, la amplificación de los loci de TCR o Ig reordenados y el primer conjunto de plantillas sintéticas se realiza mediante el uso de una pluralidad de cebadores oligonucleotídicos. En una modalidad, los cebadores oligonucleotídicos comprenden una pluralidad de cebadores oligonucleotídicos de segmentos V que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región V de TCR de Ig o al complemento de este. En una modalidad, cada cebador de segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región V de TCR o Ig. En una modalidad, la pluralidad de cebadores de segmentos V se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región V de TCR o Ig que están presentes en la composición. En una modalidad, la pluralidad de cebadores incluye además una pluralidad de cebadores oligonucleotídicos de segmentos J que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región J de TCR o Ig o al complemento de este. En una modalidad, cada cebador de segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región J de TCR o Ig. En una modalidad, la pluralidad de cebadores de segmentos J se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región J de TCR o Ig que están presentes en la composición.

En una modalidad, la pluralidad de cebadores oligonucleotídicos de segmentos V y dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprenden las secuencias que se exponen en las SEQ ID NO:1-764. En una modalidad, la pluralidad de cebadores oligonucleotídicos de segmentos V comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740, y/o la pluralidad de cebadores oligonucleotídicos de segmentos J comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID

NO:121-146, 159-166, 277-406, 579-592, y 741-764.

5 En una modalidad, la muestra es biopsia de tumor. En una modalidad 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 modalidad, 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 modalidad, el segmento V de Ig comprende un segmento génico V de IGH, un segmento génico V de IGL, o un segmento génico V de IGK. En una modalidad, el segmento de región J de Ig comprende un segmento génico J de IGH, un segmento génico J de IGL, o un segmento génico V de IGK.

10 En una modalidad, las secuencias de salida obtenidas son cada una de alrededor de 100 a 300 nucleótidos de longitud.

15 Los métodos descritos en la presente descripción involucran la generación y uso de moléculas plantilla sintéticas. En un caso, las moléculas plantilla sintéticas pueden incluir secuencias oligonucleotídicas que son complementarias a una molécula diana, una secuencia oligonucleotídica aleatoria de longitud N, y una secuencia de código de barras única. Las secuencias oligonucleotídicas aleatorias pueden generarse aleatoriamente durante la síntesis de la molécula.

20 En la presente descripción se describen métodos que se proporcionan para estimar el número de genomas iniciales en una muestra. En un caso, el método involucra amplificar por PCR múltiple y secuenciar una o más secuencias biológicas para obtener un número total de secuencias biológicas de salida y un conjunto de plantillas sintéticas que contienen una o más secuencias biológicas correspondientes a las secuencias biológicas amplificadas. En un caso, el conjunto de plantillas sintéticas incluye, además de la una o más secuencias biológicas correspondientes, un código de barras único que identifica la(s) plantilla(s) sintética(s) como sintética(s) y un segmento de ácidos nucleicos aleatorios. En una modalidad, cada miembro del conjunto de plantillas sintéticas se encuentra representado solo una vez en el conjunto amplificado. En un caso, se determina un factor de amplificación para cada una de la una o más secuencias biológicas mediante la división del número total de secuencias sintéticas amplificadas y secuenciadas por el número total inicial de plantillas sintéticas únicas amplificadas y secuenciadas. En un caso adicional, el número de genomas iniciales en la muestra se estima mediante la división del número total de secuencias biológicas de salida para cada una de la una o más secuencias biológicas amplificadas y secuenciadas por el factor de amplificación correspondiente para esa secuencia biológica.

30 En un caso, la muestra comprende células T y/o células B y proporciona un estimado del número de genomas de células T y/o células B iniciales totales. En un caso, el método incluye amplificar por PCR múltiple y secuenciar una o más secuencias oligonucleotídicas de CDR3 reordenadas de receptor de células T (loci de TCR) de células T y o loci de inmunoglobulina (Ig) de células B. En un caso, cada secuencia oligonucleotídica de CDR3 comprende un segmento V y un segmento J. En un caso, el número total de células T y/o células B se determina mediante la adición del número de genomas estimados para cada uno de los loci de TCR y/o Ig reordenados.

40 En un caso, el método incluye amplificar por PCR múltiple y secuenciar una o más regiones de control genómico. En un caso, el método incluye amplificar por PCR múltiple y secuenciar dos o más regiones de control genómico. En un caso, el método incluye amplificar por PCR múltiple y secuenciar tres o más regiones de control genómico. En un caso, el método incluye amplificar por PCR múltiple y secuenciar cuatro o más regiones de control genómico. En un caso, el método incluye amplificar por PCR múltiple y secuenciar cinco o más regiones de control genómico. En un caso, el método incluye amplificar por PCR múltiple y secuenciar uno o más de ACTB, B2M, C1orf34, CHMP2A, GPI, GUSB, HMBS, HPRT1, PSMB4, RPL13A, RPLP0, SDHA, SNRPD3, UBC, VCP, VPS29, PPIA, PSMB2, RAB7A, UBC, VCP, REEP5 y EMC7. En un caso, el método incluye amplificar por PCR múltiple y secuenciar PSMB2, RAB7A, PPIA, REEP5, y EMC7. En un caso, el número total de genomas iniciales se calcula tomando un promedio mediante el uso de cada uno de los cinco factores de amplificación determinados para cada uno de PSMB2, RAB7A, PPIA, REEP5, y EMC7 amplificados y secuenciados. En un caso adicional, el número mínimo y máximo calculado de genomas iniciales se descarta antes de tomar el promedio.

50 En un caso, el método involucra amplificar por PCR múltiple y secuenciar un conjunto de plantillas sintéticas de la fórmula I: 5'-U1-B1-V-B2-J-B3-U2-3'. En un caso, V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) de TCR o Ig, o el complemento de esta y cada plantilla en dicho primer conjunto de plantillas sintéticas tiene una secuencia oligonucleotídica de región V única. En un caso adicional, J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica que codifica la región de unión (J) de TCR o Ig, o el complemento de esta y cada plantilla en dicho primer conjunto de plantillas sintéticas comprende una secuencia oligonucleotídica de región J única. Aún en un caso adicional, U1 comprende una secuencia oligonucleotídica que se selecciona de (i) una primera secuencia oligonucleotídica adaptadora universal; y (ii) una primera secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 5' a una primera secuencia oligonucleotídica adaptadora universal. En un caso adicional, U2 comprende una secuencia oligonucleotídica que se selecciona de (i) una segunda secuencia oligonucleotídica adaptadora universal; y (ii) una segunda secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 5' a una segunda secuencia oligonucleotídica adaptadora universal. Aún en un caso adicional, B1, B2 y B3 comprenden cada uno independientemente nada o una secuencia oligonucleotídica de código de barras de 3-25 ácidos nucleicos que identifica de manera única, como una combinación pareada (i) dicha secuencia oligonucleotídica de región V única; y dicho oligonucleótido de región J único, en donde al menos uno de B1, B2 y B3 está presente en cada

plantilla sintética contenida en dicho conjunto de oligonucleótidos y en donde dichas plantillas sintéticas comprenden un segmento de nucleótidos aleatorios único. Aún en un caso adicional, cada una de las plantillas sintéticas comprende un segmento de nucleótidos aleatorios único. En un caso, el segmento de nucleótidos aleatorios comprende de 4 a 50 nucleótidos. En un caso adicional, el segmento de nucleótidos aleatorios comprende 8 nucleótidos.

5

En un caso, se proporciona un método para determinar la relación de células T o B en una muestra con relación al número total de genomas iniciales. En un caso, el método proporciona la amplificación por PCR múltiple y la secuenciación de loci de receptor de células T (TCR) de células T o loci de inmunoglobulina (Ig) de células V reordenados en una muestra para obtener un número total de secuencias biológicas de salida. En un caso adicional, se proporcionan métodos para amplificar por PCR múltiple y secuenciar un primer conjunto de plantillas sintéticas que comprenden cada una un segmento V de TCR o Ig y un segmento J o C de TCR de Ig y un código de barras único que identifica dicha plantilla sintética como sintética. En un caso, cada plantilla sintética comprende una combinación única de segmentos V y J o C. En un caso adicional el método proporciona la determinación de un factor de amplificación para cada plantilla sintética que se encuentra representada por el número total de primeras plantillas sintéticas amplificadas y secuenciadas dividido por el número inicial total de primeras plantillas sintéticas únicas. En un caso, el método proporciona la determinación del número total de células T o células B en la muestra mediante la división del número total de secuencias biológicas de salida por el factor de amplificación correspondiente a esa secuencia biológica.

10

15

20

25

30

En un caso el método proporciona además la amplificación por PCR múltiple y la secuenciación de una o más regiones de control genómico del ADN obtenido a partir de una muestra para obtener el número total de secuencias biológicas de salida para cada región de control genómico. En un caso adicional se proporcionan métodos para amplificar por PCR múltiple y secuenciar un segundo conjunto de plantillas sintéticas, que comprenden cada una la secuencia de una o más de dichas regiones de control genómico, un código de barras único y un segmento de ácidos nucleicos aleatorios. En un caso cada plantilla sintética en el segundo conjunto de plantillas sintéticas se encuentra representado solo una vez. En un caso adicional, el método proporciona la determinación de un factor de amplificación para cada una de la una o más regiones de control genómico mediante la división del número total de segundas plantillas sintéticas amplificadas y secuenciadas por el número inicial total de segundas plantillas sintéticas únicas. En un caso, el método proporciona además un método para determinar el número total de genomas iniciales mediante la división del número total de secuencias biológicas de salida de cada región de control genómico por el factor de amplificación correspondiente para esa región de control genómico.

En un caso, la muestra se obtiene a partir de un sujeto mamífero. En otro caso, la muestra comprende una mezcla de células que comprende células T y/o células B y células que no son células T y/o células B.

35

En un caso, el número total de plantillas sintéticas en el primer conjunto de plantillas sintéticas que se somete a amplificación se determina mediante el uso de una dilución limitante de dichas plantillas sintéticas que comprenden cada una una región V y J o C de TCR de Ig única de manera que cada plantilla sintética única se encuentra en copia única.

40

En un caso, el número total de plantillas sintéticas en el primer conjunto de plantillas sintéticas que se somete a amplificación se determina mediante el recuento del número de plantillas sintéticas únicas en base a los nucleótidos aleatorios únicos contenidos en cada plantilla sintética.

45

En un caso, el método proporciona la amplificación de dos o más regiones de control genómico. En otro caso, el método proporciona la amplificación de tres o más regiones de control genómico. Aún en otro caso, el método proporciona la amplificación de cuatro o más regiones de control genómico. Aún en otro caso, el método proporciona la amplificación de cinco o más regiones de control genómico. En un caso, el método proporciona la amplificación de cinco regiones de control genómico y el cálculo de los factores de amplificación para cada una. En un caso, el factor de amplificación promedio se determina tomando el promedio de los factores de amplificación para cada región de control genómico. En un caso, el factor de amplificación de región de control genómico mínimo y máximo se descarta antes de tomar un promedio. En un caso, las regiones de control genómico son una o más de PPIA, PSMB2, RAB7A, UBC, VCP, REEP5, EMC7, VPS29, SNRPD3, SDHA, RPLPO, RPL13A, PSMB4, HPRT1, HMBS, GUSB, GPI, CHMP2A, C1orf43, B2M, y ACT3. En un caso, las regiones de control genómico son PSMB2, RAB7A, PPIA, REEP5, y EMC7.

50

55

En un caso, la PCR múltiple y la secuenciación de los loci de TCR o Ig reordenados y las primeras plantillas sintéticas se realizan en una reacción de PCR múltiple mientras que la amplificación de las regiones de control genómico y el segundo conjunto de plantillas sintéticas se realizan en una segunda reacción de PCR múltiple. En otro caso, los loci de TCR y o Ig reordenados, el primer conjunto de plantillas sintéticas, las regiones de control genómico y el segundo conjunto de plantillas sintéticas se amplifican y secuencian en la misma reacción de PCR múltiple.

60

En un caso, la amplificación de los loci de TCR o Ig reordenados y el primer conjunto de plantillas sintéticas se realiza mediante el uso de una pluralidad de cebadores oligonucleotídicos. En un caso, los cebadores oligonucleotídicos comprenden una pluralidad de cebadores oligonucleotídicos de segmentos V que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región V de TCR de Ig o al complemento de este. En un caso, cada cebador de segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región V de TCR o Ig. En un caso, la pluralidad de cebadores de segmentos V se hibrida específicamente a sustancialmente

65

- 5 todos los segmentos génicos funcionales que codifican la región V de TCR o Ig que están presentes en la composición. En un caso, la pluralidad de cebadores incluye además una pluralidad de cebadores oligonucleotídicos de segmentos J que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región J de TCR o Ig o al complemento de este. En un caso, cada cebador de segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región J de TCR o Ig. En un caso, la pluralidad de cebadores de segmentos J se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región J de TCR o Ig que están presentes en la composición.
- 10 En un caso, la pluralidad de cebadores oligonucleotídicos de segmentos V y dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprenden las secuencias que se exponen en las SEQ ID NO:1-764. En un caso, la pluralidad de cebadores oligonucleotídicos de segmentos V comprende secuencias que tienen al menos 90 % identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740, y/o la pluralidad de cebadores oligonucleotídicos de segmentos J comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740.
- 20 En un caso, la muestra es tejido fresco, congelado o fijado. En un caso, la muestra comprende células humanas, células de ratón o células de rata. En un caso, la muestra comprende tejido somático.
- 25 En un caso, la muestra es biopsia de tumor. En un caso 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 un caso, 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 un caso, el segmento V de Ig comprende un segmento génico V de IGH, un segmento génico V de IGL, o un segmento génico V de IGK. En una modalidad, el segmento de región J de Ig comprende un segmento génico J de IGH, un segmento génico J de IGL, o un segmento génico V de IGK.
- 30 En un caso, las secuencias biológicas de salida de los loci de TCR o Ig y las plantillas sintéticas contenidas en el primer conjunto de plantillas sintéticas son cada una de alrededor de 100-300 nucleótidos de longitud. En otro caso, las secuencias de salida de cada región de control genómico y las plantillas sintéticas contenidas en el segundo conjunto de plantillas sintéticas son cada una de alrededor de 100-300 nucleótidos de longitud. Aún en otro caso, las secuencias biológicas de salida de los loci de TCR o Ig, las plantillas sintéticas contenidas en el primer conjunto de plantillas sintéticas, las secuencias de salida de cada región de control genómico y las plantillas sintéticas contenidas en el segundo conjunto de plantillas sintéticas son cada una de alrededor de 100-300 nucleótidos de longitud.
- 35 En un caso, se determina un factor de amplificación para (i) una pluralidad de moléculas biológicas de ácidos nucleicos reordenados que codifican un receptor de inmunidad adaptativa que comprende un receptor de células T (TCR) o inmunoglobulina (Ig) de dicha muestra biológica, cada molécula biológica de ácido nucleico reordenado comprende un segmento génico único que codifica la región variable (V) y un segmento génico único que codifica la región de unión (J), y (ii) una pluralidad de moléculas de oligonucleótidos plantilla sintéticos, que comprenden cada una una combinación pareada de un segmento génico de región V único y un segmento génico de región J único que se encuentra en una de la pluralidad de moléculas biológicas de ácidos nucleicos reordenados.
- 40 En un caso adicional, se determina un número total de moléculas biológicas de ácidos nucleicos reordenados iniciales mediante la comparación del número de secuencias de salida de moléculas biológicas de ácidos nucleicos reordenados obtenidas a partir de la secuenciación de las moléculas biológicas de ácidos nucleicos reordenados amplificadas producidas a partir de dicha PCR múltiple con dicho factor de amplificación. Aún en un caso adicional, la representación relativa de células de la inmunidad adaptativa (en una muestra biológica que comprende una mezcla de células que comprende células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa) se determina mediante la comparación de dicho número de moléculas biológicas de ácidos nucleicos reordenados iniciales con dicho número de moléculas biológicas de ácidos nucleicos reordenados iniciales.
- 45 En algunos casos, la determinación de dicho factor de amplificación comprende dividir (1) dicho número de secuencias oligonucleotídicas plantilla sintéticas de salida obtenidas a partir de la secuenciación de las moléculas de oligonucleótidos plantilla sintéticos amplificadas generadas a partir de la PCR múltiple por (2) dicho número de oligonucleótidos plantilla sintéticos iniciales añadidos a dicha PCR múltiple. En otros casos, la determinación de un número de moléculas biológicas de ácidos nucleicos reordenados iniciales comprende dividir (1) un número total de secuencias de salida de moléculas biológicas de ácidos nucleicos reordenados obtenidas a partir de la secuenciación de moléculas biológicas de ácidos nucleicos reordenados amplificadas producidas a partir de dicha PCR múltiple por (2) dicho factor de amplificación. Aún en otros casos, la comparación de dicho número de moléculas biológicas de ácidos nucleicos reordenados iniciales con dicho número de moléculas biológicas de ácidos nucleicos reordenados iniciales comprende dividir el número de moléculas biológicas de ácidos nucleicos reordenados iniciales por dicho número de moléculas biológicas de ácidos nucleicos reordenados iniciales.
- 50 En un caso, dicho número de oligonucleótidos plantilla sintéticos iniciales añadidos en dicha PCR múltiple se determina mediante la amplificación de una mezcla de oligonucleótidos plantilla sintéticos sin diluir mediante el uso de PCR simple
- 55
- 60
- 65

para obtener una pluralidad de amplicones de plantillas sintéticas, la secuenciación de dicha pluralidad de amplicones de plantillas sintéticas para determinar una frecuencia de cada oligonucleótido plantilla sintético único en la mezcla, la cuantificación de una relación en base a simulaciones *in silico* de dicha frecuencia de cada oligonucleótido plantilla sintético único en la mezcla, entre un número total de secuencias oligonucleotídicas plantilla sintéticas únicas observadas en un subconjunto de la mezcla y el número de oligonucleótidos plantilla sintéticos totales presentes en dicho subconjunto, y la determinación de un número de oligonucleótidos plantilla sintéticos totales iniciales en dicha PCR múltiple, dicha PCR múltiple incluye una dilución limitante de dicha mezcla de oligonucleótidos plantilla sintéticos, dicha determinación es en base al número de oligonucleótidos plantilla sintéticos únicos observados en el resultado de la secuenciación de dicha PCR simple y en dicha relación cuantificada. En un caso adicional, dicho número de oligonucleótidos plantilla sintéticos iniciales añadidos en dicha PCR múltiple se determina además mediante la adición de una cantidad conocida de dicha mezcla de oligonucleótidos plantilla sintéticos diluidos a dicha PCR múltiple para producir un número de oligonucleótidos plantilla sintéticos totales amplificados.

En un caso, dicha PCR múltiple se realiza mediante el uso de una pluralidad de conjuntos de cebadores oligonucleotídicos que comprenden: (a) una pluralidad de cebadores oligonucleotídicos de segmentos V que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región V del receptor de inmunidad adaptativa o al complemento de este, en donde cada cebador de segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región V del receptor de inmunidad adaptativa y en donde dicha pluralidad de cebadores de segmentos V se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región V del receptor de inmunidad adaptativa que están presentes en la composición, y (b) una pluralidad de cebadores oligonucleotídicos de segmentos J que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región J del receptor de inmunidad adaptativa o al complemento de este, en donde cada cebador de segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región J del receptor de inmunidad adaptativa y en donde dicha pluralidad de cebadores de segmentos J se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región J del receptor de inmunidad adaptativa que están presentes en la composición, de manera que dicha pluralidad de cebadores oligonucleotídicos de segmentos V y segmentos J son capaces de amplificar en dicha PCR múltiple: (i) sustancialmente todos los oligonucleótidos plantilla sintéticos para producir una pluralidad de moléculas de oligonucleótidos plantilla sintéticos amplificadas, y (ii) sustancialmente todas las moléculas biológicas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa en dicha muestra biológica para producir una pluralidad de moléculas biológicas de ácidos nucleicos reordenados amplificadas, dicha pluralidad de moléculas biológicas de ácidos nucleicos reordenados amplificadas es suficiente para cuantificar la diversidad de dichas moléculas de ácidos nucleicos reordenados de dicha muestra biológica. En un caso adicional, dicha pluralidad de cebadores oligonucleotídicos de segmentos V y dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprenden las secuencias que se exponen en las SEQ ID NO: 1-764.

En otro caso, ya sea una o ambas de: (i) dicha pluralidad de cebadores oligonucleotídicos de segmentos V comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740, y (ii) dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740. En algunos casos, dicha pluralidad de moléculas de oligonucleótidos plantilla sintéticos comprende un número de al menos a o al menos b secuencias oligonucleotídicas únicas, cualquiera que sea mayor, en donde a es el número de segmentos génicos únicos que codifican la región V del receptor de inmunidad adaptativa en el sujeto y b es el número de segmentos génicos únicos que codifican la región J del receptor de inmunidad adaptativa en el sujeto. En un caso adicional, a está en el intervalo de 1 a un número máximo de segmentos génicos V en el genoma de dicho sujeto mamífero. En un caso adicional, b está en el intervalo de 1 a un número máximo de segmentos génicos J en el genoma de dicho sujeto mamífero. En otros casos, dicha pluralidad de moléculas de oligonucleótidos plantilla sintéticos comprende al menos una secuencia oligonucleotídica plantilla sintética para cada secuencia oligonucleotídica de región V única y al menos una secuencia oligonucleotídica plantilla sintética para cada secuencia oligonucleotídica de región J única. En algunos casos, dichas células de la inmunidad adaptativa son células T o células B. En otros casos, dicha muestra biológica es tejido fresco, tejido congelado, o tejido fijado, y dicha muestra biológica comprende células humanas, células de ratón, o células de rata. En casos adicionales, dicha muestra biológica comprende tejido somático.

En un caso, dicho segmento génico que codifica la región V comprende un segmento $V\delta$ de TCR, un segmento $V\gamma$ de TCR, un segmento $V\alpha$ de TCR, o un segmento $V\beta$ de TCR. En otro caso, dicho segmento génico que codifica la región J comprende un segmento $J\delta$ de TCR, un segmento $J\gamma$ de TCR, un segmento $J\alpha$ de TCR, o un segmento $J\beta$ de TCR. En algunos casos, dicho segmento génico que codifica la región V comprende un segmento génico V de IGH, un segmento génico V de IGL, o un segmento génico V de IGK. En otros casos, dicho segmento génico que codifica la región J comprende un segmento génico J de IGH, un segmento génico J de IGL, o un segmento génico V de IGK. En algunos casos, dicha pluralidad de secuencias oligonucleotídicas plantilla sintéticas comprende secuencias seleccionadas de las SEQ ID NO:707-3003. En otros casos, V de la fórmula (1) 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 que codifica la región V del receptor de inmunidad adaptativa, o el complemento de esta. En otros casos, J de la fórmula (1) 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 que codifica la región J del receptor de inmunidad adaptativa, o el complemento de esta.

5 En algunos casos, J de la fórmula (I) comprende una secuencia que comprende una región constante de secuencia génica que codifica la región J. En otros casos, cada secuencia oligonucleotídica plantilla sintética es de menos de 1.000, 900, 800, 700, 600, 500, 400, 300 o 200 nucleótidos de longitud.

10 En la presente descripción se describen además kits que comprenden reactivos que comprenden una composición que comprende una pluralidad de oligonucleótidos plantilla sintéticos y un conjunto de cebadores oligonucleotídicos como se describió anteriormente, e instrucciones para cuantificar una representación relativa de células de la inmunidad adaptativa en una muestra biológica que comprende una mezcla de células que comprende células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa, mediante la cuantificación de: (i) un número de productos de plantilla sintética de moléculas de oligonucleótidos plantilla sintéticos amplificadas, y (ii) un número de productos biológicos reordenados de un número de secuencias de salida.

15 Debido a que el número de secuencias de ADN posibles de longitud N es 4^N , un segmento de ADN aleatorio incluso de una longitud modesta podría codificar muchas secuencias de ADN únicas posibles. Mediante la inclusión de una secuencia oligonucleotídica aleatoria dentro de las moléculas plantilla sintéticas, se concibe una molécula que actúa (en términos de apareamiento a cebadores de PCR y secuenciación) exactamente como una molécula sintética sin secuencias oligonucleotídicas aleatorias, pero que puede cuantificarse exactamente. Esto depende del hecho de que el conjunto de secuencias sintéticas aleatorias posibles es mucho mayor que el número de moléculas añadidas a una reacción de amplificación por PCR, por lo tanto, cada secuencia oligonucleotídica aleatoria única observada en el resultado de la secuenciación representa una molécula individual de material inicial. Esto permite añadir simultáneamente moléculas suficientes para obtener información estadística excelente acerca del sesgo de amplificación (que requiere la adición de muchas moléculas) y ser capaz de cuantificar exactamente el número inicial de moléculas de ADN añadidas a la reacción (por ejemplo, a diferencia del uso de una dilución limitante y estadística de Poisson, que requiere la adición de muy pocas moléculas).

30 Breve descripción de los dibujos

Una mejor comprensión de las características novedosas de la invención y las ventajas de la presente invención se obtendrá en referencia a la siguiente descripción que expone las modalidades ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos en los que:

35 La Figura 1 representa dos moléculas plantilla sintéticas de la presente descripción. Los adaptadores universales pueden usarse para caracterizar las plantillas sintéticas con el uso de cebadores con los adaptadores universales y de Illumina anexados y secuenciarse con adaptadores de Illumina (Figura 1a). El uso de cebadores de PCR múltiple VF y JR puede usarse para caracterizar las secuencias que se encuentran entre, y que incluyen, los genes V y J (Figura 1b).

40 La Figura 2 representa la amplificación por PCR de vBloques y gBloques en dos corridas separadas. Cada punto representa el sesgo de amplificación promedio observado para plantillas sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. La leyenda de cada gráfica muestra la correlación de Pearson al cuadrado (R^2) entre las mediciones del sesgo de amplificación de vBloques y gBloques. La correlación es más fuerte en la gráfica a la izquierda debido a que la Corrida 1 de PCR incluyó un número mayor de vBloques.

45 La Figura 3 representa la consistencia de las mediciones del sesgo de amplificación en diferentes genes V y J de "Referencia" (Ref 1 y Ref 2) en ambos experimentos de PCR (corridas). Al igual que antes, cada punto representa el sesgo de amplificación promedio observado para plantillas sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. Las correlaciones de Pearson al cuadrado (R^2) se calcularon entre las mediciones del sesgo de amplificación de diferentes genes V y J de referencia en una corrida de PCR dada. La correlación es más fuerte en la gráfica a la izquierda puesto que la Corrida 1 de PCR incluyó un número mayor de vBloques.

50 La Figura 4 representa que tanto los vBloques como los gBloques producen mediciones estables del sesgo de amplificación en diferentes experimentos de PCR (corridas). Cada punto representa el sesgo de amplificación promedio observado para plantillas sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. Aquí, se calcularon las correlaciones de Pearson al cuadrado (R^2) entre las mediciones del sesgo de amplificación de gBloques (izquierda) y vBloques (derecha). La correlación es más fuerte en la gráfica a la izquierda puesto que se usaron números mayores de gBloques que de vBloques en las dos corridas.

60 La Figura 5 representa la organización de los controles sintéticos para medir las secuencias iniciales relativas en una muestra biológica, en donde la secuencia de nucleótidos aleatorios y el código de barras están, unidos entre sí y flanqueados por secuencias de un gen constitutivo seleccionado.

65 La Figura 6 representa la capacidad de los métodos que utilizan regiones de control genómico de calcular con exactitud el número de genomas iniciales y el número de células T en base al número de secuencias iniciales.

Descripción detallada de la invención

Los métodos de la descripción se proporcionan para la determinación y corrección exacta del sesgo de amplificación en la amplificación múltiple de segmentos V y/o J de células de la inmunidad adaptativa. Los métodos y composiciones se proporcionan para determinar el número de genomas iniciales de células de la inmunidad adaptativa en una mezcla compleja de células. Además, la presente descripción se refiere a métodos para la determinación cuantitativa de la presencia de linfocitos en tejidos complejos, tales como tejidos sólidos. Los métodos descritos en la presente descripción incluyen una cuantificación de la representación relativa de genomas de linfocitos infiltrantes de tumor (TIL) como una proporción relativa de todos los genomas celulares representados en una muestra, tal como una muestra de tejido sólido o tumor sólido, o la cuantificación de los genomas de linfocitos que se han infiltrado en 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 representados en una muestra de ADN de tejido.

Las composiciones descritas en la presente descripción incluyen pares de cebadores que amplifican una región del genoma y una plantilla sintética que incluye sitios de apareamiento a cebador y una etiqueta de secuencia que identifica la plantilla como sintética. Los pares de cebadores amplifican la región genómica y las plantillas sintéticas con la misma eficacia, lo que da como resultado una genoteca mixta que incluye amplicones tanto de las plantillas sintéticas como biológicas. Las plantillas sintéticas se describen en más detalle en el documento WO2013/169957.

El experto en la técnica (por ejemplo, aquellos capacitados en biología molecular) entiende el diseño de pares de cebadores para amplificar regiones conservadas del genoma. Específicamente, un par de cebadores óptimo amplifica una región conservada del genoma, evitando específicamente las regiones que tienen polimorfismos de nucleótido simple comunes y variantes de número de copias. Adicionalmente, los investigadores pueden desear cebadores para amplificar una región del genoma, pero siempre y cuando los pares de cebadores amplifiquen de manera consistente el mismo número de regiones ya sean una o dos o tres, el ensayo puede funcionar. El experto en la técnica usará la experiencia y la bibliografía publicada para identificar posibles regiones diana y comprobar si los cebadores diseñados cumplen los requisitos mediante el uso de recursos usados comúnmente como el navegador de genomas de la UCSC (<http://genome.ucsc.edu/>) o Primer BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Además, los pares de cebadores deberán amplificar una región del genoma que tiene aproximadamente el mismo tamaño que la región de interés. Por ejemplo, se ha usado como diana una región de interés en las regiones CDR3 de cadenas de TRB reordenadas. Esta región de interés se adquiere solamente en los linfocitos T, no en todos los tipos celulares. La descripción acerca del diseño de cebadores de segmentos V y segmentos J para la amplificación de regiones CDR3 se encuentra en los documentos US2010/0330571 y US2012/0058902.

En la presente descripción se proporcionan además composiciones y métodos que son útiles para cuantificar y determinar de manera confiable las secuencias de poblaciones grandes y estructuralmente diversas de genes reordenados que codifican receptores de inmunidad adaptativa, tales como inmunoglobulinas (IG) y/o receptores de células T (TCR). Estos genes reordenados pueden estar presentes en una muestra biológica que contiene ADN de células linfoides de un sujeto o fuente biológica, que incluye un sujeto humano, y/o los transcritos de ARNm de estos genes reordenados pueden estar presentes en una muestra de este tipo y usarse como plantillas para la síntesis de ADNc por transcripción inversa.

Se proporcionan métodos para cuantificar una cantidad de oligonucleótidos plantilla sintéticos en una muestra para determinar un número total de genomas iniciales de células de la inmunidad adaptativa en una muestra biológica. En un caso, una muestra de oligonucleótidos plantilla sintéticos se usa para determinar una relación del número de moléculas de oligonucleótidos plantilla sintéticos iniciales en comparación con el número de oligonucleótidos plantilla sintéticos de salida (amplicones) totales. Una dilución limitante de esta muestra se usa para enriquecer una muestra biológica (al inicio de un ensayo de PCR múltiple) y se usa para determinar el número total de genomas iniciales de células de la inmunidad adaptativa en la muestra biológica. En ciertos casos, las plantillas sintéticas en la muestra comprenden un segmento de ácidos nucleicos aleatorios, por ejemplo, un segmento de nucleótidos aleatorios de 8 nucleótidos. Por consiguiente, pueden prepararse diluciones limitantes de manera que cada plantilla sintética en la muestra esté presente solo una vez y pueda identificarse mediante el segmento de nucleótidos aleatorios de 8 nucleótidos contenido en estas. La invención no se limita por el uso de un segmento de nucleótidos aleatorios de 8 nucleótidos, sin embargo. Pueden usarse segmentos de nucleótidos aleatorios de varias longitudes, por ejemplo 4-15, o más nucleótidos de acuerdo con los métodos de la presente invención.

El método incluye además determinar la representación relativa de células de la inmunidad adaptativa en una muestra que contiene una mezcla de células, donde la mezcla comprende células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa. En ciertas modalidades, puede cuantificarse una representación relativa del ADN de células de la inmunidad adaptativa (por ejemplo, linfocitos T y/o B que tienen genes reordenados de receptor de inmunidad adaptativa, que incluyen células de los linajes T y B de diferentes etapas de maduración tales como precursores, blastocitos, progenie o similares) entre el ADN total de una muestra de tipos celulares mixtos. Por ejemplo, ciertas modalidades permiten la determinación, en el ADN extraído de una muestra biológica, de la representación relativa del ADN de linfocitos infiltrantes de tumor (TIL) en el ADN de la muestra biológica, donde la muestra comprende la totalidad o una porción de un tumor que contiene células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa (que incluyen células tumorales). Ciertas otras modalidades, por ejemplo, permiten la determinación, en el ADN extraído de una muestra biológica, de la representación relativa del ADN de linfocitos infiltrantes en el ADN de la

muestra biológica, donde la muestra comprende la totalidad o una porción de un tejido somático que contiene células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa, tales como células de un tejido sólido. Métodos alternativos para cuantificar la representación relativa de células de la inmunidad adaptativa en una mezcla de células se describen en los documentos US2013/0288237 y WO2013/059725.

Las células en la mezcla de células pueden no ser todas células de la inmunidad adaptativa, y ciertas ventajas no previstas de las modalidades descritas en la presente descripción se obtienen cuando no es necesario que todas las células en la mezcla de células sean células de la inmunidad adaptativa. Como se describe en la presente descripción, se proporcionan composiciones y métodos para cuantificar la proporción de genomas celulares en una muestra que comprende moléculas de ácidos nucleicos (por ejemplo, ADN) contribuidas por células de la inmunidad adaptativa con relación al número total de genomas celulares en la muestra, a partir de una muestra de ADN que se ha extraído de una mezcla de tipos celulares, tales como un tumor sólido o un tejido sólido.

En ciertas modalidades, las moléculas de ácidos nucleicos reordenados de receptores de inmunidad adaptativa se amplifican en una sola PCR múltiple mediante el uso de conjuntos de cebadores oligonucleotídicos específicos de receptor de inmunidad adaptativa reordenado para producir secuencias de ADN específicas de células de la inmunidad adaptativa, que se usan para determinar la contribución relativa de células de la inmunidad adaptativa en comparación con el ADN total extraído de una muestra de tipos celulares mixtos. En otros casos, las moléculas de ARNm reordenadas de células de la inmunidad adaptativa se amplifican mediante el uso de rt-qPCR y conjuntos de cebadores oligonucleotídicos específicos de receptor de inmunidad adaptativa reordenado para cuantificar las señales de ADNc de receptor de inmunidad adaptativa reordenado y determinar la contribución relativa de células de la inmunidad adaptativa al número total de genomas extraídos de una muestra de tipos celulares mixtos. Los métodos de uso de qPCR para determinar la representación relativa de células de la inmunidad adaptativa en una mezcla de células se describen en los documentos US2013/0288237 y WO2013/059725.

Además, en otras modalidades, donde la muestra incluye moléculas de ARNm, los métodos de la invención incluyen el uso de un ensayo de reacción en cadena de la polimerasa cuantitativa en tiempo real (qPCR) con conjuntos de cebadores oligonucleotídicos que específicamente amplifican sustancialmente todos los genes reordenados de receptor de inmunidad adaptativa (*por ejemplo*, porciones que contienen polinucleótidos que codifican CDR3 de genes reordenados de receptor de células T y/o inmunoglobulina) que pueden estar presentes en una muestra, para generar una primera señal de ADN detectable que refleja cuantitativamente la producción de una multiplicidad de moléculas de ADN reordenadas amplificadas que codifican el receptor de inmunidad adaptativa. En ciertas modalidades, la amplificación por qPCR puede seguirse en uno o una pluralidad de puntos de tiempo durante el curso de la reacción de qPCR, *es decir*, en “tiempo real”. El seguimiento en tiempo real permite la determinación de la cantidad de ADN que se genera mediante la comparación de una señal de cuantificación del ADN que codifica receptor de inmunidad adaptativa así medida con una señal de cuantificación de plantilla sintética apropiada (o ADN plantilla de control), que puede usarse como un estándar de calibración. Los métodos para la cuantificación mediante el uso de qPCR se describen en detalle en los documentos US2013/0288237 y WO2013/059725.

En la presente descripción se describen además enfoques inesperadamente ventajosos para determinar la representación relativa de células de la inmunidad adaptativa en una muestra biológica mediante el uso de PCR múltiple para generar una población de moléculas de ADN amplificadas a partir de una muestra biológica que contiene genes reordenados que codifican receptores de inmunidad adaptativa, antes de la secuenciación cuantitativa de alta productividad de tales productos amplificados. La amplificación múltiple y la secuenciación de alta productividad de secuencias de ADN que codifican TCR y BCR (IG) reordenados se describen, por ejemplo, en Robins y *otros*, 2009 *Blood* 114, 4099; Robins y *otros*, 2010 *Sci. Translat. Med.* 2:47ra64; Robins y *otros*, 2011 *J. Immunol. Meth.* doi:10.1016/j.jim.2011.09.001; Sherwood y *otros* 2011 *Sci. Translat. Med.* 3:90ra61; publicación de Estados Unidos núm. 2012/0058902, publicación de Estados Unidos núm. 2010/0330571, documentos WO/2010/151416, WO/2011/106738, WO2012/027503, WO/2013/169957, WO/2013/188831.

En la presente descripción se describen, además, composiciones y métodos para el uso de oligonucleótidos plantilla sintéticos destinados a la inclusión directa en 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 exactitud de la corrección del sesgo de la amplificación por PCR múltiple y la cuantificación absoluta de plantillas iniciales. El sesgo de amplificación se describe en más detalle en el documento WO/2013/169957 y en Carlson, C.S. y *otros*, Using synthetic templates to design an unbiased multiplex PCR assay, *Nature Communications* 4, 2680, doi: 10.1038/ncomms3680 (2013).

En la presente descripción se describe la cuantificación del ADN de células de la inmunidad adaptativa que están presentes en tejidos sólidos, y para tumores sólidos, de manera que la presencia relativa de células de la inmunidad adaptativa puede determinarse como una proporción de todos los tipos celulares que pueden estar presentes en el tejido (*por ejemplo*, tumor). Estos casos y los relacionados son en parte un resultado de ciertas ventajas sorprendentes y hasta el momento no reconocidas, descritas en más detalle a continuación, que se derivan de la sensibilidad exquisita brindada, para la detección de células de la inmunidad adaptativa, por el diseño de PCR múltiple mediante el uso de los conjuntos de cebadores oligonucleotídicos descritos en la presente descripción. Estos conjuntos de cebadores oligonucleotídicos permiten la producción de moléculas de ADN reordenadas amplificadas y moléculas plantilla sintéticas que codifican porciones de receptores de inmunidad adaptativa. Estos casos y los relacionados muestran la selección de una pluralidad

de cebadores oligonucleotídicos que se hibridan específicamente a secuencias polinucleotídicas que codifican polipéptidos de región V y secuencias polinucleotídicas que codifican polipéptidos de región J del receptor de inmunidad adaptativa (*por ejemplo*, receptor de células T, TCR; o inmunoglobulina, Ig). En la presente descripción se describen cebadores universales que son específicos para secuencias adaptadoras universales y se unen a amplicones que comprenden secuencias adaptadoras universales. Los cebadores promueven la amplificación por PCR de moléculas de ácidos nucleicos, tales como ADN, que incluyen sustancialmente todas las regiones génicas reordenadas que codifican CDR3 de TCR o que codifican CDR3 de Ig que pueden estar presentes en una muestra biológica de prueba, donde la muestra contiene una mezcla de células que comprende células de la inmunidad adaptativa (*por ejemplo*, células de los linajes linfocitarios T y B) y células que no son células de la inmunidad adaptativa. Por ejemplo, una mezcla de células puede obtenerse a partir de un tumor sólido que comprende células tumorales y TIL.

Receptores de células de la inmunidad adaptativa

El TCR nativo es una proteína de superficie celular heterodimérica de la superfamilia de las inmunoglobulinas, que se asocia con proteínas invariantes del complejo CD3 involucradas en la mediación de la transducción de señales. El TCR existe en las formas $\alpha\beta$ y $\gamma\delta$, que son estructuralmente similares, pero tienen ubicaciones anatómicas y probablemente funciones bien diferenciadas. Los ligandos de MHC clase I y clase II, que se unen al TCR, también son proteínas de la superfamilia de las inmunoglobulinas, pero están especializadas en la presentación de antígenos, con un sitio de unión a péptido altamente polimórfico que les permite presentar una gama diversa de fragmentos peptídicos cortos en la superficie celular de las APC.

Las porciones extracelulares de los TCR $\alpha\beta$ y $\gamma\delta$ heterodiméricos nativos consisten en dos polipéptidos cada uno de los cuales tiene un dominio constante próximo a la membrana, y un dominio variable distal a la membrana. Cada uno de los dominios constante y variable incluye un puente disulfuro intracatenario. Los dominios variables contienen los lazos altamente polimórficos análogos a las regiones determinantes de complementariedad (CDR) de los anticuerpos. La CDR3 de los TCR $\alpha\beta$ interactúa 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 por medio del reordenamiento somático de genes variables (V), de diversidad (D), de unión (J), y constantes unidos.

Los loci de los genes de Ig y TCR contienen muchos segmentos génicos variables (V), de diversidad (D), y de unión (J) diferentes, que se someten a procesos de reordenamiento durante la diferenciación linfóide temprana. Las secuencias de los segmentos génicos V, D y J de Ig y TCR se conocen en la técnica y están disponibles en bases de datos públicas tales como GENBANK. Los reordenamientos Y-D-J están mediados por un complejo de enzimas recombinasas en el que las proteínas RAG1 y RAG2 desempeñan un papel importante mediante el reconocimiento y el corte del ADN en las secuencias señales de recombinación (RSS). Las RSS se ubican corriente abajo de los segmentos génicos V, a ambos lados de los segmentos génicos D, y corriente arriba de los segmentos génicos J. Las RSS inapropiadas reducen o incluso impiden completamente el reordenamiento. La RSS consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3', y nonámero, 5'-ACAAAAACC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Un número de posiciones de nucleótidos se han identificado como importantes para la recombinación, que incluyen el dinucleótido CA en la posición uno y dos del heptámero, y una C en la posición tres del heptámero también ha demostrado ser fuertemente preferida, así como un nucleótido de A en las posiciones 5, 6, 7 del nonámero. (Ramsden y otros 1994 *Nucl. Ac. Res.* 22:1785; Akamatsu y otros 1994 *J. Immunol.* 153:4520; Hesse y otros 1989 *Genes Dev.* 3:1053). Las mutaciones de otros nucleótidos tienen efectos mínimos o inconsistentes. El espaciador, aunque más variable, también tiene un impacto sobre la recombinación, y los reemplazos de nucleótidos simples han demostrado tener un impacto significativo sobre la eficiencia de recombinación (Fanning y otros 1996 *Cell. Immunol. Immunopath.* 79:1; Larijani y otros 1999 *Nucl. Ac. Res.* 27:2304; Nadel y otros 1998 *J. Immunol.* 161:6068; Nadel y otros 1998 *J. Exp. Med.* 187:1495). Se han descrito criterios para identificar secuencias polinucleotídicas de RSS que tienen eficiencias de recombinación significativamente diferentes (Ramsden y otros 1994 *Nucl. Ac. Res.* 22:1785; Akamatsu y otros 1994 *J. Immunol.* 153:4520; Hesse y otros 1989 *Genes Dev.* 3:1053, y Lee y otros, 2003 *PLoS* 1(1):E1).

El proceso de reordenamiento generalmente comienza con un reordenamiento D a J seguido de un reordenamiento V a D-J en el caso de los genes de cadena pesada de Ig (IgH), TCR beta (TCRB), y TCR delta (TCRD) o se refiere a reordenamientos directos 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 génicos en reordenamiento generalmente se eliminan en forma de un producto de escisión circular, también llamado círculo de escisión de TCR (TREC) o círculo de escisión de receptor de células B (BREC).

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

Los linfocitos B maduros extienden aún más su repertorio de Ig tras el reconocimiento del antígeno en los centros foliculares por medio de hipermutación somática, un proceso, que conduce a la maduración de la afinidad de las moléculas de Ig. El proceso de hipermutación somática se centra en el exón V- (D-) J de los genes de IgH y cadena ligera de Ig y se

refiere a mutaciones de nucleótidos simples y en ocasiones también a inserciones o deleciones de nucleótidos. Los genes de Ig que se someten a hipermutación somática también se encuentran en enfermedades malignas de células B maduras de origen folicular o posfolicular.

5 Definiciones

Como se usa en la presente descripción, el término "gen" se refiere a un segmento de ADN que puede expresarse como una cadena polipeptídica. La cadena polipeptídica puede ser la totalidad o una porción de un polipéptido de TCR o Ig (*por ejemplo*, un polipéptido que contiene CDR3). El gen puede incluir regiones precedentes y posteriores a la región codificante ("líder y seguidora"), secuencias intermedias (intrones) entre segmentos codificantes individuales (exones), elementos reguladores (*por ejemplo*, promotores, potenciadores, sitios de unión a represor y similares), y secuencias señales de recombinación (RSS), como se describe en la presente descripción.

Los "ácidos nucleicos" o "moléculas de ácidos nucleicos" o "polinucleótidos" u "oligonucleótidos" pueden estar en forma de ácidos ribonucleicos (ARN), o en forma de ácidos desoxirribonucleicos (ADN). Como se refiere en la presente descripción, el ARN incluye ARNm. El ADN incluye ADNc, ADN genómico, y ADN sintético. El ADN puede ser bicatenario o monocatenario, y si es monocatenario puede ser la cadena codificante o la cadena no codificante (antisentido). Una secuencia codificante que codifica un TCR o una inmunoglobulina o una región de estos (*por ejemplo*, una región V, un segmento D, una región J, una región C, *etc.*) puede ser idéntica a la secuencia codificante conocida en la técnica para las regiones génicas o dominios polipeptídicos (*por ejemplo*, dominios de región V, dominios CDR3, *etc.*) de cualquier TCR o inmunoglobulina dados. En otras modalidades, la secuencia codificante puede ser una secuencia codificante diferente, que, como resultado de la redundancia o reiteración del código genético, codifica la misma región o polipéptido de TCR o inmunoglobulina.

El término "cebador," como se usa en la presente descripción, se refiere a un oligonucleótido capaz de actuar como punto de iniciación de la síntesis de ADN en condiciones adecuadas. Tales condiciones incluyen aquellas en las que la síntesis de un producto de extensión de cebador complementario a una cadena de ácido nucleico se induce en la presencia de cuatro nucleósidos trifosfatados diferentes y un agente para la 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 típicamente está en el intervalo de 6 a 50 nucleótidos, o en ciertas modalidades, de 15-35 nucleótidos. Las moléculas cebadoras cortas generalmente requieren temperaturas más bajas para formar complejos híbridos lo suficientemente estables con la plantilla. No es necesario que un cebador refleje la secuencia exacta del ácido nucleico plantilla, pero debe ser lo suficientemente complementario para hibridarse con la plantilla. El diseño de cebadores adecuados para la amplificación de una secuencia diana dada se conoce bien en la técnica y se describe en la bibliografía citada en la presente descripción.

Como se describe en la presente descripción, los cebadores pueden incorporar características adicionales que permiten la detección o inmovilización del cebador, pero no alteran la propiedad básica del cebador, la de actuar como un punto de iniciación de la síntesis de ADN. Por ejemplo, los cebadores pueden contener una secuencia de ácido nucleico adicional en el extremo 5' que no se hibrida con el ácido nucleico diana, pero que facilita la clonación, la detección, o la secuenciación del producto amplificado. La región del cebador que es suficientemente complementaria para hibridarse a la plantilla se denomina en la presente descripción la región de hibridación.

Como se usa en la presente descripción, 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 al ácido nucleico diana. Típicamente, un cebador es específico para una secuencia diana si la estabilidad del dúplex de cebador con diana es mayor que la estabilidad de un dúplex formado entre el cebador y cualquier otra secuencia que se encuentra en la muestra. El experto en la técnica reconocerá que varios factores, tales como las condiciones salinas, así como la composición de bases del cebador y la ubicación de los errores de apareamiento, afectarán la especificidad del cebador, y que en muchos casos se necesitará la confirmación experimental de rutina de la especificidad del cebador. Pueden seleccionarse condiciones de hibridación en las que el cebador puede formar dúplex estables solamente con una secuencia diana. Por lo tanto, el uso de cebadores específicos para la diana en condiciones de amplificación adecuadamente rigurosas permite la amplificación selectiva de las secuencias diana que contienen los sitios de unión a cebador en la diana.

El término "aliviar" se refiere a cualquier resultado terapéuticamente beneficioso en el tratamiento de un estado patológico, por ejemplo, un estadio de cáncer, un estado patológico autoinmunitario, que incluye profilaxis, reducción de la gravedad o progresión, remisión, o cura de este.

El término "in vivo" se refiere a procesos que se producen en un organismo vivo.

El término "mamífero" como se usa en la presente descripción incluye tanto a seres humanos como no humanos e incluye pero sin limitarse a seres humanos, primates no humanos, caninos, felinos, murinos, bovinos, equinos, y porcinos.

5 El término porcentaje de "identidad", en el contexto de dos o más secuencias de ácido nucleico o polipéptido, se refiere a dos o más secuencias o subsecuencias que tienen un porcentaje especificado de nucleótidos o residuos de aminoácidos que son iguales, cuando se comparan y alinean para la correspondencia máxima, como se mide mediante el uso de uno de los algoritmos de comparación de secuencias descritos a continuación (*por ejemplo*, BLASTP y BLASTN u otros algoritmos disponibles para los expertos) o por inspección visual. En dependencia de la aplicación, el porcentaje de "identidad" puede existir en una región de la secuencia comparada, *por ejemplo*, en un dominio funcional, o, alternativamente, existe en la longitud completa de las dos secuencias a comparar.

10 Para la comparación de secuencias, típicamente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, si es necesario se designan coordenadas posteriores, y se designan los parámetros del programa del algoritmo de secuencia. Después, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la(s) secuencia(s) de prueba(s) con relación a la secuencia de referencia, basado en los parámetros de programa designados.

15 El alineamiento óptimo de las secuencias para la comparación puede realizarse, *por ejemplo*, por el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), por el algoritmo de alineamiento por homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), por el método de búsqueda de similitud de Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el paquete de software de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, Wis.), o por inspección visual (ver generalmente Ausubel y *otros*, más abajo).

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

25 Debe señalarse que, como se usa en la descripción y en las reivindicaciones adjuntas, las formas singulares "un", "una" y "el/la" incluyen referencias plurales a menos que el contexto lo indique claramente de cualquier otra manera.

Muestras (Tejidos y Uso)

35 Como se usa en la presente descripción, una muestra, muestra de prueba o muestra biológica de prueba se refiere a tejidos biológicos (por ejemplo, un agregado de células que tienen estructura y función similar) obtenidos a partir de un sujeto de interés. La muestra puede incluir una mezcla compleja de células de la inmunidad adaptativa (*por ejemplo*, células de linaje linfocitario T y B) y células que no son células de la inmunidad adaptativa (por ejemplo, células de tumores sólidos).

40 En ciertas modalidades, una muestra biológica de prueba de interés comprende tejido somático. El tejido somático puede comprender un tejido sólido. En algunas modalidades, el tejido sólido puede ser un sitio de patología de enfermedad autoinmunitaria, tal como un tejido sobre el que actúa de manera inapropiada el sistema inmunitario de un huésped para una respuesta inmunitaria "contra lo propio". En ciertas otras modalidades, el tejido somático puede comprender un tejido sólido que es un sitio de una infección, tal como una infección bacteriana, de levadura, viral o de otro microbio (*por ejemplo*, una infección por virus del herpes simple (HSV)). Aún en otras modalidades, el tejido somático se obtiene a partir de un órgano trasplantado (*por ejemplo*, un hígado, pulmón, riñón, corazón, bazo, páncreas, piel, intestino y timo trasplantado).

45 Las muestras pueden obtenerse a partir de los tejidos antes, durante, y/o después del tratamiento. Las muestras pueden usarse en el diagnóstico, pronóstico, seguimiento de la enfermedad, seguimiento de la eficacia terapéutica y otros contextos, lo que proporciona de este modo información importante, tal como la cuantificación de la representación de células de la inmunidad adaptativa en tejidos complejos que comprenden una mezcla de células. La cuantificación de células de la inmunidad adaptativa (*por ejemplo*, la cuantificación de la representación relativa de células de la inmunidad adaptativa en muestras) o la cuantificación del ADN de células de la inmunidad adaptativa (*por ejemplo*, la cuantificación de la representación relativa de ADN de células de la inmunidad adaptativa en muestras que contienen ADN de una mezcla de células) en tejidos antes y después, y/o durante el curso de tratamiento de un sujeto, puede proporcionar información de relevancia para el diagnóstico y 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 significativas) en la presencia de células de la inmunidad adaptativa en uno o más tejidos.

50 En algunas modalidades, la muestra se obtiene a partir de un tumor sólido en un sujeto. Múltiples muestras pueden obtenerse antes, durante y/o después de la administración de un régimen terapéutico al sujeto. Una muestra puede obtenerse, por ejemplo, por escisión del tejido de un sujeto antes o después del tratamiento.

65

En otras modalidades, la muestra que comprende tejido se evalúa o analiza de acuerdo con otros criterios aceptados en la técnica. Los indicadores del estado (*por ejemplo*, evidencia de presencia o ausencia de patología, o de eficacia de un tratamiento terapéutico administrado anteriormente o simultáneamente) pueden ser, por ejemplo, compuestos indicadores detectables, nanopartículas, nanoestructuras u otras composiciones que comprenden una molécula reportera que proporciona una señal detectable que indica el estado fisiológico de una célula o tejido, tal como un colorante vital (*por ejemplo*, azul Tripán), un indicador colorimétrico de pH, un compuesto fluorescente que puede exhibir fluorescencia distinta en función de cualquiera de una serie de parámetros fisiológicos celulares (*por ejemplo*, pH, Ca²⁺ intracelular u otra concentración iónica de relevancia fisiológica, potencial de membrana mitocondrial, potencial de membrana plasmática, etc., ver Haugland, *The Handbook: A Guide to Fluorescent Probes and Labeling Technologies* (10^{ma} Ed.) 2005, Invitrogen Corp., Carlsbad, CA), un sustrato enzimático, una sonda oligonucleotídica específica, un gen reportero, o similares.

Sujetos y fuente

El sujeto o fuente biológica, del cual puede obtenerse una muestra biológica de prueba, puede ser un animal humano o no humano, o un organismo transgénico o clonado o producido por ingeniería de tejidos (que incluye a través del uso de células madre). En ciertas modalidades preferidas de la invención, puede conocerse que el sujeto o fuente biológica tiene, o puede sospecharse que tiene o está en riesgo de tener, un tumor sólido u otra afección maligna, o una enfermedad autoinmunitaria, o una afección inflamatoria, y en ciertas modalidades preferidas de la invención puede conocerse que el sujeto o fuente biológica está libre de riesgo o presencia de tal enfermedad.

Ciertas modalidades preferidas contemplan un sujeto o fuente biológica que es un sujeto humano tal como un paciente con diagnóstico de tener o estar en riesgo de desarrollar o adquirir cáncer de acuerdo con los criterios de diagnóstico clínico aceptados en la técnica, tales como los del Instituto Nacional del Cáncer de Estados Unidos (Bethesda, MD, Estados Unidos) o como se describe en DeVita, Hellman, and Rosenberg's *Cancer: Principles and Practice of Oncology* (2008, Lippincott, Williams y Wilkins, Filadelfia/ Ovid, Nueva York); Pizzo y Poplack, *Principles and Practice of Pediatric Oncology* (Cuarta edición, 2001, Lippincott, Williams y Wilkins, Filadelfia/ Ovid, Nueva York); y Vogelstein y Kinzler, *The Genetic Basis of Human Cancer* (Segunda edición, 2002, McGraw Hill Professional, Nueva York); ciertas modalidades contemplan un sujeto humano conocido por estar libre de un riesgo de tener, desarrollar o adquirir cáncer según tales criterios.

Ciertas modalidades contemplan un sujeto o fuente biológica no humana, que incluye, pero sin limitarse a, un primate no humano, tal como un macaco, chimpancé, gorila, mono verde, orangután, babuino, u otro primate no humano, que incluye sujetos no humanos de este tipo que pueden conocerse en la técnica como modelos preclínicos, que incluyen modelos preclínicos para tumores sólidos y/o otros cánceres. Ciertas otras modalidades contemplan un sujeto no humano que es un mamífero, por ejemplo, un ratón, rata, conejo, cerdo, oveja, caballo, bovino, cabra, jervo, hámster, cobaya u otro mamífero. Muchos de tales mamíferos pueden ser sujetos que se conocen en la técnica como modelos preclínicos para ciertas enfermedades o trastornos, que incluyen tumores sólidos y/u otros cánceres (*por ejemplo*, Talmadge y otros, 2007 *Am. J. Pathol.* 170:793; Kerbel, 2003 *Canc, Biol Therap.* 2(4 Supl 1):S 134; Man y otros, 2007 *Canc. Met. Rev.* 26:737; Cespedes y otros, 2006 *Clin. Transl. Oncol.* 8:318). Sin embargo, no se pretende limitar así la gama de modalidades, de manera que también se contemplan otras modalidades en las que el sujeto o fuente biológica puede ser un vertebrado que no es mamífero, por ejemplo, otra especie de vertebrado superior, o un ave, anfibio o reptil, u otro sujeto o fuente biológica.

Las muestras biológicas pueden proporcionarse mediante la obtención de una muestra de sangre, espécimen de biopsia, explante de tejido, cultivo de órgano, fluido biológico o cualquier otra preparación de tejido o células de un sujeto o una fuente biológica. En ciertas modalidades preferidas, una muestra biológica de prueba puede obtenerse a partir de un tejido sólido (*por ejemplo*, un tumor sólido), por ejemplo, por resección quirúrgica, biopsia con aguja u otros medios para la obtención de una muestra biológica de prueba que contiene una mezcla de células.

Los tejidos sólidos se conocen bien en las técnicas médicas y pueden incluir cualquier compartimento anatómico definido no fluido, cohesivo y espacialmente discreto que es sustancialmente el producto de la arquitectura multicelular, intercelular, de tejidos y/u órganos, tal como un compartimento definido de manera tridimensional que puede comprender u obtener su integridad estructural a partir del tejido conectivo asociado y puede separarse de otras áreas del cuerpo por una membrana delgada (*por ejemplo*, membrana meníngea, membrana pericárdica, membrana pleural, membrana mucosal, membrana basal, epiplón, membrana encapsuladora de órganos, o similares). Los tejidos sólidos ilustrativos no limitantes pueden incluir cerebro, hígado, pulmón, riñón, próstata, ovario, bazo, ganglios linfáticos (que incluyen amígdalas), piel, tiroide, páncreas, corazón, músculo esquelético, intestino, laringe, esófago y estómago. Las ubicaciones anatómicas, las propiedades morfológicas, la caracterización histológica, y el acceso invasivo y/o no invasivo de estos y otros tejidos sólidos son bien conocidos para aquellos familiarizados con las técnicas en cuestión.

Los tumores sólidos de cualquier tipo se contemplan como adecuados para la caracterización de los TIL mediante el uso de las composiciones y métodos descritos en la presente descripción. En ciertas modalidades preferidas, el tumor sólido puede ser un tumor benigno o un tumor maligno, que puede ser además un tumor primario, un tumor invasivo o un tumor metastásico. Ciertas modalidades contemplan un tumor sólido que comprende una de 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

5 cerebral, una célula de cáncer renal, una célula de cáncer de piel (tal como carcinoma de células escamosas, carcinoma de células basales, o melanoma) y una célula de cáncer de ovario, pero no se pretende limitar así la invención y pueden usarse otros tipos de tumores sólidos y tipos de células cancerosas. Por ejemplo, el tumor puede comprender un cáncer seleccionado de adenoma, adenocarcinoma, carcinoma de células escamosas, carcinoma de células basales, melanoma (por ejemplo, melanoma maligno), carcinoma de células pequeñas, carcinoma indiferenciado de células grandes, condrosarcoma y fibrosarcoma, o similares. Como también se señala en otra parte en la presente descripción, se han establecido criterios de diagnóstico clínico aceptados en la técnica para estos y otros tipos de cáncer, tales como los promulgados por el Instituto Nacional del Cáncer de Estados Unidos (Bethesda, MD, Estados Unidos) o como se describe en *DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology* (2008, Lippincott, Williams y Wilkins, Filadelfia/ Ovid, Nueva York); Pizzo y Piplack, *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). Otros ejemplos no limitantes de clasificación y caracterización de cánceres particulares se describen, por ejemplo, en Ignatiadis y otros (2008 *Pathobiol* 75:104); Kunz (2008 *Curr. Drug Discov. Technol.* 5:9); y Auman y otros (2008 *Drug Metab. Rev.* 40:303).

15 Las células B y células T pueden obtenerse a partir de una muestra biológica, tal como a partir de una variedad de muestras de tejidos y fluidos biológicos que incluyen médula ósea, timo, glándulas linfáticas, ganglios linfáticos, tejidos y sangre periférica, pero la sangre periférica es la de acceso más fácil. Cualquier tejido periférico puede muestrearse para detectar la presencia de células B y T y por consiguiente se contempla para el uso en los métodos descritos en la presente descripción. Los tejidos y fluidos biológicos a partir de los cuales pueden obtenerse células de la inmunidad adaptativa incluyen, pero sin limitarse a piel, tejidos epiteliales, colon, bazo, una secreción mucosal, mucosa oral, mucosa intestinal, mucosa vaginal o una secreción vaginal, tejido cervical, ganglios, saliva, fluido cerebroespinal (CSF), médula ósea, sangre de cordón umbilical, suero, fluido serosal, plasma, linfa, orina, fluido ascítico, fluido pleural, fluido pericárdico, fluido peritoneal, fluido abdominal, medio de cultivo, medio de cultivo condicionado o fluido de lavado. En ciertas modalidades, las células de la inmunidad adaptativa pueden aislarse de una muestra de aféresis. Las muestras de sangre periférica pueden obtenerse por flebotomía de sujetos. Las células mononucleares de sangre periférica (PBMC) se aíslan por técnicas conocidas para los expertos en la técnica, por ejemplo, por separación en gradiente de densidad de Ficoll-Hypaque®. En ciertas modalidades, se usan las PBMC completas para el análisis.

30 En ciertas modalidades relacionadas, las muestras que comprenden predominantemente linfocitos (por ejemplo, células T y B) o que comprenden predominantemente células T o predominantemente células B, pueden prepararse para el uso como se proporciona en la presente descripción, de acuerdo con metodologías establecidas, aceptadas en la técnica.

35 En otras modalidades relacionadas, pueden aislarse subpoblaciones específicas de células T o B antes del análisis, mediante el uso de los métodos descritos en la presente descripción. Varios métodos y kits disponibles en el mercado para aislar diferentes subpoblaciones de células T y B se conocen en la técnica e incluyen, pero sin limitarse a, separación con microesferas inmunomagnéticas para la selección de subconjuntos o clasificación de células por inmunocitometría de flujo mediante el uso de anticuerpos específicos para uno o más de cualquiera de una variedad de marcadores de superficie de células T y B conocidos. Los marcadores ilustrativos incluyen, pero sin limitarse a, uno o una combinación de CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD25, CD28, CD45RO, CD45RA, CD54, CD62, CD62L, CDw137 (41BB), CD154, GITR, FoxP3, CD54, y CD28. Por ejemplo, como conoce el experto en la técnica, los marcadores de superficie celular, tales como CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD45RA, y CD45RO pueden usarse para determinar linajes y subpoblaciones de T, B, y monocitos mediante el uso de citometría de flujo. De manera similar, la dispersión frontal de la luz, la dispersión lateral, y/o los marcadores de superficie celular, tales como CD25, CD62L, CD54, CD137, y CD154, pueden usarse para determinar el estado de activación y las propiedades funcionales de las células.

45 Las combinaciones ilustrativas útiles en ciertos métodos descritos en la presente descripción pueden incluir CD8⁺CD45RO⁺ (células T citotóxicas de memoria), CD4⁺CD45RO⁺ (células T auxiliaoras de memoria), CD8⁺CD45RO⁻ (CD8⁺CD62L⁺CD45RA⁺ (células T citotóxicas tipo virgen); CD4⁺CD25⁺CD62L^{hi}GITR⁺FoxP3⁺ (células T reguladoras). Los anticuerpos ilustrativos para el uso en separaciones inmunomagnéticas de células o clasificación de células por inmunocitometría de flujo incluyen anticuerpos antihumanos con etiquetas fluorescentes, 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 puede realizarse con la combinación apropiada de anticuerpos, seguido del lavado de las células antes del análisis. Los subconjuntos de linfocitos pueden aislarse por clasificación de células activadas por fluorescencia (FACS), por ejemplo, con 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 por métodos conceptualmente similares que involucran anticuerpos específicos inmovilizados a superficies o microesferas.

60 Para la extracción de ácidos nucleicos, el ADN genómico total puede extraerse de las células mediante el uso de métodos conocidos en la técnica y/o kits disponibles en el mercado, por ejemplo, mediante el uso del kit minipreparativo de ADN en sangre QIAamp® (QIAGEN®). La masa aproximada de un solo genoma haploide es 3 picogramos (pg). En algunas modalidades, un solo genoma diploide es de aproximadamente 6,5 picogramos. En una modalidad, el número absoluto de células T puede estimarse asumiendo una célula total de material inicial por 6,5 picogramos de datos genómicos. En algunas modalidades, se usan al menos 100.000 a 200.000 células para el análisis, es decir, alrededor de 0,6 a 1,2 µg de ADN de células T o B diploides.

PCR múltiple

5 Como se describe en la presente descripción, se proporciona un método para cuantificar la representación relativa del ADN de células de la inmunidad adaptativa en el ADN de una muestra biológica de prueba de tipos celulares mixtos, y por lo tanto para estimar el número relativo de células T o B en una mezcla compleja de células. De acuerdo con ciertos casos, el método para cuantificar la representación relativa del ADN de células de la inmunidad adaptativa en una mezcla compleja de células involucra un método de PCR múltiple que usa un conjunto de cebadores directos que se hibridan específicamente a los segmentos V y un conjunto de cebadores inversos que se hibridan específicamente a los segmentos J, donde la reacción de PCR múltiple permite la amplificación de todas las combinaciones VJ (y VDJ) posibles dentro de una población dada de células T o B. En algunas modalidades, el método de PCR múltiple incluye el uso del conjunto de cebadores directos de segmentos V y el conjunto de cebadores inversos de segmentos J para amplificar una población dada de oligonucleótidos plantilla sintéticos que comprenden las combinaciones VJ y VDJ. Debido a que la reacción de PCR múltiple amplifica sustancialmente todas las combinaciones posibles de segmentos V y J, es posible determinar, mediante el uso de PCR múltiple, el número relativo de genomas de células T o de células B en una muestra que comprende una población mixta de células.

Extracción de ácidos nucleicos

20 En una modalidad, el ADN genómico total puede extraerse de las células mediante el uso de métodos estándar conocidos en la técnica y/o kits disponibles en el mercado, *por ejemplo*, mediante el uso del kit minipreparativo de ADN en sangre QIAamp® (QIAGEN®). La masa aproximada de un solo genoma haploide es 3 pg. Preferentemente, se usan al menos 100.000 a 200.000 células para el análisis de la diversidad, es *decir*, alrededor de 0,6 a 1,2 µg de ADN de células T o B diploides.

25 Alternativamente el ácido nucleico total, que incluye tanto ADN genómico como ARNm, puede aislarse de las células. Si la diversidad se medirá a partir del ARNm en el extracto de ácidos nucleicos, el ARNm debe convertirse a ADNc antes de la medición. Esto puede hacerse fácilmente mediante métodos de un experto, por ejemplo, mediante el uso de transcriptasa inversa de acuerdo con procedimientos conocidos.

30 En algunas modalidades, el ADN o el ARNm puede extraerse de una muestra que comprende una población mixta de células. En ciertas modalidades, la muestra puede ser una muestra de tejido neoplásico o tejido somático. Las muestras ilustrativas para el uso en los presentes métodos incluyen cualquier tipo de tumor sólido, en particular, un tumor sólido de cánceres colorrectal, hepatocelular, de vesícula biliar, pancreático, esofágico, de pulmón, de mama, de próstata, de cabeza y cuello, carcinoma de células renales, de ovario, endometrial, cervical, de vejiga y urotelial. Cualquier tumor sólido en el que se evaluarán los linfocitos infiltrantes de tumor se contempla para el uso en los presentes métodos. Los tejidos somáticos que son la diana de una reacción autoinmunitaria incluyen, pero sin limitarse a, tejidos de las articulaciones, 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 adrenal, tendón, mucus, membrana, ganglios linfáticos, tiroide, endometrio, tejido conectivo, y médula ósea. En ciertas modalidades, el ADN o ARN puede extraerse de un órgano trasplantado, tal como un hígado, pulmón, riñón, corazón, bazo, páncreas, piel, intestino, y timo trasplantado.

45 En otras modalidades, dos o más muestras pueden obtenerse a partir de un solo tejido (*por ejemplo*, un solo tejido neoplásico) y las representaciones relativas de células de la inmunidad adaptativa en las dos o más muestras se cuantifican para considerar las variaciones en diferentes secciones de un tejido de prueba. En ciertas otras modalidades, la determinación de la representación relativa de células de la inmunidad adaptativa en una muestra de un tejido de prueba es suficiente debido a variaciones mínimas entre diferentes secciones del tejido de prueba.

Composiciones (cebadores para PCR múltiple)

50 Se proporcionan composiciones para el uso en una PCR múltiple que comprenden una pluralidad de cebadores de segmentos V y una pluralidad de cebadores de segmentos J que son capaces de promover la amplificación de sustancialmente todas las regiones reordenadas de manera productiva que codifican CDR3 del receptor de inmunidad adaptativa en una muestra para producir una multiplicidad de moléculas de ADN reordenadas amplificadas a partir de una población de células T (para TCR) o células B (para Ig) en la muestra.

60 Los genes de TCR e Ig pueden generar millones de proteínas distintas por medio de mutación somática. Debido a este mecanismo de generación de diversidad, las regiones determinantes de complementariedad hipervariables de estos genes pueden codificar secuencias que pueden interactuar con millones de ligandos, y estas regiones se unen a una región constante que puede transmitir una señal a la célula que indica la unión del ligando afín a la proteína. El sistema inmunitario adaptativo emplea varias estrategias para generar un repertorio de receptores de antígeno de células T y B con diversidad suficiente para reconocer el universo de patógenos potenciales. En células T $\alpha\beta$ y $\gamma\delta$, que reconocen principalmente antígenos peptídicos presentados por moléculas del MHC, la mayor parte de esta diversidad de receptores está contenida dentro de la tercera región determinante de complementariedad (CDR3) de las cadenas α y β (o las cadenas γ y δ) del receptor de células T (TCR).

5 En el genoma humano, actualmente se considera que existen alrededor de 70 segmentos génicos V α y alrededor de 61 J α de TCR, alrededor de 52 segmentos génicos V β , alrededor de 2 D β y alrededor de 13 J β de TCR, alrededor de 9 segmentos génicos V γ y alrededor de 5 J γ de TCR, y alrededor de 46 segmentos génicos V μ , alrededor de 23 D μ y alrededor de 6 J μ de cadena pesada de inmunoglobulina (IGH). En consecuencia, cuando se conocen las secuencias genómicas de estos loci de manera que pueden producirse fácilmente sondas moleculares específicas para cada una de ellas, se considera, de acuerdo con la teoría no limitante, que las presentes composiciones y métodos se refieren sustancialmente a la totalidad (*por ejemplo*, más de 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 %) de estos segmentos génicos que codifican regiones V, D y J de receptor de inmunidad adaptativa conocidos y fácilmente detectables.

10 En un caso, las composiciones descritas en la presente descripción proporcionan una pluralidad de cebadores de segmentos V y una pluralidad de cebadores de segmentos J que son capaces de amplificar sustancialmente todas las combinaciones de los segmentos V y J de un locus de receptor inmunitario reordenado. El término “sustancialmente todas las combinaciones” se refiere a al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más de la totalidad de las combinaciones de los segmentos V y J de un locus de receptor inmunitario reordenado. En ciertas modalidades, la pluralidad de cebadores de segmentos V y la pluralidad de cebadores de segmentos J amplifican la totalidad de las combinaciones de los segmentos V y J de un locus de receptor inmunitario reordenado. En ciertas modalidades, cada uno de la pluralidad de cebadores de segmentos V y segmentos J puede comprender o consistir en una secuencia de ácido nucleico que es igual a, complementaria a, o sustancialmente complementaria a una secuencia contigua de un segmento que codifica la región V o J diana (*es decir*, la porción de polinucleótido genómico que codifica un polipéptido, o una porción de ARNm de región V o de región J).

25 En algunas modalidades, los cebadores de segmentos V y segmentos J son “completamente complementarios” a una secuencia contigua de un segmento que codifica la región V o J diana, respectivamente. En otras modalidades, los cebadores de segmentos V y segmentos J son “sustancialmente complementarios” con respecto a la secuencia contigua de un segmento que codifica la región V o J diana. Generalmente no hay más de 4, 3 o 2 pares de bases con error de apareamiento tras la hibridación, mientras se retiene la capacidad de hibridación en las condiciones más relevantes para su aplicación final.

30 En ciertas modalidades, se diseñan dos mezclas de cebadores para el uso en una reacción de PCR altamente múltiple. La primera mezcla “directa” puede incluir cebadores oligonucleotídicos que son cada uno específicos para (*por ejemplo*, que tiene una secuencia de nucleótidos complementaria a una región de secuencia única de) cada segmento que codifica la región V (“segmento V”) en el locus génico de TCR o Ig respectivo. En ciertas modalidades, se usan cebadores dirigidos a una región altamente conservada, para capturar simultáneamente muchos segmentos V, lo que reduce de este modo el número de cebadores requeridos en la PCR múltiple. De esta manera, un cebador de segmento V puede ser complementario a (*por ejemplo*, se hibrida a) más de un segmento funcional que codifica la región V de TCR o Ig y actúa como un cebador promiscuo. En otras modalidades, cada cebador de segmento V es específico para un segmento funcional que codifica la región V de TCR o Ig, diferente.

40 Los cebadores de la mezcla “inversa” pueden incluir cebadores oligonucleotídicos que son cada uno específicos para (*por ejemplo*, que tienen una secuencia de nucleótidos complementaria a una región de secuencia única de) cada segmento que codifica la región J (“segmento J”) en el locus génico de TCR o Ig respectivo. En algunas modalidades, el cebador de J puede aparearse con una secuencia conservada en el segmento de unión (“J”). En ciertas modalidades, un cebador de segmento J puede ser complementario a (*por ejemplo*, se hibrida a) más de un segmento J. En otras modalidades, cada cebador de segmento J es específico para un segmento funcional que codifica la región J de TCR o Ig, diferente. A modo de ilustración y no de limitación, los cebadores de segmentos V pueden usarse como cebadores “directos” y los cebadores de segmentos J pueden usarse como cebadores “inversos”, de acuerdo con la terminología de PCR usada comúnmente, pero el experto apreciará que en ciertas otras modalidades los cebadores de segmentos J pueden considerarse cebadores “directos” cuando se usan con cebadores “inversos” de segmentos V.

50 En algunas modalidades, el cebador de segmento V o segmento J es de al menos 15 nucleótidos de longitud. En otras modalidades, el cebador de segmento V o 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 complementario a, una secuencia contigua del segmento que codifica la región V o J diana. En algunas modalidades, la longitud de los cebadores puede ser más larga, tal como alrededor de 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, en dependencia del uso o necesidad específica. Todas las longitudes intermedias de los cebadores descritos en la presente descripción se contemplan para el uso en la presente descripción. Como reconocerá el experto, los cebadores pueden comprender secuencias adicionales (*por ejemplo*, nucleótidos que pueden no ser iguales o complementarios al segmento de polinucleótido que codifica la región V o J diana), tales como sitios de reconocimiento de enzimas de restricción, secuencias adaptadoras universales para la secuenciación, secuencias de código de barras, modificaciones químicas, y similares (*ver por ejemplo*, las secuencias de cebadores proporcionadas en el listado de secuencias en la presente descripción).

65 En otras modalidades, los cebadores de segmentos V o segmentos J comprenden secuencias que comparten un alto grado de identidad de secuencia con los cebadores oligonucleotídicos para los que se presentan las secuencias de nucleótidos en la presente descripción, que incluyen los que se exponen en el listado de secuencias. En ciertas

modalidades, los cebadores de segmentos V o segmentos J comprenden variantes de cebadores que pueden tener identidad sustancial con las secuencias de cebadores de segmentos V o segmentos J del receptor de inmunidad adaptativa descritas en la presente descripción. Por ejemplo, tales variantes de cebadores oligonucleotídicos pueden comprender al menos 70 % de identidad de secuencia, preferentemente al menos 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o 99 % o mayor identidad de secuencia en comparación con una secuencia oligonucleotídica de referencia, tal como las secuencias de cebadores oligonucleotídicos descritas en la presente descripción, mediante el uso de los métodos descritos en la presente descripción (*por ejemplo*, análisis BLAST mediante el uso de parámetros estándar). El experto en esta técnica reconocerá que estos valores pueden ajustarse de manera apropiada para determinar la capacidad correspondiente de una variante de cebador oligonucleotídico de aparearse a un polinucleótido que codifica un segmento de receptor de inmunidad adaptativa teniendo en cuenta la redundancia de codones, la ubicación del marco de lectura y similares. Típicamente, las variantes de cebadores oligonucleotídicos contendrán una o más sustituciones, adiciones, deleciones y/o inserciones, preferentemente de manera que la capacidad de apareamiento de la variante de oligonucleótido no disminuya sustancialmente con relación a la de una secuencia de cebador de segmento V o segmento J de receptor de inmunidad adaptativa que se expone específicamente en la presente descripción. En otras modalidades, los cebadores de segmentos V o segmentos J se diseñan de modo que sean capaces de amplificar una secuencia de TCR o IGH reordenada que incluye la región codificante de CDR3.

En algunas modalidades, como se describe en la presente descripción, cada uno de la pluralidad de cebadores de segmentos V y segmentos J comprende secuencias adicionales en el extremo 5', tales como secuencias adaptadoras universales, secuencias de código de barras, secuencias oligonucleotídicas aleatorias, y similares. Las secuencias pueden ser secuencias de origen no natural y/o secuencias que no aparecen naturalmente adyacentes o contiguas a un segmento que codifica la región V o J diana.

En ciertas modalidades, la pluralidad de cebadores de segmentos V y segmentos J se diseñan para producir moléculas de ADN reordenadas amplificadas que tienen menos de 600 nucleótidos de longitud, para excluir de este modo los productos de amplificación de loci de receptor de inmunidad adaptativa no reordenados. En algunas modalidades, 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, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, o 600 nucleótidos de longitud. En una modalidad, la molécula de ADN reordenada amplificada es de al menos 250 nucleótidos de longitud. En otra modalidad, la molécula de ADN reordenada amplificada es de aproximadamente 200 nucleótidos de longitud. La molécula de ADN reordenada amplificada puede denominarse amplicón, molécula amplificada, producto de PCR, o producto de amplificación, por ejemplo.

Un ensayo de PCR múltiple ilustrativo usa una pluralidad de cebadores directos de segmentos V y una pluralidad de cebadores inversos de segmentos J para amplificar selectivamente los reordenamientos VDJ de cada célula. Aunque estos cebadores pueden aparearse con segmentos génicos V y J tanto de la línea germinal como reordenados, la amplificación por PCR se limita a segmentos génicos reordenados, debido al sesgo de tamaño (*por ejemplo*, producto de PCR de 250 pb mediante el uso de segmentos génicos reordenados como plantillas vs. producto de PCR de >10 Kb mediante el uso de segmentos génicos de la línea germinal como plantillas).

En algunas modalidades, la selección de los cebadores y el diseño del conjunto de cebadores puede realizarse de manera que detecte preferentemente segmentos génicos V y J productivos, y excluya pseudogenes de TCR o IG. Los pseudogenes pueden incluir segmentos V que contienen un codón de terminación en marco dentro de la secuencia codificante de segmento V, un desplazamiento del marco de lectura entre el codón de inicio y la secuencia que codifica CDR3, una o más inserciones de elementos repetidos, y deleciones 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 de segmentos V, de los cuales 26 son huérfanos 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 abiertos de lectura) que carecen de al menos un residuo de aminoácido altamente conservado, y 81 son pseudogenes.

Para detectar reordenamientos de TCR o IG funcionales en una muestra mientras se evitan señales de amplificación potencialmente superfluas que pueden atribuirse a segmentos génicos V y/o J no productivos tales como pseudogenes y/o huérfanos, por consiguiente se contempla de acuerdo con ciertas modalidades el uso de un subconjunto de cebadores oligonucleotídicos que se diseñan de modo que incluyan solamente los segmentos V que participan en un reordenamiento funcional para codificar un TCR o IG, sin tener que incluir cebadores de amplificación específicos para las secuencias de pseudogenes y/o huérfanos o similares. Por lo tanto, se obtienen eficiencias ventajosas con respecto, *inter alia*, a tiempo y gastos.

La pluralidad de cebadores de segmentos V y cebadores de segmentos J se diseñan de modo que se ubiquen fuera de las regiones donde se producen deleciones sin plantilla. Estas posiciones de cebadores de segmentos V y cebadores de segmentos J son con relación a la secuencia señal de recombinación del gen V (V-RSS) y la secuencia señal de recombinación del gen J (J-RSS) en el segmento génico. En algunas modalidades, los cebadores de segmentos V y los cebadores de segmentos J se diseñan para proporcionar información de secuencia adecuada en el producto amplificado para identificar de manera única los genes V y J.

En algunas modalidades, cada uno de los cebadores de segmentos V comprende una primera secuencia y una segunda secuencia, en donde la primera secuencia se ubica 3' a la segunda secuencia en el cebador de segmento V. En ciertas modalidades, la primera secuencia es complementaria a una porción de una primera región de al menos un segmento V, y la primera región del segmento V se ubica inmediatamente 5' a una segunda región del segmento V donde se producen delecciones sin plantilla durante el reordenamiento de genes de TCR o IG. La segunda región del segmento V es adyacente a y 5' a una secuencia señal de recombinación V (V-RSS) del segmento V. La segunda región donde se producen delecciones sin plantilla en el segmento V puede ser de al menos 10 pares de bases (pb) de longitud. En una modalidad, el extremo 3' del cebador de segmento V puede ubicarse al menos 10 pb corriente arriba de la V-RSS. En algunas modalidades, el cebador de segmento V se ubica a más de 40 pares de bases de secuencia corriente arriba de la V-RSS.

En otras modalidades, cada uno de los cebadores de segmentos J tiene una primera secuencia y una segunda secuencia, en donde la primera secuencia se ubica 3' a la segunda secuencia en el cebador de segmento J. La primera secuencia del cebador de segmento J es complementaria a una porción de una primera región de un segmento J, y la primera región del segmento J se ubica inmediatamente 3' a una segunda región del segmento J donde se producen delecciones sin plantilla durante el reordenamiento de genes de TCR o IG. La segunda región del segmento J es adyacente a y 3' a una secuencia señal de recombinación J (J-RSS) de dicho segmento J, y la segunda región del segmento J puede ser de al menos 10 pares de bases de longitud. En algunas modalidades, el extremo 3' de los cebadores de segmentos J se ubica al menos 10 pares de bases corriente abajo de la J-RSS. En ciertas modalidades, al igual que en los segmentos génicos J β de TCR, la primera región del segmento J incluye una etiqueta única de cuatro bases en las posiciones +11 a +14 corriente abajo del sitio RSS. En otras modalidades, las delecciones de segmentos J son de 4 pb +/- 2,5 pb de longitud, y los cebadores de segmentos J se ubican al menos 4 pb corriente abajo de la J-RSS. En algunas modalidades, el cebador de segmento J se ubica a más de 30 pares de bases corriente abajo de la J-RSS.

La descripción adicional acerca del diseño, colocación y ubicación de los cebadores de segmentos V y los cebadores de segmentos J, y cebadores ilustrativos puede encontrarse en los documentos US2010/0330571, W02010/151416, presentado el 4 de junio de 2010, y US2012/0058902, y Robins y otros, 2009 *Blood* 114, 4099.

Amplificación por PCR múltiple

Un sistema de PCR múltiple puede usarse para amplificar loci reordenados de receptores de células de la inmunidad adaptativa a partir de ADN genómico y de oligonucleótidos plantilla sintéticos, preferentemente a partir de una región CDR3. En ciertas modalidades, 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 a partir de un locus de Ig, tal como un locus de IgH o IgL (lambda o kappa).

En general, un sistema de PCR múltiple comprende una pluralidad de cebadores directos de segmentos V y una pluralidad de cebadores inversos de segmentos J. La pluralidad de cebadores directos de segmentos 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 modalidades, al menos 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, o 39, y en otras modalidades 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 a o es complementario a una secuencia correspondiente a uno o más segmentos de región V.

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

Los oligonucleótidos que son capaces de hibridarse o aparearse específicamente a una secuencia de ácido nucleico diana por complementariedad de bases de nucleótidos pueden hacerlo en condiciones de rigurosidad moderadas a altas. Con propósitos de ilustración, las condiciones de rigurosidad moderadas a altas adecuadas para la amplificación específica por PCR de una secuencia de ácido nucleico diana estarían entre 25 y 80 ciclos de PCR, con cada ciclo que consiste en una etapa de desnaturalización (*por ejemplo*, alrededor de 10-30 segundos al menos alrededor de 95 °C), una etapa de apareamiento (*por ejemplo*, alrededor de 10-30 s a alrededor de 60-68 °C), y una etapa de extensión (*por ejemplo*, alrededor de 10-60 s a alrededor de 60-72 °C), opcionalmente de acuerdo con ciertas modalidades con las etapas de apareamiento y extensión combinadas para proporcionar una PCR en dos etapas. Como reconocerá el experto, otros reactivos de PCR pueden añadirse o cambiarse en la reacción de PCR para aumentar la especificidad de apareamiento de los cebadores y la amplificación, tal como la alteración de la concentración de magnesio, opcionalmente la adición de DMSO, y/o el uso de cebadores bloqueados, nucleótidos modificados, ácidos nucleicos peptídicos, y similares.

En ciertas modalidades, pueden usarse técnicas de hibridación de ácidos nucleicos para evaluar la especificidad de hibridación de los cebadores descritos en la presente descripción. Las técnicas de hibridación se conocen bien en la

técnica de biología molecular. Con propósitos de ilustración, las condiciones de rigurosidad moderada adecuadas para probar la hibridación de un polinucleótido como se proporciona en la presente descripción con otros polinucleótidos incluyen el lavado previo en una solución de SSC 5X, 0,5 % SDS, 1,0 mM EDTA (pH 8,0); hibridación a 50 °C-60 °C, SSC 5X, durante toda la noche; seguido de lavado dos veces a 65 °C durante 20 minutos con cada uno de SSC 2X, 0,5X y 0,2X que contiene 0,1 % SDS. El experto en la técnica entenderá que la rigurosidad de la hibridación puede manipularse fácilmente, tal como mediante la alteración del contenido de sales de la solución de hibridación y/o la temperatura a la que se realiza la hibridación. Por ejemplo, en otra modalidad, las condiciones de hibridación altamente rigurosas incluyen las descritas anteriormente, con la excepción de que la temperatura de hibridación se aumenta, *por ejemplo*, a 60 °C-65 °C o 65 °C-70 °C.

En ciertas modalidades, los cebadores se diseñan de modo que no crucen una frontera intrón/exón. Los cebadores directos en ciertas modalidades se aparean con los segmentos V en una región de conservación de secuencia relativamente fuerte entre segmentos V para maximizar la conservación de la secuencia entre estos cebadores. En consecuencia, esto minimiza el potencial de propiedades de apareamiento diferencial de cada cebador, y de modo que la región amplificada entre los cebadores de segmentos V y J contiene información de secuencia V de TCR o Ig suficiente para identificar el segmento génico V específico usado. En una modalidad, los cebadores de segmentos J se hibridan con un elemento conservado del segmento J, y tienen una fuerza de apareamiento similar. En una modalidad particular, los cebadores de segmentos J se aparean con el mismo motivo de región marco conservado.

Los oligonucleótidos (por ejemplo, cebadores) pueden prepararse por cualquier método adecuado, que incluye síntesis química directa mediante un método tal como el método de fosfotriéster de Narang y *otros*, 1979, Meth. Enzymol. 68:90-99; el método del fosfodiéster de Brown y *otros*, 1979, Meth. Enzymol. 68:109-151; el método de dietilfosforamida de Beaucage y *otros*, 1981, Tetrahedron Lett. 22:1859-1862; y el método de soporte sólido de la patente de Estados Unidos núm. 4,458,066, cada uno incorporado en la presente descripción como referencia. Una reseña de los métodos de síntesis de conjugados de oligonucleótidos y nucleótidos modificados se proporciona en Goodchild, 1990, Bioconjugate Chemistry 1(3): 165-187.

Secuenciación de alta productividad

Oligonucleótidos de secuenciación

En una modalidad, los cebadores de segmentos V y los cebadores de segmentos J descritos en la presente descripción incluyen una segunda subsecuencia situada en sus extremos 5' que incluye una secuencia adaptadora universal complementaria a y que puede hibridarse a secuencias adaptadoras de secuenciación para el uso en un secuenciador de ADN, tal como Illumina.

En ciertas modalidades, cada uno de los segmentos génicos que codifican la región J tiene una etiqueta de identificación definida por secuencia única de 2, 3, 4, 5, 6, 7, 8, 9, 10 o alrededor de 15, 20 o más nucleótidos, situada en una posición definida con relación a un sitio RSS. Por ejemplo, puede usarse una etiqueta de cuatro bases, en el segmento que codifica la región J β de regiones que codifican CDR3 de TCR β amplificadas, en las posiciones +11 a +14 corriente abajo del sitio RSS. Sin embargo, no es necesario limitar así estas modalidades y las relacionadas y también se contemplan otras etiquetas de identificación definidas por secuencias de nucleótidos relativamente cortas que pueden detectarse en segmentos génicos que codifican la región J y definidas en base a sus posiciones con relación a un sitio RSS. Estas pueden variar entre diferentes loci que codifican receptores de inmunidad adaptativa.

La secuencia señal de recombinación (RSS) consiste en dos secuencias conservadas (heptámero, 5'-CACAGTG-3', y nonámero, 5'-ACAAAAAC-3'), separadas por un espaciador de 12 +/- 1 pb ("señal 12") o 23 +/- 1 pb ("señal 23"). Una serie de posiciones de nucleótido se han identificado como importantes para la recombinación que incluyen el dinucleótido CA en la posición uno y dos del heptámero, y una C en la posición tres del heptámero también ha demostrado ser fuertemente preferida, así como un nucleótido de A en las posiciones 5, 6, 7 del nonámero. (Ramsden y otros 1994; Akamatsu y otros 1994; Hesse y otros 1989). Las mutaciones de otros nucleótidos tienen efectos mínimos o inconsistentes. El espaciador, aunque más variable, también tiene un impacto sobre la recombinación, y reemplazos de nucleótidos simples han demostrado tener un impacto significativo sobre la eficiencia de recombinación (Fanning y otros 1996, Larjani y otros 1999; Nadel y otros 1998). Se han descrito criterios para identificar secuencias polinucleotídicas RSS que tienen eficiencias de recombinación significativamente diferentes (Ramsden y otros 1994; Akamatsu y otros 1994; Hesse y otros 1989 y Cowell y otros 1994). En consecuencia, los oligonucleótidos de secuenciación pueden hibridarse adyacentes a una etiqueta de cuatro bases dentro de los segmentos génicos que codifican J amplificados en las posiciones +11 a +14 corriente abajo del sitio RSS. Por ejemplo, los oligonucleótidos de secuenciación para TCR β pueden diseñarse de modo que se apareen con un motivo de nucleótidos consenso observado justo corriente abajo de esta "etiqueta", de modo que las primeras cuatro bases de una lectura de secuencia identificarán de manera única el segmento génico que codifica J. Las secuencias oligonucleotídicas de secuenciación ilustrativas se encuentran a continuación en la Tabla 1 y en las SEQ ID NO:765-786.

La información usada para asignar identidades a los segmentos que codifican J y V de una lectura de secuencia se encuentra completamente contenida dentro de la secuencia amplificada, y no se basa en la identidad de los cebadores de PCR. En particular, los métodos descritos en la presente descripción permiten la amplificación de todas las

combinaciones V-J posibles en un locus de TCR o Ig y la secuenciación de las moléculas amplificadas individuales permite la identificación y cuantificación del ADN reordenado que codifica las regiones CDR3. La diversidad de las células de la inmunidad adaptativa de una muestra dada puede inferirse a partir de las secuencias generadas mediante el uso de los métodos y algoritmos descritos en la presente descripción.

5

Métodos de secuenciación de alta productividad

Los métodos de la invención comprenden además la secuenciación de las moléculas de ADN amplificadas que codifican el receptor de inmunidad adaptativa que se producen. La secuenciación puede realizarse a los productos de amplificación producidos a partir de una muestra biológica que comprende células de la inmunidad adaptativa, y/o de los oligonucleótidos plantilla sintéticos que se describen a continuación.

10

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

15

La secuenciación puede realizarse mediante el uso de cualquiera de una variedad de máquinas y sistemas de secuenciación de moléculas individuales de alta productividad disponibles. Los sistemas de secuenciación ilustrativos incluyen sistemas de secuenciación por síntesis tales como el instrumento Genome Analyzer de Illumina, el MiSeq de Illumina, e instrumentos asociados (Illumina, Inc., San Diego, CA), el sistema de análisis genético de Helicos (Helicos BioSciences Corp., Cambridge, MA), PacBio de Pacific Biosciences RS (Pacific Biosciences, Menlo Park, CA), u otros sistemas que tienen capacidades similares. La secuenciación se logra mediante el uso de un conjunto de oligonucleótidos de secuenciación que se hibridan a una región definida dentro de las moléculas de ADN amplificadas. Los oligonucleótidos de secuenciación se diseñan de manera que los segmentos génicos que codifican V y J puedan identificarse de manera única mediante las secuencias que se generan, en base a la presente descripción y en vista de las secuencias génicas de receptor de inmunidad adaptativa que aparecen en bases de datos disponibles al público.

20

25

En ciertas modalidades, se secuencian al menos 30, 40, 50, 60, 70, 80, 90, 100, 101-150, 151-200, 201-300, 301-500, y no más de 1.000 nucleótidos contiguos de las moléculas de ADN amplificadas que codifican el receptor de inmunidad adaptativa. En algunas modalidades, los amplicones y oligonucleótidos plantilla sintéticos que se secuencian son de menos de 600 pb de longitud. En modalidades adicionales, las lecturas de secuenciación resultantes son de aproximadamente 130 pb de longitud. Aún en modalidades adicionales, se producen aproximadamente 30 millones de lecturas de secuenciación por ensayo de secuenciación.

30

Las composiciones y métodos para la secuenciación de secuencias génicas reordenadas de receptor de inmunidad adaptativa y para la determinación del clonotipo del receptor de inmunidad adaptativa se describen en más detalle en Robins y otros, 2009 *Blood* 114, 4099; Robins y otros, 2010 *Sci. Translat. Med.* 2:47ra64; Robins y otros, 2011 *J. Immunol. Meth.* doi: 10.1016/j.jim.2011.09.001; Sherwood y otros 2011 *Sci. Translat. Med.* 3:90ra61; publicación de Estados Unidos núm. 2012/0058902, publicación de Estados Unidos núm. 2010/0330571, documentos WO/2010/151416, WO/2011/106738, WO2012/027503.

35

40

En ciertas modalidades, cada uno de los segmentos génicos que codifican la región J amplificados puede tener una etiqueta de identificación definida por secuencia única de 2, 3, 4, 5, 6, 7, 8, 9, 10 o alrededor de 15, 20 o más nucleótidos, situada en una posición definida con relación a un sitio RSS. Por ejemplo, puede usarse una etiqueta de cuatro bases, en el segmento que codifica la región J β de regiones que codifican CDR3 de TCR β amplificadas, en las posiciones +11 a +14 corriente abajo del sitio RSS. Sin embargo, no es necesario limitar así estas modalidades y las relacionadas y también se contemplan otras etiquetas de identificación definidas por secuencias de nucleótidos relativamente cortas que pueden detectarse en segmentos génicos que codifican la región J y definidas en base a sus posiciones con relación a un sitio RSS. Estas pueden variar entre diferentes loci que codifican receptores de inmunidad adaptativa.

45

50

La secuencia 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"). Una serie de posiciones de nucleótido se han identificado como importantes para la recombinación que incluyen el dinucleótido CA en la posición uno y dos del heptámero, y una C en la posición tres del heptámero también ha demostrado ser fuertemente preferida, así como un nucleótido de A en las posiciones 5, 6, 7 del nonámero. (Ramsden y otros 1994; Akamatsu y otros 1994; Hesse y otros 1989). Las mutaciones de otros nucleótidos tienen efectos mínimos o inconsistentes. El espaciador, aunque más variable, también tiene un impacto sobre la recombinación, y reemplazos de nucleótidos simples han demostrado tener un impacto significativo sobre la eficiencia de recombinación (Fanning y otros 1996, Larijani y otros 1999; Nadel y otros 1998). Se han descrito criterios para identificar secuencias polinucleotídicas RSS que tienen eficiencias de recombinación significativamente diferentes (Ramsden y otros 1994; Akamatsu y otros 1994; Hesse y otros 1989 y Cowell y otros 1994). En consecuencia, los oligonucleótidos de secuenciación pueden hibridarse adyacentes a una etiqueta de cuatro bases dentro de los segmentos génicos que codifican J amplificados en las posiciones +11 a +14 corriente abajo del sitio RSS. Por ejemplo, los oligonucleótidos de secuenciación para TCRB pueden diseñarse de modo que se apareen con un motivo de nucleótidos consenso observado justo corriente abajo de esta "etiqueta", de modo que las primeras cuatro bases de una lectura de secuencia identificarán de manera única el segmento génico que codifica J. Los cebadores ilustrativos de J de TCRB se encuentran en las SEQ ID NO: 121-146 (que

55

60

65

muestran los cebadores inversos de segmentos J de TCRB (específicos de gen) y los cebadores inversos de segmentos J de TCRB con una secuencia adaptadora universal.

La información usada para asignar identidades a los segmentos que codifican J y V de una lectura de secuencia se encuentra completamente contenida dentro de la secuencia amplificada, y no se basa en la identidad de los cebadores de PCR. En particular, los métodos descritos en la presente descripción permiten la amplificación de todas las combinaciones V-J posibles en un locus de TCR o Ig y la secuenciación de las moléculas amplificadas individuales permite la identificación y cuantificación del ADN reordenado que codifica las regiones CDR3. La diversidad de las células de la inmunidad adaptativa de una muestra dada puede inferirse a partir de las secuencias generadas mediante el uso de los métodos y algoritmos descritos en la presente descripción. Una ventaja sorprendente proporcionada en ciertas modalidades preferidas por las composiciones y métodos de la presente descripción fue la capacidad de amplificar con éxito todas las combinaciones V-J posibles de un locus de receptor de células de la inmunidad adaptativa en una sola reacción de PCR múltiple.

En ciertas modalidades, los oligonucleótidos de secuenciación descritos en la presente descripción pueden seleccionarse de manera que el cebado promiscuo de una reacción de secuenciación para un segmento génico que codifica J por un oligonucleótido específico para otro segmento génico que codifica J distinto genere datos de secuencia que comienzan exactamente en el mismo nucleótido que los datos de secuencia del oligonucleótido de secuenciación correcto. De esta manera, el apareamiento promiscuo de los oligonucleótidos de secuenciación no afecta la calidad de los datos de secuencia generados.

La longitud promedio de la región que codifica CDR3, para el TCR, definida como los nucleótidos que codifican el polipéptido TCR entre la segunda cisteína conservada del segmento V y la fenilalanina conservada del segmento J, es 35 ± 3 nucleótidos. En consecuencia y en ciertas modalidades, la amplificación por PCR mediante el uso de cebadores de segmentos V y cebadores de segmentos J que comienza desde la etiqueta del segmento J de una región J de TCR o IgH particular (*por ejemplo*, J β de TCR, J γ de TCR o JH de IgH como se describe en la presente descripción) casi siempre capturará la unión V-D-J completa en una lectura de 50 pares de bases. La longitud promedio de la región CDR3 de IgH, definida como los nucleótidos entre la cisteína conservada en el segmento V y la fenilalanina conservada en el segmento J, está menos restringida que en el locus de TCR β , pero estará típicamente entre alrededor de 10 y alrededor de 70 nucleótidos. En consecuencia y en ciertas modalidades, la amplificación por PCR mediante el uso de cebadores de segmentos V y cebadores de segmentos J que comienza a partir de la etiqueta 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 aparean con y soportan la extensión del polinucleótido en secuencias plantilla con errores de apareamiento se denominan cebadores promiscuos. En ciertas modalidades, los cebadores de PCR inversos de segmentos J de TCR e Ig pueden diseñarse de modo que se minimice la superposición con los oligonucleótidos de secuenciación, para minimizar el cebado promiscuo en el contexto de la PCR múltiple. En una modalidad, los cebadores inversos de segmentos J de TCR e Ig pueden anclarse en el extremo 3' mediante el apareamiento al motivo de sitio de corte y empalme consenso, con superposición mínima de los cebadores de secuenciación. Generalmente, los cebadores de segmentos V y J de TCR e Ig pueden seleccionarse de modo que funcionen en la PCR a temperaturas de apareamiento consistentes mediante el uso de programas de diseño y análisis de secuencias/cebadores con parámetros predeterminados. Para la reacción de secuenciación, los cebadores de J de IGH ilustrativos usados para la secuenciación se encuentran en las SEQ ID NO:579-592 (que muestran los cebadores inversos de segmentos J de IGH (específicos de gen) y los cebadores inversos de segmentos J de IGH con una secuencia adaptadora universal).

Procesamiento de los datos de secuencia

Como se describe en la presente descripción, también se proporcionan métodos para analizar las secuencias de la mezcla diversa de regiones reordenadas que codifican CDR3 que se generan mediante el uso de las composiciones y métodos que se describen en la presente descripción. En particular, se proporciona un algoritmo para corregir el sesgo de PCR, los errores de secuenciación y de PCR y para estimar la distribución correcta de clonotipos específicos (*por ejemplo*, un TCR o Ig que tiene una secuencia CDR3 reordenada de manera única) en una muestra. Un algoritmo preferido se describe en más detalle en la presente descripción. Como reconocerá el experto, los algoritmos proporcionados en la presente descripción pueden modificarse de manera apropiada para adaptarse a situaciones experimentales o clínicas particulares.

El uso de una etapa de PCR para amplificar las regiones CDR3 de TCR o Ig antes de la secuenciación podría introducir potencialmente un sesgo sistemático en la abundancia relativa inferida de las secuencias, debido a las diferencias en la eficiencia de amplificación por PCR de regiones CDR3 que utilizan segmentos génicos V y J diferentes. Como se analiza en más detalle en los Ejemplos, cada ciclo de amplificación por PCR introduce potencialmente un sesgo de magnitud promedio $1,5^{1/15} = 1,027$. Por lo tanto, los 25 ciclos de PCR introducen un sesgo total de magnitud promedio $1,027^{25} = 1,95$ en la abundancia relativa inferida de secuencias de región CDR3 distintas.

Las lecturas secuenciadas se filtran para las que incluyen secuencias de CDR3. El procesamiento de los datos del secuenciador involucra una serie de etapas para eliminar errores en la secuencia primaria de cada lectura, y para comprimir los datos. Un filtro de complejidad elimina aproximadamente 20 % de las secuencias que son errores de lectura del secuenciador. Después, se requiere que las secuencias tengan un mínimo de coincidencia de seis bases tanto con

una de las regiones J como con una de las regiones V de TCR o Ig. Al aplicar el filtro al carril de control que contiene secuencia de fago, como promedio solo una secuencia en 7-8 millones pasó estas etapas. Por último, se usa un algoritmo del vecino más cercano para colapsar los datos en secuencias únicas mediante la fusión de las secuencias estrechamente relacionadas, para eliminar tanto el error de PCR como el error de secuenciación.

5 En el análisis de los datos, la relación de secuencias en el producto de PCR se obtiene trabajando a la inversa a partir de los datos de secuencia antes de estimar la distribución correcta de los clonotipos (*por ejemplo*, secuencias clonales únicas) en la sangre. Para cada secuencia observada un número de veces dado en los datos en la presente descripción, se estima la probabilidad de que esa secuencia se haya muestreado a partir de una mezcla de PCR de tamaño particular. Debido a que las regiones CDR3 secuenciadas se muestrean aleatoriamente a partir de una mezcla masiva de productos de PCR, el número de observaciones de cada secuencia se obtiene a partir de distribuciones de Poisson. Los parámetros de Poisson se cuantifican de acuerdo con el número de genomas de células T que proporcionó la plantilla de PCR. Un modelo de mezcla de Poisson simple estima estos parámetros y estima una probabilidad pareada para cada secuencia que se obtiene a partir de cada distribución. Este es un método de maximización de expectativas que reconstruye las abundancias de cada secuencia que se obtuvo a partir de la sangre.

Para estimar el número total de secuencias únicas de CDR3 de receptor de inmunidad adaptativa que están presentes en una muestra, puede emplearse un enfoque de cálculo que emplea la fórmula de "especies ocultas" (Efron y Thisted, 1976 *Biometrika* 63, 435-447). Este enfoque estima el número de especies únicas (*por ejemplo*, secuencias únicas de receptor de inmunidad adaptativa) en una población grande, compleja (*por ejemplo*, una población de células de la inmunidad adaptativa tales como células T o células B), en base al número de especies únicas observadas en una muestra aleatoria, finita de una población (Fisher y otros, 1943 *J. Anim. Ecol.* 12:42-58; Ionita-Laza y otros, 2009 *Proc. Nat. Acad. Sci. USA* 106:5008). El método emplea una expresión que predice el número de especies "nuevas" que podrían observarse si se analizara una segunda muestra aleatoria, finita y de tamaño idéntico de la misma población. Especies "ocultas" se refiere a al número de secuencias de receptor de inmunidad adaptativa nuevas que podrían detectarse si las etapas de amplificación de las secuencias que codifican receptor de inmunidad adaptativa en una muestra y determinación de la frecuencia de aparición de cada secuencia única en la muestra se repitieran un número infinito de veces. A modo de teoría no limitante, se asume operativamente para los propósitos de estos estimados que las células de la inmunidad adaptativa (*por ejemplo*, células T, células B) circulan libremente en el compartimento anatómico del sujeto que es la fuente de la muestra a partir de la cual se estima la diversidad (*por ejemplo*, sangre, linfa, etc.).

Para aplicar esta fórmula, los clonotipos únicos de receptores de inmunidad adaptativa (*por ejemplo*, TCR β , TCR α , TCR γ , TCR δ , IgH) toman el lugar de las especies. La solución matemática proporciona eso para S, el número total de receptores de inmunidad adaptativa que tienen secuencias únicas (*por ejemplo*, las "especies" o clonotipos TCR β , TCR γ , IgH, que en ciertas modalidades pueden ser secuencias de CDR3 únicas), un experimento de secuenciación observa x_s copias de secuencia s. Para la totalidad de 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 parámetro λ_s . El número de genomas de células T o B secuenciados en la primera medición se define como l, y el número de genomas de células T o B secuenciados en la segunda medición se define como t.

Debido a que existe un gran número de secuencia únicas, se usa una integral en lugar de una suma. Si $G(\lambda)$ es la función de distribución empírica de los parámetros λ_i , λ_s , y n_x es el número de clonotipos (*por ejemplo*, secuencias de TCR o Ig únicas, o secuencias de CDR3 únicas) observados exactamente x veces, entonces el número total de clonotipos, es decir, la medición de la diversidad E, se proporciona mediante la siguiente fórmula (I):

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

En consecuencia, la fórmula (I) puede usarse para estimar la diversidad total de especies en la fuente completa de la cual se toman muestras de tamaño idéntico. Sin desear limitarse por la teoría, el principio es que el número de clonotipos muestreados en una muestra de cualquier tamaño dado contiene información suficiente para estimar la distribución subyacente de clonotipos en la fuente completa. El valor de $\Delta(t)$, el número de clonotipos nuevos observados en una segunda medición, puede determinarse, preferentemente mediante el uso de la siguiente ecuación (II):

$$\Delta(t) = \sum_x E(n_x)_{msmt1+msmt2} - \sum_x E(n_x)_{msmt1} = S \int_0^{\infty} e^{-\lambda} (1 - e^{-\lambda t}) dG(\lambda) \quad (II)$$

en la que $msmt1$ y $msmt2$ 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 por $\Delta(t)$ produce:

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

5 que puede aproximarse mediante el reemplazo de las expectativas ($E(n_x)$) con los números reales de secuencias observadas exactamente x veces en la primera medición de la muestra. La expresión para $\Delta(t)$ oscila ampliamente dado que t tiende al infinito, así que $\Delta(t)$ se regulariza para producir un límite inferior de $\Delta(\infty)$, por ejemplo, mediante el uso de la transformación de Euler (Efron y otros, 1976 *Biometrika* 63:435).

10 De acuerdo con ciertas modalidades descritas expresamente en la presente descripción, también se proporcionan métodos en los que el grado de clonalidad de las células de la inmunidad adaptativa que están presentes en una muestra, tal como una muestra que comprende una mezcla de células donde solo algunas de ellas son células de la inmunidad adaptativa, puede determinarse ventajosamente sin la necesidad de clasificación de las células o de secuenciación del ADN. Estas modalidades y las relacionadas superan los problemas de eficiencia, tiempo y costo que, antes de la presente descripción, han obstaculizado la capacidad de determinar si la presencia de células de la inmunidad adaptativa en una muestra (*por ejemplo*, TIL) es monoclonal u oligoclonal (*por ejemplo*, si todos los TIL son la progenie de una o de un número relativamente limitado de células de la inmunidad adaptativa), o si en lugar de esto la presencia de células de la inmunidad adaptativa en la muestra es policlonal (*por ejemplo*, los TIL son la progenie de un número relativamente grande de células de la inmunidad adaptativa).

20 De acuerdo con la teoría no limitante, estas modalidades explotan la comprensión actual en la técnica (también descrita anteriormente) de que una vez que una célula de la inmunidad adaptativa (*por ejemplo*, un linfocito T o B) ha reordenado sus genes que codifican el receptor de inmunidad adaptativa (*por ejemplo*, TCR o Ig), sus células progenie poseen el mismo reordenamiento de genes que codifica el receptor de inmunidad adaptativa, lo que origina por lo tanto una población clonal que puede identificarse por la presencia en ella de segmentos génicos V y J reordenados que codifican CDR3 que pueden amplificarse mediante una combinación pareada específica de cebadores oligonucleotídicos específicos para V y J como se describe en la presente descripción.

30 Composiciones de oligonucleótidos plantilla sintéticos para el uso en la cuantificación de genomas iniciales de células de la inmunidad adaptativa, y determinación de la representación relativa de células de la inmunidad adaptativa

Composiciones de plantillas sintéticas útiles para cuantificar los números de moléculas iniciales en una muestra

35 Pueden diseñarse oligonucleótidos plantilla sintéticos para cuantificar un número de moléculas iniciales en una muestra biológica. Como se usa en la presente descripción "plantilla sintética" significa un oligonucleótido que contiene secuencias que incluyen secuencias sustancialmente idénticas a secuencias biológicas (es decir segmentos V, J o C de TCR o regiones de control genómico) además de secuencias de origen no natural (es decir códigos de barras, segmentos de nucleótidos aleatorios, adaptadores, etc.). La secuencia de nucleótidos completa de las plantillas sintéticas, por consiguiente, no aparece en la naturaleza y, en lugar de esto, son secuencias diseñadas y producidas en el laboratorio. Se determina una relación del número de moléculas de oligonucleótidos plantilla sintéticos iniciales en una muestra en comparación con el número de lecturas de secuenciación de salida totales de oligonucleótidos plantilla sintéticos (secuenciados a partir de amplicones de plantillas sintéticas) en la muestra. En una modalidad, una dilución limitante de oligonucleótidos plantilla sintéticos (que permite la determinación del número de moléculas de oligonucleótidos plantilla sintéticos totales presentes mediante la medición del número de secuencias oligonucleotídicas plantilla sintéticas únicas observadas) se añade a una muestra biológica para la PCR múltiple, y asumiendo que para las plantillas biológicas se cumple la misma relación que para las sintéticas, la relación se usa para determinar el número de moléculas reordenadas de receptor de células T o B, y por lo tanto el número de células T o B, en la muestra biológica.

50 En la presente descripción se describe una composición de plantillas sintéticas que comprende una pluralidad de oligonucleótidos plantilla de la fórmula general (I) o (II):

5'-U1-B1-V-B2-J-B3-U2-3' (I).

5'-U1-B1-V-I-B2-N-J-B3-U2-3' (II).

55 Los oligonucleótidos plantilla constituyentes son diversos con respecto a las secuencias de nucleótidos de los oligonucleótidos plantilla individuales.

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

65 En una modalidad, I representada en la fórmula general II es una secuencia oligonucleotídica de marcador interno que comprende al menos 2 nucleótidos, y no más de 100 nucleótidos.

5 B1, B2, y B3 pueden ser cada uno independientemente nada o cada uno comprende un oligonucleótido "B" que comprende una secuencia oligonucleotídica de código de barras 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 (que incluyen todos los valores enteros intermedios). En algunas modalidades, cada uno de B1, B2, y B3 puede comprender una secuencia oligonucleotídica única que identifica de manera única, o identifica como una combinación pareada, (i) la secuencia oligonucleotídica V única del oligonucleótido plantilla y (ii) la secuencia oligonucleotídica J única del oligonucleótido plantilla.

10 La ubicación relativa de los oligonucleótidos código de barras B1, B2, y B3 y los adaptadores universales U1 y U2 permite ventajosamente la identificación y cuantificación rápida de los productos de amplificación de un oligonucleótido plantilla único dado mediante lecturas de secuencias cortas y secuenciación de extremos pareados en secuenciadores de ADN automatizados (*por ejemplo*, HiSeq™ de Illumina o MiSEQ® de Illumina, o GeneAnalyzer™-2, Illumina Corp., San Diego, CA). En particular, estas modalidades y las relacionadas permiten la determinación rápida de alta productividad de combinaciones específicas de una secuencia de segmento V y una secuencia de segmento J que están presentes en un producto de amplificación, para caracterizar de este modo la eficiencia de amplificación relativa de cada cebador específico de V y cada cebador específico de J que pueden estar presentes en un conjunto de cebadores, que es capaz de amplificar el ADN reordenado que codifica TCR o BCR en una muestra. La comprobación de las identidades y/o cantidades de los productos de amplificación puede lograrse con lecturas de secuencias más largas, que incluyen opcionalmente lecturas de secuencias que se extienden hasta B2.

20 V puede ser nada o un polinucleótido que comprende al menos 20, 30, 60, 90, 120, 150, 180, o 210, y no más de 1.000, 900, 800, 700, 600 o 500 nucleótidos contiguos de una secuencia de ADN. En algunas modalidades, la secuencia de ADN es de una secuencia génica que codifica la región variable (V) del receptor de inmunidad adaptativa, o el complemento de esta, y en cada una de la pluralidad de secuencias oligonucleotídicas plantilla V comprende una secuencia oligonucleotídica única.

30 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 modalidades, la secuencia de ADN es de una secuencia génica que codifica la región de unión (J) del receptor de inmunidad adaptativa, o el complemento de esta, y en cada una de la pluralidad de secuencias oligonucleotídicas plantilla J comprende una secuencia oligonucleotídica única.

35 En la construcción de las porciones "V" y "J" de los oligonucleótidos plantilla sintéticos de la fórmula I o II, pueden usarse varias secuencias génicas de región variable (V) y región de unión (J) del receptor de inmunidad adaptativa. Un gran número de secuencias génicas de región V y J se conocen como secuencias de nucleótidos y/o de aminoácidos, que incluyen secuencias de ADN genómico no reordenadas de loci de TCR e Ig, y secuencias de ADN reordenadas de manera productiva en tales loci y sus productos codificados, y que incluyen además pseudogenes en estos loci, y que incluyen además huérfanos relacionados. *Ver, por ejemplo*, los documentos US2012/0058902, US2010/0330571, WO2011/106758, WO2012/027503. Por otra parte, las secuencias genómicas de los genes de región V de TCR y BCR de seres humanos y otras especies se conocen y están disponibles en bases de datos públicas tales como Genbank. Las secuencias génicas de región V incluyen secuencias polinucleotídicas que codifican los productos de genes de TCR y BCR expresados, reordenados y también incluyen secuencias polinucleotídicas de pseudogenes que se han identificado en los loci de región V. Las diversas secuencias polinucleotídicas V que pueden incorporarse en las plantillas descritas en la presente descripción de la fórmula general I o II pueden variar ampliamente en longitud, en composición de nucleótidos (*por ejemplo*, contenido de GC), y en la secuencia polinucleotídica lineal real, y se conoce, por ejemplo, que incluyen "puntos calientes" o regiones hipervariables que exhiben una diversidad de secuencias particular. Estas y otras secuencias conocidas en la técnica pueden usarse de acuerdo con la presente descripción para el diseño y producción de oligonucleótidos plantilla a incluir en la composición de plantillas proporcionada en la presente descripción para estandarizar la eficiencia de amplificación de un conjunto de cebadores oligonucleotídicos, y para el diseño y producción del conjunto de cebadores oligonucleotídicos que es capaz de amplificar ADN reordenado que codifica cadenas polipeptídicas de TCR o Ig, cuyo ADN reordenado puede estar presente en una muestra biológica que comprende ADN de células linfoides.

55 La secuencia polinucleotídica completa de cada polinucleótido V en la fórmula general I o II puede, pero no necesariamente, consistir exclusivamente en nucleótidos contiguos de cada gen V distinto. Por ejemplo y de acuerdo con ciertas modalidades, en la composición de plantillas descrita en la presente descripción, cada polinucleótido V de la fórmula I o II solo necesita tener al menos una región que comprende una secuencia oligonucleotídica V única que se encuentra en un gen V y con la cual puede aparearse específicamente un solo cebador de región V en el conjunto de cebadores. Por lo tanto, el polinucleótido V de la fórmula I o II puede comprender la totalidad o cualquier porción 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 génica V de origen natural (que incluye una secuencia de pseudogén V), siempre que se incluya al menos una región de secuencia oligonucleotídica V única (*por ejemplo*, el sitio de apareamiento al cebador) que no esté incluida en ningún otro polinucleótido V plantilla.

65 En algunas modalidades, la pluralidad de polinucleótidos V que están presentes en la composición de plantillas sintéticas tienen longitudes que simulan las longitudes completas de secuencias de nucleótidos de genes V conocidas, de origen

natural, aun cuando las secuencias de nucleótidos específicas difieren entre la región V plantilla y cualquier gen V de origen natural. Las longitudes de las regiones V en las plantillas sintéticas pueden diferir de las longitudes de las secuencias génicas V de origen natural en no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 por ciento. Opcionalmente y de acuerdo con ciertas modalidades, el polinucleótido V del oligonucleótido plantilla sintético descrito en la presente descripción incluye un codón de terminación en o cerca del extremo 3' de V en la fórmula general I o II.

Por lo tanto el polinucleótido V en la fórmula (I) puede comprender, en ciertas modalidades, una secuencia de nucleótidos que tiene una longitud que es igual o similar a la longitud de un gen típico desde su codón de inicio hasta su región que codifica CDR3 y puede, pero no necesariamente, incluir una secuencia de nucleótidos que codifica la región CDR3. Las secuencias de nucleótidos que codifican CDR3 y las longitudes de las secuencias pueden variar considerablemente y se han caracterizado mediante varios esquemas de numeración diferentes (*por ejemplo*, Lefranc, 1999 *The Immunologist* 7:132; Rabat y otros, 1991 *En: Sequences of Proteins of Immunological Interest*, Publicación de NIH 91-3242; Chothia y otros, 1987 *J. Mol. Biol.* 196:901; Chothia y otros, 1989 *Nature* 342:877; Al-Lazikani y otros, 1997 *J Mol Biol.* 273:927; ver además, *por ejemplo*, Rock y otros, 1994 *J. Exp. Med.*, 179:323; Saada y otros, 2007 *Immunol. Cell Biol.* 85:323).

Brevemente, la región CDR3 típicamente abarca la porción de polipéptido que se extiende desde un residuo de cisteína altamente conservado (codificado por el codón trinucleótido TGY; Y = T o C) en el segmento hasta un residuo de fenilalanina altamente conservado (codificado por TTY) en el segmento J de los TCR, o hasta un triptófano altamente conservado (codificado por TGG) en IGH. Más de 90 % de los reordenamientos productivos naturales, en el locus de TCRB tienen una longitud que codifica CDR3 según este criterio de entre 24 y 54 nucleótidos, correspondientes a entre 9 y 17 aminoácidos codificados. Las longitudes de CDR3 de los oligonucleótidos plantilla sintéticos descritos en la presente descripción deben residir, para cualquier locus de TCR o BCR dado, dentro del mismo intervalo que el 95 % de los reordenamientos de origen natural. Por lo tanto, por ejemplo, en una composición de plantillas sintéticas descrita en la presente descripción, la porción que codifica CDR3 del polinucleótido V tiene una longitud de 24 a 54 nucleótidos, que incluye cada número entero intermedio. Los esquemas de numeración para las regiones que codifican CDR3 descritas anteriormente denotan las posiciones de los codones de cisteína, fenilalanina y triptófano, conservados y estos esquemas de numeración también pueden aplicarse a pseudogenes en los que uno o más codones que codifican estos aminoácidos conservados pueden haberse reemplazado con un codón que codifica un aminoácido diferente. Para los pseudogenes que no usan estos aminoácidos conservados, la longitud de CDR3 puede definirse con relación a la posición correspondiente en la que el residuo conservado se habría observado en ausencia de la sustitución, de acuerdo con uno de los esquemas de numeración de posición de secuencia de CDR3 establecidos a los que se hizo referencia anteriormente.

La secuencia polinucleotídica completa de cada polinucleótido J en la fórmula general I o II puede, pero no necesariamente, consistir exclusivamente en nucleótidos contiguos de cada gen J distinto. Por ejemplo y de acuerdo con ciertas modalidades, en la composición de plantillas descrita en la presente descripción, cada polinucleótido J de la fórmula I o II solo necesita tener al menos una región que comprende una secuencia oligonucleotídica J única que se encuentra en un gen J y con la cual puede aparearse específicamente un solo cebador de región V en el conjunto de cebadores. Por lo tanto, el polinucleótido V de la fórmula I o II puede comprender la totalidad o cualquier porción 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 génica V de origen natural (que incluye una secuencia de pseudogén V) siempre que se incluya al menos una región de secuencia oligonucleotídica V única (el sitio de apareamiento al cebador) que no esté incluida en ningún otro polinucleótido J plantilla.

Puede preferirse en ciertas modalidades que la pluralidad de polinucleótidos J que están presentes en la composición de plantillas descrita en la presente descripción tengan longitudes que simulen las longitudes completas de secuencias de nucleótidos de genes J conocidas, de origen natural, aun cuando las secuencias de nucleótidos específicas difieren entre la región J plantilla y cualquier gen J de origen natural. Las longitudes de las regiones J en las plantillas descritas en la presente descripción pueden diferir de las longitudes de las secuencias génicas J de origen natural en no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 por ciento.

Por lo tanto el polinucleótido J en la fórmula I o II puede comprender, en ciertas modalidades, una secuencia de nucleótidos que tiene una longitud que es igual o similar a la longitud de un gen J de origen natural típico y puede, pero no necesariamente, incluir una secuencia de nucleótidos que codifica la región CDR3, como se analizó anteriormente.

Las secuencias genómicas de los genes de región J de TCR y BCR de seres humanos y otras especies se conocen y están disponibles en bases de datos públicas tales como Genbank; las secuencias génicas de región J incluyen secuencias polinucleotídicas que codifican los productos de genes de TCR y BCR reordenados expresados y no expresados. Las diversas secuencias polinucleotídicas J que pueden incorporarse en las plantillas descritas en la presente descripción de la fórmula general I o II pueden variar ampliamente en longitud, en composición de nucleótidos (*por ejemplo*, contenido de GC), y en secuencia polinucleotídica lineal real.

Las alternativas a las secuencias V y J descritas en la presente descripción, para el uso en la construcción de los oligonucleótidos plantilla y/o cebadores oligonucleotídicos de segmentos V y segmentos J descritos en la presente descripción, pueden seleccionarse por un experto en base a la presente descripción mediante el uso del conocimiento en

la técnica con respecto a secuencias génicas publicadas para las regiones que codifican V y J de los genes para cada subunidad de TCR e Ig. Las entradas a Genbank de las referencias de secuencias humanas de receptor de inmunidad adaptativa incluyen: TCR α : (TCRA/D): NC_000014.8 (cro14:22090057..23021075); TCR β (TCRB): NC_000007.13 (cro7:141998851..142510972); TCR γ : (TCRG): NC_000007.13 (cro7:38279625..38407656); cadena pesada de inmunoglobulina, IgH (IGH): NC_000014.8 (cro14:106032614..107288051); cadena ligera kappa de inmunoglobulina, IgL κ (IGK): NC_000002.11 (cro2:89156874..90274235); y cadena ligera lambda de inmunoglobulina, IgL λ (IGL): NC_000022.10 (cro22:22380474..23265085). Las entradas a Genbank de las referencias de secuencias de ratón de loci de receptor de inmunidad adaptativa incluyen: TCR β : (TCRB): NC_000072.5 (cro6: 40841295..41508370), y cadena pesada de inmunoglobulina, IgH (IGH): NC_000078.5 (cro12: 114496979.. 117248165).

Los análisis de diseño de plantillas y cebadores y las consideraciones de selección de los sitios diana pueden realizarse, por ejemplo, mediante el uso del software de análisis de cebadores OLIGO y/o el software de algoritmo BLASTN 2.0.5 (Altschul y otros, *Nucleic Acids Res.* 1997, 25(17):3389-402), u otros programas similares disponibles en la técnica.

En consecuencia, en base a la presente descripción y en vista de estas secuencias génicas de receptor de inmunidad adaptativa conocidas y las metodologías de diseño de oligonucleótidos, para la inclusión en los presentes oligonucleótidos plantilla los expertos en la técnica pueden diseñar una pluralidad de secuencias polinucleotídicas específicas de región V y específicas de región J que contienen cada una independientemente secuencias oligonucleotídicas que son únicas para un gen V y J dado, respectivamente. De manera similar, a partir de la presente descripción y en vista de las secuencias de receptor de inmunidad adaptativa conocidas, los expertos en la técnica pueden diseñar además un conjunto de cebadores que comprende una pluralidad de cebadores oligonucleotídicos específicos de región V y específicos de región J que son cada uno independientemente capaces de aparearse con una secuencia específica que es única para un gen V y J dado, respectivamente, por medio de lo cual la pluralidad de cebadores es capaz de amplificar sustancialmente todos los genes V y sustancialmente todos los genes J en un locus que codifica receptor de inmunidad adaptativa dado (por ejemplo, un locus de TCR o IGH humano). Tales conjuntos de cebadores permiten la generación, en PCR múltiple (por ejemplo, mediante el uso de múltiples pares de cebadores directos e inversos), de productos de amplificación que tienen un primer extremo que está codificado por un segmento génico reordenado que codifica la región V y un segundo extremo que está codificado por un segmento génico que codifica la región J.

Típicamente, tales productos de amplificación pueden incluir una secuencia que codifica CDR3. Los cebadores pueden diseñarse preferentemente para producir productos de amplificación que tienen porciones suficientes de secuencias V y J y/o de secuencias de código de barras (B) de V-J como se describe en la presente descripción, de manera que mediante la secuenciación de los productos (amplicones), es posible identificar en base a las secuencias que son únicas para cada segmento génico (i) el gen V particular, y (ii) el gen J particular en cuya proximidad el gen V experimentó reordenamiento para producir un gen que codifica un receptor de inmunidad adaptativa funcional. Típicamente, y en modalidades preferidas, los productos de amplificación de PCR tendrán un tamaño de no más de 600 pares de bases, lo que de acuerdo con la teoría no limitante excluirá los productos de amplificación de genes de receptor de inmunidad adaptativa no reordenados. En ciertas otras modalidades preferidas los productos de amplificación tendrán un tamaño de no más de 500, 400, 300, 250, 200, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30 o 20 pares de bases, tal como puede proporcionar ventajosamente la cuantificación rápida, de alta productividad de amplicones de secuencias distintas mediante lecturas de secuencias cortas.

En una modalidad de la fórmula I o II, V es una secuencia polinucleotídica que codifica al menos 10-70 aminoácidos contiguos de una región V del receptor de inmunidad adaptativa, o el complemento de esta; J es una secuencia polinucleotídica que codifica al menos 5-30 aminoácidos contiguos de una región J del receptor de inmunidad adaptativa, o el complemento de esta; U1 y U2 son cada uno nada o comprenden un oligonucleótido que comprende una secuencia de nucleótidos que se selecciona de (i) una secuencia oligonucleotídica adaptadora universal, y (ii) una secuencia oligonucleotídica específica de plataforma de secuenciación unida a y posicionada 5' a la secuencia oligonucleotídica adaptadora universal; B1, B2, y B3 son cada uno independientemente nada o cada uno comprende un oligonucleótido B que comprende una secuencia oligonucleotídica de código de barras de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos contiguos, en donde en cada una de la pluralidad de secuencias oligonucleotídicas, B comprende una secuencia oligonucleotídica única que identifica de manera única, como una combinación pareada, (i) la secuencia oligonucleotídica V única y (ii) la secuencia oligonucleotídica J única.

En otra modalidad de la 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 génica de región variable (V) de receptor de inmunidad adaptativa (por ejemplo, TCR o BCR), o el complemento de esta, y en cada una de la pluralidad de secuencias oligonucleotídicas V comprende una secuencia oligonucleotídica única.

La descripción adicional acerca de oligonucleótidos plantilla sintéticos puede encontrarse en el documento WO2013/169957.

La Figura 1A ilustra un ejemplo de un oligonucleótido plantilla sintético, de acuerdo con una modalidad de la invención. En una modalidad, un oligonucleótido plantilla 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 específico de plantilla

(BC) (104), una secuencia que comprende una porción de o la totalidad de una secuencia génica única que codifica la región variable (V) del receptor de inmunidad adaptativa (gen V) (106), un marcador interno de plantilla sintética (IM) (108), una repetición del código de barras (BC) (104), una repetición del marcador interno (IM) (108), una secuencia que comprende una porción de o la totalidad de una secuencia génica única que codifica la región variable (J) del receptor de inmunidad adaptativa (gen J) (110), una tercera repetición del código de barras (BC) (104), y una secuencia de cebador universal inverso (UB) (112). Cada oligonucleótido plantilla sintético incluye una secuencia génica única que codifica la región variable (V) del receptor de inmunidad adaptativa y una secuencia génica única que codifica la región de unión (J) del receptor de inmunidad adaptativa. La combinación de secuencias V y J en los oligonucleótidos plantilla sintéticos son iguales a las que se encuentran en moléculas biológicas que comprenden combinaciones únicas de secuencias V y J reordenadas en la muestra.

En un ejemplo, el oligonucleótido plantilla sintético puede ser una secuencia de 495 pb que comprende una secuencia de cebador universal (UA) (102), un código de barras específico de plantilla (BC) de 16 pb (104), una secuencia génica que codifica la región variable (V) del receptor de inmunidad adaptativa (gen V) de 300 pb (106), un marcador interno de plantilla sintética (IM) de 9 pb (108), una repetición del código de barras (BC) (104), una repetición del marcador interno (IM) (108), una secuencia génica que codifica la región variable (J) del receptor de inmunidad adaptativa (gen J) de 100 pb (110), una tercera repetición del código de barras (BC) (104), y una secuencia de cebador universal inverso (UB) (112). Varias longitudes de las secuencias y el orden de las regiones pueden usarse en el diseño de los oligonucleótidos plantilla sintéticos, como conoce el experto en la técnica.

Los oligonucleótidos plantilla sintéticos de la Fórmula I pueden incluir además secuencias adaptadoras. Las secuencias adaptadoras pueden añadirse a los oligonucleótidos plantilla sintéticos mediante el diseño de cebadores que incluyen secuencias adaptadoras en sus extremos 5' y que se hibridan específicamente a las regiones adaptadoras UA y UB en los oligonucleótidos plantilla sintéticos (ver la Figura 1(A)). Un ejemplo de una secuencia adaptadora es una secuencia adaptadora de Illumina, como se describe en la sección "Adaptadores" a continuación.

En una modalidad, los amplicones de oligonucleótidos plantilla sintéticos resultantes tienen la estructura de la fórmula general I y pueden incluir una secuencia adaptadora o secuencias adaptadoras (secuencia de Illumina), de manera que la secuencia del oligonucleótido plantilla sintético comprende lo siguiente: una secuencia adaptadora, una secuencia de cebador universal (UA) (102), un código de barras específico de plantilla (BC) (104), una secuencia génica que codifica la región variable (V) del receptor de inmunidad adaptativa (gen V) (106), un marcador interno de plantilla sintética (IM) (108), una repetición del código de barras (BC) (104), una repetición del marcador interno (IM) (108), una secuencia génica que codifica la región variable (J) del receptor de inmunidad adaptativa (gen J) (110), una tercera repetición del código de barras (BC) (104), una secuencia de cebador universal inverso (UB) (112), y una segunda secuencia adaptadora.

Número de oligonucleótidos plantilla sintéticos en la muestra

En ciertas modalidades, la composición de plantillas sintéticas comprende una pluralidad de oligonucleótidos plantilla sintéticos distintos y únicos. En una modalidad, la pluralidad de oligonucleótidos plantilla sintéticos comprende al menos a o al menos b secuencias oligonucleotídicas únicas, cualquiera que sea mayor, en donde a es el número de segmentos génicos únicos que codifican la región V del receptor de inmunidad adaptativa en el sujeto y b es el número de segmentos génicos únicos que codifican la región J del receptor de inmunidad adaptativa en el sujeto, y la composición comprende al menos un oligonucleótido plantilla para cada polinucleótido V único y al menos un oligonucleótido plantilla para cada polinucleótido J único.

En otra modalidad, la pluralidad de oligonucleótidos plantilla comprende al menos $(a \times b)$ secuencias oligonucleotídicas únicas, donde a es el número de segmentos génicos únicos que codifican la región V del receptor de inmunidad adaptativa en el sujeto y b es el número de segmentos génicos únicos que codifican la región J del receptor de inmunidad adaptativa en el sujeto, y la composición comprende al menos un oligonucleótido plantilla para cada combinación posible de un segmento génico que codifica la región V y un segmento génico que codifica la región J.

En consecuencia, la composición puede dar cabida al menos a una aparición de cada secuencia polinucleotídica V única y al menos una aparición de cada secuencia polinucleotídica J única, donde en algunos casos la al menos una aparición de un polinucleótido V único particular estará presente en el mismo oligonucleótido plantilla en el que puede encontrarse al menos una aparición de un polinucleótido J único particular. Por lo tanto, por ejemplo, "al menos un oligonucleótido plantilla para cada polinucleótido V único y al menos un oligonucleótido plantilla para cada polinucleótido J único" puede referirse en ciertos casos a un solo oligonucleótido plantilla en el que están presentes un polinucleótido V único y un polinucleótido J único.

En una modalidad, a es 1 hasta un número máximo de segmentos génicos V en el genoma de mamífero del sujeto. En otra modalidad, b es 1 hasta un número máximo de segmentos génicos J en el genoma de mamífero del sujeto. En otras modalidades, a es 1. En otras modalidades, b es 1.

En algunas modalidades, a puede estar en el intervalo 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 modalidades, *b* puede estar en el intervalo 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 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 ciertas modalidades, una mezcla de oligonucleótidos plantilla sintéticos que comprende cada combinación posible de un segmento génico que codifica la región V y un segmento génico que codifica la región J comprende 248 tipos de plantillas sintéticas únicas para TCRA/D, 858 tipos de plantillas sintéticas únicas para TCRB, 70 tipos de plantillas sintéticas únicas para TCRG, 1.116 tipos de plantillas sintéticas únicas para IGH, y 370 tipos de plantillas sintéticas únicas para IGK/L.

La tabla a continuación enumera el número de segmentos génicos V (*a*) y de segmentos génicos J (*b*) para cada uno de los loci de receptor de inmunidad adaptativa humanos, que incluyen segmentos V y J funcionales.

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

* Total de genes de segmentos variables y de unión

** Genes de segmentos variables y de unión con al menos un alelo funcional

En algunas modalidades, el polinucleótido J del oligonucleótido plantilla 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 de inmunidad adaptativa, o el complemento de esta.

No se pretende limitar así la invención que se contempla en la presente descripción, sin embargo, de manera que en ciertas modalidades, puede usarse ventajosamente un número de oligonucleótidos plantilla sustancialmente menor. En estas modalidades y las relacionadas, donde *a* es el número de segmentos génicos únicos que codifican la región V del receptor de inmunidad adaptativa en un sujeto y *b* es el número de segmentos génicos únicos que codifican la región J del receptor de inmunidad adaptativa en el sujeto, el número mínimo de secuencias oligonucleotídicas únicas que constituyen la pluralidad de oligonucleótidos plantilla sintéticos puede determinarse por cualquiera que sea el mayor de *a* y *b*, siempre que cada secuencia polinucleotídica V única y cada secuencia polinucleotídica J única esté presente en al menos un oligonucleótido plantilla sintético en la composición de plantillas. Por lo tanto, de acuerdo con ciertas modalidades relacionadas, la composición de plantillas puede comprender al menos un oligonucleótido plantilla sintético para cada polinucleótido V único, *por ejemplo*, que incluye uno solo de cada polinucleótido V único de acuerdo con la fórmula general I o II, y al menos un oligonucleótido plantilla sintético para cada polinucleótido J único, *por ejemplo*, que incluye uno solo de cada polinucleótido J único de acuerdo con la fórmula general I o II.

En ciertas otras modalidades, la composición de plantillas comprende al menos un oligonucleótido plantilla sintético con el que puede aparearse cada cebador oligonucleotídico de amplificación en un conjunto de cebadores de amplificación.

Es decir, en ciertas modalidades, la composición de plantillas comprende al menos un oligonucleótido plantilla sintético que tiene una secuencia oligonucleotídica de la fórmula general (I) a la que puede hibridarse específicamente cada cebador oligonucleotídico de segmento V, y al menos un oligonucleótido plantilla sintético que tiene una secuencia oligonucleotídica de la fórmula general (I) a la que puede hibridarse específicamente cada cebador oligonucleotídico de segmento J.

De acuerdo con tales modalidades, el conjunto de cebadores oligonucleotídicos que es capaz de amplificar el ADN reordenado que codifica uno o una pluralidad de receptores de inmunidad adaptativa comprende una pluralidad *a'* de cebadores oligonucleotídicos de segmentos V únicos y una pluralidad *b'* de cebadores oligonucleotídicos de segmentos J únicos. La pluralidad *a'* de cebadores oligonucleotídicos de segmentos V son cada uno independientemente capaces

de aparearse o hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región V del receptor de inmunidad adaptativa o al complemento de este, en donde cada cebador de segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico que codifica la región V del receptor de inmunidad adaptativa. La pluralidad *b'* de cebadores oligonucleotídicos de segmentos J son cada uno independientemente capaces de aparearse o hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región J del receptor de inmunidad adaptativa o al complemento de este, en donde cada cebador de segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico que codifica la región J del receptor de inmunidad adaptativa.

En algunas modalidades, *a'* es igual que *a* (descrito anteriormente para oligonucleótidos plantilla sintéticos). En otras modalidades, *b'* es igual que *b* (descrito anteriormente para oligonucleótidos plantilla sintéticos).

Por lo tanto, en ciertas modalidades y como se analiza además en otra parte en la presente descripción, la presente composición de plantillas sintéticas puede usarse en reacciones de amplificación con cebadores de amplificación diseñados para amplificar todas las secuencias génicas reordenadas que codifican el receptor de inmunidad adaptativa, que incluyen aquellas que no se expresan. En ciertas otras modalidades, la composición de plantillas y los cebadores de amplificación pueden diseñarse de modo que no produzcan productos de amplificación de genes reordenados que no se expresan (*por ejemplo*, pseudogenes, huérfanos). Por consiguiente, se apreciará que en ciertas modalidades solo un subconjunto de genes reordenados que codifican el receptor de inmunidad adaptativa puede amplificarse convenientemente, de manera que pueden diseñarse subconjuntos de cebadores de amplificación adecuados y emplearse para amplificar solamente las secuencias V-J reordenadas que son de interés. En estas modalidades y las relacionadas, en consecuencia, puede usarse una composición de plantillas sintéticas que comprende solamente un subconjunto de interés de secuencias V-J reordenadas, siempre que la composición de plantillas sintéticas comprenda al menos un oligonucleótido plantilla sintético con el que puede aparearse cada cebador oligonucleotídico de amplificación en un conjunto de cebadores de amplificación. Por lo tanto, el número real de oligonucleótidos plantilla sintéticos en la composición de plantillas puede variar considerablemente entre las modalidades contempladas, en función del conjunto de cebadores de amplificación que se usará.

Por ejemplo, en ciertas modalidades relacionadas, en la composición de plantillas, la pluralidad de oligonucleótidos plantilla sintéticos comprende las SEQ ID NO:707-3003.

Cebadores para el uso con oligonucleótidos plantilla sintéticos

El polinucleótido V en la fórmula general I o II (o su complemento) incluye secuencias con las que pueden aparearse específicamente los miembros de los conjuntos de cebadores oligonucleotídicos específicos para los genes de TCR o BCR. Los conjuntos de cebadores que son capaces de amplificar el ADN reordenado que codifica una pluralidad de TCR o BCR se describen, por ejemplo, en los documentos US2012/0058902, US2010/0330571, WO2011/106738 o WO2012/027503; o similares; o como se describe allí pueden diseñarse de modo que incluyan secuencias oligonucleotídicas que puedan hibridarse específicamente a cada gen V único y a cada gen J en un locus génico de TCR o BCR particular (*por ejemplo*, TCR α , β , γ o δ , o IgH μ , γ , δ , α o ϵ , o IgL κ o λ).

Por ejemplo, a modo de ilustración y no de limitación, un cebador oligonucleotídico de un conjunto de cebadores oligonucleotídicos de amplificación que es capaz de amplificar el ADN reordenado que codifica uno o una pluralidad de TCR o BCR puede incluir típicamente una secuencia de nucleótidos de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40 nucleótidos contiguos, o más, y puede aparearse 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 uno J como se proporciona en la presente descripción. En ciertas modalidades los cebadores pueden comprender al menos 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 nucleótidos, y en cierta modalidad los cebadores pueden comprender secuencias de no más de 40 nucleótidos contiguos. Los cebadores y los sitios de apareamiento al cebador de otras longitudes también se contemplan expresamente, como se describe en la presente descripción.

El polinucleótido J en la fórmula general (I) (o su complemento) incluye secuencias con las que pueden aparearse específicamente los miembros de los conjuntos de cebadores oligonucleotídicos específicos para los genes de TCR o BCR. Los conjuntos de cebadores que son capaces de amplificar el ADN reordenado que codifica una pluralidad de TCR o BCR se describen, por ejemplo, en los documentos US2012/0058902, US2010/0330571, WO2011/106758 o WO2010/027503; o similares; o como se describe allí pueden diseñarse de modo que incluyan secuencias oligonucleotídicas que puedan hibridarse específicamente a cada gen V único y a cada gen J único en un locus génico de TCR o BCR particular (*por ejemplo*, TCR α , β , γ o δ , o IgH μ , γ , δ , α o ϵ , o IgL κ o λ).

Estos cebadores oligonucleotídicos de segmentos V y segmentos J pueden comprender secuencias adaptadoras universales en sus extremos 5' para la secuenciación de los amplicones resultantes, como se describió anteriormente y en los documentos US2012/0058902, US2010/0330571, WO2011/106758 o WO2012/027503. La Figura 1B ilustra cebadores que se hibridan a regiones específicas de las secuencias de segmentos V y segmentos J e incluyen además secuencias adaptadoras universales.

En ciertas modalidades, los conjuntos de cebadores oligonucleotídicos para la amplificación pueden proporcionarse en cantidades sustancialmente equimolares. Como también se describe en la presente descripción, de acuerdo con ciertas otras modalidades, la concentración de uno o más cebadores en un conjunto de cebadores puede ajustarse deliberadamente de modo que ciertos cebadores no estén presentes en cantidades equimolares o en cantidades sustancialmente equimolares.

Adaptadores

Los oligonucleótidos plantilla descritos en la presente descripción de la fórmula general (I) también pueden comprender en ciertas modalidades una primera (U1) (102) y segunda (U2) (112) secuencias oligonucleotídicas adaptadoras universales, o pueden carecer de cualquiera o ambas de U1 (102) y U2 (112). Por lo tanto U1 (102) puede comprender nada o un oligonucleótido que tiene una secuencia que se selecciona de (i) una primera secuencia oligonucleotídica adaptadora universal, y (ii) una primera secuencia oligonucleotídica específica de plataforma de secuenciación unida a y posicionada 5' a una primera secuencia oligonucleotídica adaptadora universal, y U2 (112) puede comprender nada o un oligonucleótido que tiene una secuencia que se selecciona de (i) una segunda secuencia oligonucleotídica adaptadora universal, y (ii) una segunda secuencia oligonucleotídica específica de plataforma de secuenciación unida a y posicionada 5' a una segunda secuencia oligonucleotídica adaptadora universal.

U1 (102) y/o U2 (112) pueden comprender, por ejemplo, secuencias oligonucleotídicas adaptadoras universales y/o secuencias oligonucleotídicas específicas de plataforma de secuenciación que son específicas para una tecnología de secuenciación de moléculas individuales que se emplea, por ejemplo, los sistemas HiSeq™ o GeneAnalyzer™-2 (GA-2) (Illumina, Inc., San Diego, CA) u otro paquete de secuenciación adecuado de instrumentación, reactivos y software. La inclusión de tales secuencias adaptadoras específicas de plataforma permite la secuenciación cuantitativa directa de la composición de plantillas descrita en la presente descripción, que comprende una pluralidad de oligonucleótidos plantilla diferentes de la fórmula general (I), mediante el uso de una metodología de secuenciación de nucleótidos tal como el sistema HiSeq™ o GA2 o equivalentes. Por consiguiente, esta característica permite ventajosamente la caracterización cualitativa y cuantitativa de la composición de plantillas.

En particular, la capacidad de secuenciar todos los componentes de la composición de plantillas permite directamente la comprobación de que cada oligonucleótido plantilla en la pluralidad de oligonucleótidos plantilla está presente en una cantidad sustancialmente equimolar. Por ejemplo, puede generarse un conjunto de los oligonucleótidos plantilla descritos en la presente descripción que tienen secuencias adaptadoras universales en ambos extremos, de modo que las secuencias adaptadoras pueden usarse para incorporar además oligonucleótidos específicos de plataforma de secuenciación en cada extremo de cada plantilla.

Sin desear limitarse por la teoría, los oligonucleótidos específicos de plataforma pueden añadirse en los extremos de tales plantillas modificadas mediante el uso de oligonucleótidos 5' (5'-secuencia de plataforma-secuencia adaptadora universal-1-3') y 3' (5'-secuencia de plataforma-secuencia adaptadora universal-2-3') en tan poco como dos ciclos de desnaturalización, apareamiento y extensión, de modo que la representación relativa en la composición de plantillas de cada uno de los oligonucleótidos plantilla componentes no se altera cuantitativamente. Las secuencias de identificación únicas (*por ejemplo*, secuencias de código de barras B que comprenden secuencias oligonucleotídicas V y B únicas que se asocian con y por lo tanto identifican, respectivamente, regiones V y J individuales, como se describe en la presente descripción) se ubican adyacentes a las secuencias adaptadoras, lo que permite por lo tanto la secuenciación cuantitativa en lecturas de secuencias cortas, para caracterizar la población de plantillas según el criterio de la cantidad relativa de cada secuencia plantilla única que está presente.

Cuando tal secuenciación cuantitativa directa indica que uno o más oligonucleótidos particulares puede estar sobrerrepresentado o subrepresentado en una preparación de la composición de plantillas, puede realizarse el ajuste de la composición de plantillas en consecuencia para obtener una composición de plantillas en la que todos los oligonucleótidos estén presentes en cantidades sustancialmente equimolares. La composición de plantillas en la que todos los oligonucleótidos están presentes en cantidades sustancialmente equimolares puede usarse entonces como un estándar de calibración para conjuntos de cebadores de amplificación, tal como en los métodos descritos en la presente descripción para determinar y corregir la amplificación no uniforme potencial entre los miembros de un conjunto de cebadores.

Cuando a los cebadores se les anexan los adaptadores universales + Illumina y se secuencian con los adaptadores de Illumina (ver la Figura 1), estas plantillas se comportan de la misma manera que las plantillas sintéticas típicas. Cuando se amplifican mediante el uso de cebadores de PCR múltiple 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 génica J (de alrededor de 15 pares de bases), (2) un marcador interno de plantilla sintética (IM) de 9 pares de bases, (3) un código de barras (BC) de V-J de 16 pares de bases, (4) un segundo marcador interno de plantilla sintética (IM) de 9 pares de bases, y (5) un gen V (de alrededor de 15 pares de bases).

Además de las secuencias adaptadoras descritas en las SEQ ID NO:765-786, otras secuencias oligonucleotídicas que pueden usarse como secuencias adaptadoras universales serán conocidas para aquellos familiarizados con la técnica en vista de la presente descripción, que incluye la selección de secuencias oligonucleotídicas adaptadoras que son distintas

a las secuencias que se encuentran en otras porciones de las plantillas descritas en la presente descripción.

Códigos de barras

5 Como se describe en la presente descripción, ciertas modalidades contemplan el diseño de secuencias oligonucleotídicas
 plantilla de modo que contengan secuencias de identificación cortas que permiten la identificación inequívoca de la
 secuencia plantilla, y por tanto de al menos un cebador responsable de la amplificación de esa plantilla, sin tener que
 secuenciar el producto de amplificación completo. En los oligonucleótidos plantilla sintéticos descritos en la presente
 descripción de la fórmula general (I), B1, B2, B3, y B4 son cada uno independientemente nada o cada uno comprende un
 10 oligonucleótido B que comprende una secuencia oligonucleotídica de código de barras 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
 o más nucleótidos contiguos (que incluyen todos los valores enteros intermedios), en donde en cada una de la pluralidad
 de secuencias oligonucleotídicas plantilla B comprende una secuencia oligonucleotídica única que identifica de manera
 única, como una combinación pareada, (i) la secuencia oligonucleotídica V única del oligonucleótido plantilla y (ii) la
 15 secuencia oligonucleotídica J única del oligonucleótido plantilla.

Por lo tanto, por ejemplo, los oligonucleótidos plantilla sintéticos que tienen secuencias de identificación de código de
 barras pueden permitir 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,
 20 45, 40, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4 o menos nucleótidos, seguido de hacer
 corresponder esta información de secuencias de código de barras con las secuencias V y J asociadas que se incorporan
 en la plantilla que tiene el código de barras como parte del diseño de plantillas. Mediante este enfoque, un gran número
 de productos de amplificación puede secuenciarse de manera parcial simultáneamente mediante secuenciación en
 paralelo de alta productividad, para identificar los cebadores que son responsables del sesgo de amplificación en un
 25 conjunto de cebadores complejo.

Los códigos de barras ilustrativos pueden comprender un primer oligonucleótido código de barras de 5, 6, 7, 8, 9, 10, 11,
 12, 13, 14, 15 o 16 nucleótidos que identifica de manera única cada polinucleótido V en la plantilla y un segundo
 oligonucleótido código de barras de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 nucleótidos que identifica de manera única
 cada polinucleótido J en la plantilla, para proporcionar códigos de barras de, respectivamente, 10, 11, 12, 13, 14, 15, 16,
 30 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 o 32 nucleótidos de longitud, pero no se pretende limitar así estas
 modalidades y las relacionadas. Los oligonucleótidos 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 plantilla que tienen secuencias de cualquier otra
 35 manera distintas (*por ejemplo* secuencias V y J).

Por lo tanto, la longitud mínima del código de barras, para evitar tal redundancia entre los códigos de barras que se usan
 para identificar de manera única pares de secuencias V-J diferentes, es X nucleótidos, donde 4^X es mayor que el número
 de especies plantilla distintas a diferenciar en base a que tienen secuencias no idénticas. Por ejemplo, para el conjunto
 de 858 oligonucleótidos plantilla que se expone en la presente descripción en las SEQ ID NO:1888-3003, la longitud
 40 mínima del código de barras sería cinco nucleótidos, lo que permitiría un total teórico de 1.024 secuencias de
 pentanucleótidos diferentes posibles (*es decir*, mayor que 871). En la práctica, las longitudes de lecturas de secuencias
 oligonucleotídicas código de barras pueden limitarse solamente por los límites de longitud de lectura de secuencia del
 instrumento de secuenciación de nucleótidos a emplear. Para ciertas modalidades, los oligonucleótidos código de barras
 45 diferentes que distinguirán especies individuales de oligonucleótidos plantilla deberán tener al menos dos errores de
 coincidencia de nucleótidos (*por ejemplo*, una distancia de Hamming mínima de 2) cuando se alinean para maximizar el
 número de nucleótidos que coinciden en posiciones particulares las secuencias oligonucleotídicas de código de barras.

En modalidades preferidas, para cada oligonucleótido plantilla distinto las especies que tienen una secuencia única dentro
 de la composición de plantillas de la fórmula general (I), B1, B2, B3, y B4 serán idénticas.
 50

El experto estará familiarizado con el diseño, la síntesis, y la incorporación en un constructo de oligonucleótido o
 polinucleótido más grande, de secuencias oligonucleotídicas de código de barras 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, 300,
 55 500 o más nucleótidos contiguos, que incluyen todos los valores enteros intermedios. Para ejemplos no limitantes del
 diseño e implementación de estrategias de identificación de secuencias oligonucleotídicas de código de barras, ver, *por
 ejemplo*, de Career y otros, 2011 *Adv. Env. Microbiol.* 77:6310; Parameswaran y otros, 2007 *Nucl. Ac. Res.* 35(19):330;
 Roh y otros, 2010 *Trends Biotechnol.* 28:291.

Típicamente, los códigos de barras se ubican en las plantillas en ubicaciones donde no se encuentran naturalmente, es
decir, los códigos de barras comprenden secuencias de nucleótidos que son distintas a cualquiera de las secuencias
 oligonucleotídicas de origen natural que pueden encontrarse en las inmediaciones de las secuencias adyacentes a las
 que se sitúan los códigos de barras (*por ejemplo*, secuencias V y/o J). Tales secuencias de código de barras pueden
 incluirse, de acuerdo con ciertas modalidades descritas en la presente descripción, como elementos B1, B2 y/o B3 del
 60 oligonucleótido plantilla descrito en la presente descripción de la fórmula general (I). En consecuencia, ciertos de los
 oligonucleótidos plantilla descritos en la presente descripción de la fórmula general (I) también pueden comprender en
 65

ciertas modalidades uno, dos o los tres códigos de barras B1, B2 y B3, mientras que en ciertas otras modalidades, algunos o la totalidad de estos códigos de barras pueden estar ausentes. En ciertas modalidades todas las secuencias de código de barras tendrán un contenido de GC idéntico o similar (*por ejemplo*, difieren en el contenido de GC en no más de 20 %, o en no más de 19, 18, 17, 16, 15, 14, 13, 12, 11 o 10 %).

5

En las composiciones de plantillas de acuerdo con ciertas modalidades descritas en la presente descripción el elemento B que contiene código de barras (*por ejemplo*, B1, B2, B3, y/o B4) comprende la secuencia oligonucleotídica que identifica de manera única una sola combinación V-J pareada. Opcionalmente y en ciertas modalidades el elemento B que contiene código de barras puede incluir además 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, 300, 500 o más nucleótidos contiguos, situada corriente arriba y/o corriente abajo de la secuencia de código de barras específica que identifica de manera única cada combinación V-J pareada específica. Cuando está presente tanto corriente arriba como corriente abajo de la secuencia de código de barras específica, el nucleótido aleatorio o la secuencia polinucleotídica aleatoria son independientes entre sí, es decir, pueden comprender pero no necesariamente el mismo nucleótido o la misma secuencia polinucleotídica.

10

15

Segmentos de nucleótidos aleatorios

En algunas modalidades, el oligonucleótido plantilla sintético comprende una secuencia oligonucleotídica generada aleatoriamente, o una secuencia de "nucleótidos aleatorios" (110). La secuencia de nucleótidos aleatorios se representa como "N" en la fórmula general II. La secuencia de nucleótidos aleatorios (110) se sitúa generalmente entre las secuencias V y J, pero puede ubicarse en otra parte a lo largo del oligonucleótido plantilla sintético. En una modalidad, la secuencia de nucleótidos aleatorios (110) solo aparece una vez en la plantilla sintética. 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.

20

25

El número de secuencias de nucleótidos posibles de longitud X es 4^X , por lo tanto, un segmento de nucleótidos aleatorios incluso de una longitud pequeña puede codificar muchas secuencias de nucleótidos únicas posibles. Por ejemplo, una secuencia de nucleótidos aleatorios (110) de 12 pares de bases podría codificar cualquiera de 16.777.216 secuencias de nucleótidos únicas. La secuencia de nucleótidos aleatorios (110) garantiza que dos oligonucleótidos plantilla sintéticos cualesquiera tengan una probabilidad de alrededor de 1 en 17 millones de contener la misma secuencia de nucleótidos aleatorios (110). Por lo tanto, decenas o cientos de miles de oligonucleótidos plantilla sintéticos pueden incluirse en la reacción de PCR con superposición mínima a ninguna en las secuencias de nucleótidos aleatorios (110) entre dos oligonucleótidos plantilla sintéticos distintos.

30

35

Las secuencias de nucleótidos aleatorios (110) permiten cuantificar exactamente cada oligonucleótido plantilla sintético. Tras la amplificación de una mezcla de oligonucleótidos plantilla sintéticos, cada secuencia de nucleótidos aleatorios única observada en el resultado de la secuenciación representa una molécula individual de material inicial. Por lo tanto, el número inicial de oligonucleótidos plantilla sintéticos añadidos a la reacción de amplificación puede determinarse mediante el recuento del número de secuencias de nucleótidos aleatorios únicas. Además, el número inicial de oligonucleótidos plantilla sintéticos asociados con un código de barras particular (y por lo tanto asociados con una combinación pareada particular de una secuencia oligonucleotídica V y una secuencia oligonucleotídica J) puede determinarse mediante el recuento del número de secuencias de nucleótidos aleatorios únicas asociadas con un código de barras particular. Los ejemplos de plantillas sintéticas que comprenden segmentos de nucleótidos aleatorios pueden encontrarse, por ejemplo, en las SEQ ID NO: 3004-3159.

40

45

Sitios de enzimas de restricción

De acuerdo con ciertas modalidades descritas en la presente descripción, el oligonucleótido plantilla puede comprender además un sitio de reconocimiento de endonucleasa de restricción (RE) que se sitúa entre las secuencias V y J y no aparece en otra parte en la secuencia oligonucleotídica plantilla. Opcionalmente el sitio de reconocimiento de RE puede estar adyacente a un sitio de código de barras que identifica la secuencia de región V. El sitio de RE puede incluirse para cualquiera de una serie de propósitos, que incluyen sin limitación como una característica estructural que puede explotarse para destruir plantillas selectivamente al ponerlas en contacto con la enzima de restricción apropiada. Puede ser conveniente degradar los presentes oligonucleótidos plantilla selectivamente al ponerlos en contacto con una RE adecuada, por ejemplo, para eliminar oligonucleótidos plantilla de otras composiciones en las que pueden haberse introducido deliberadamente o accidentalmente. Alternativamente, el sitio de RE puede explotarse de manera útil en el curso de la secuenciación de los oligonucleótidos plantilla en la composición de plantillas, y/o como un marcador de secuencia posicional en una secuencia oligonucleotídica plantilla independientemente de si se escinde o no con una enzima de restricción. Un sitio de RE ilustrativo es el motivo oligonucleotídico GTCGAC, reconocido por la enzima de restricción Sal I. Un gran número de enzimas de restricción adicionales y sus secuencias de sitio de reconocimiento de RE respectivas se conocen en la técnica 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 cualquiera de una variedad de tales sitios de reconocimiento de RE puede incorporarse en modalidades particulares de los oligonucleótidos plantilla descritos en la presente descripción.

50

55

60

65

Composiciones de plantillas sintéticas de control útiles para cuantificar una representación relativa de células de la inmunidad adaptativa en una muestra biológica

5 Los oligonucleótidos plantilla sintéticos de control pueden diseñarse para cuantificar un número de moléculas iniciales en una muestra biológica. Estos oligonucleótidos plantilla sintéticos de control son similares a los oligonucleótidos plantilla sintéticos descritos anteriormente, pero no contienen una secuencia oligonucleotídica V o una secuencia oligonucleotídica J. Cuando se hace referencia a las plantillas sintéticas, frecuentemente los oligonucleótidos que contienen regiones V y J se denominan un "primer" conjunto de plantillas sintéticas mientras que las plantillas sintéticas de control frecuentemente se denominan un "segundo" conjunto de plantillas sintéticas. En lugar de esto, una composición de plantillas sintéticas de control comprende una pluralidad de oligonucleótidos plantilla de la fórmula general (II):

10 5'-U1-B1-X1-B2-N-X2-B3-U2-3' (II).

15 Los segmentos U1, B1, B2, N, B3, y U2 son iguales a como se describió anteriormente. En una modalidad, 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, o 800 nucleótidos contiguos de una secuencia de ADN. En algunas modalidades, la secuencia de ADN es de un gen de control genómico (también denominado un "gen de control interno"), o el complemento de esta. Como se usa en la presente descripción "gen de control genómico" o "gen de control interno" es cualquier gen que se encuentra en todas las células (que incluyen tanto células de la inmunidad adaptativa como células que no son células de la inmunidad adaptativa), tal como un gen constitutivo como RNasa P, PSMB2, RAB7A, UBC, VCP, REEP5, o EMC7.

20 Los oligonucleótidos plantilla sintéticos de la fórmula (I) se usan para determinar un número total de moléculas de receptor de inmunidad adaptativa iniciales (y por lo tanto de células de la inmunidad adaptativa) en una muestra biológica. Como se explica a continuación, los oligonucleótidos plantilla sintéticos de control de la fórmula (II) pueden usarse para determinar el número total de todos los genomas iniciales en una muestra biológica, la muestra biológica incluye células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa.

25 En algunas modalidades, una composición de plantillas sintéticas de control comprende una de las secuencias que se encuentran en las SEQ ID NO: 3160-3252. Las SEQ ID NO: 3167-3194 demuestran cebadores de secuenciación ilustrativos para composiciones de plantillas sintéticas de control que contienen varios segmentos génicos de control, las SEQ ID NO: 3195-3222 demuestran secuencias de cebadores ilustrativas para secuencias adaptadoras de las composiciones de plantillas sintéticas de control, y las SEQ ID NO: 3223-3236 demuestran secuencias de cebadores ilustrativas específicas para las composiciones de plantillas sintéticas de control. La Figura 1 ilustra un ejemplo de un oligonucleótido plantilla sintético de control, de acuerdo con una modalidad de la invención. El oligonucleótido plantilla sintético de control de la Figura 1 tiene la fórmula: 5'-X1-N-B1-X2-3', que difiere ligeramente de la fórmula general (II) anterior.

30 En ciertas modalidades es ventajoso para las plantillas sintéticas de control que sean de longitud similar a la de las plantillas sintéticas que contienen segmentos V y J o C de TCR y/o Ig. Además, también es ventajoso en muchas modalidades para las plantillas sintéticas (tanto las plantillas de control como las que contienen secuencias biológicas de TCR o Ig) que sean de longitud similar a la del producto de amplificación de los loci de TCR/Ig y la región de control genómico de la muestra inicial. En algunas modalidades, la longitud de las plantillas sintéticas y los amplicones correspondientes de material biológico es de entre alrededor de 100 y alrededor de 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).

35 Métodos para determinar el número de oligonucleótidos plantilla sintéticos iniciales en una muestra y un factor de amplificación

40 En la presente descripción se describen métodos para determinar un número de oligonucleótidos plantilla sintéticos añadidos a una muestra inicial para el uso en PCR. El número de oligonucleótidos plantilla sintéticos iniciales puede estimarse mediante el uso de una dilución limitante de oligonucleótidos plantilla sintéticos en un ensayo de PCR múltiple. Este número de oligonucleótidos plantilla sintéticos iniciales en un ensayo de PCR y el número de lecturas de secuenciación de salida producidas a partir del ensayo de PCR pueden usarse después para calcular una relación de amplificación.

45 Una dilución limitante se logra cuando la cantidad de ADN en una muestra se diluye hasta el punto en que solamente un subconjunto muy pequeño de oligonucleótidos plantilla sintéticos está presente en la dilución. Por ejemplo, en una mezcla de 1.000 oligonucleótidos plantilla sintéticos únicos, la dilución limitante puede incluir solamente 100 de los 1.000 oligonucleótidos plantilla sintéticos únicos. La mayor parte de las plantillas sintéticas únicas estaría ausente en la dilución limitante. Por ejemplo, la dilución limitante puede incluir solamente 100 tipos únicos de oligonucleótido plantilla sintético y solamente 1 copia de cada oligonucleótido plantilla sintético único. Por lo tanto, una porción de los oligonucleótidos plantilla sintéticos únicos se añade como una copia única o solamente un número pequeño de copias, y el resto de los oligonucleótidos plantilla sintéticos en la mezcla se añaden a cero copias (es decir, ausentes). En ciertos casos, la dilución limitante de los oligonucleótidos plantilla sintéticos únicos incluye una molécula de cada oligonucleótido plantilla sintético único, detectable. En otros casos, la dilución limitante puede incluir dos moléculas de uno o más de los oligonucleótidos plantilla sintéticos únicos, detectables. Por lo tanto, la dilución limitante incluye una concentración muy baja de

oligonucleótidos plantilla sintéticos únicos.

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

La PCR simple permite la amplificación de cada oligonucleótido plantilla sintético único mediante el uso de un par de cebadores de PCR para todas las plantillas sintéticas en la mezcla completa de oligonucleótidos plantilla sintéticos. La PCR simple puede realizarse a los oligonucleótidos plantilla sintéticos mediante el uso de cebadores universales que incluyen las secuencias adaptadoras y se hibridan a las secuencias de cebador universal (UA (102) y UB (112), como se muestra en la Figura 1B). Después, la genoteca resultante de amplicones de oligonucleótidos plantilla sintéticos puede secuenciarse individualmente mediante el uso de las secuencias adaptadoras en cada amplicón en un secuenciador, tal como un secuenciador de Illumina. Este proceso permite la medición directa de la frecuencia de cada oligonucleótido plantilla sintético en la mezcla compleja.

En ciertos casos, se usa una simulación *in silico* para analizar la relación entre el número de amplicones de oligonucleótidos plantilla sintéticos únicos secuenciados a partir de la dilución limitante usada en una reacción de PCR múltiple y el número inicial total estimado de oligonucleótidos plantilla sintéticos añadidos a dicha reacción de PCR múltiple. La Figura 2 proporciona una simulación *in silico* de la relación entre el número de tipos únicos de oligonucleótidos plantilla sintéticos observados (por ejemplo, secuenciados a partir de la muestra) y el número de moléculas plantilla sintéticas muestreadas (por ejemplo, número de oligonucleótidos plantilla sintéticos en la muestra inicial). Por ejemplo, si se secuencian y observan 400 tipos únicos de oligonucleótidos plantilla sintéticos a partir de la muestra, puede determinarse que la muestra inicial incluía aproximadamente 500 moléculas de oligonucleótidos plantilla sintéticos. En consecuencia, el número total de oligonucleótido plantilla sintético inicial puede determinarse a partir del número de oligonucleótidos plantilla sintéticos únicos observados.

Una porción de esta mezcla de oligonucleótidos plantilla sintéticos puede añadirse después a una reacción de PCR múltiple que comprende moléculas biológicas de ácidos nucleicos de TCR o IG reordenados obtenidas a partir de linfocitos en una muestra dada. El número determinado de oligonucleótidos plantilla sintéticos añadidos ("de enriquecimiento") y la relación de amplificación calculada pueden usarse para determinar un número total de linfocitos en la muestra.

Como se describe en detalle en la presente descripción, después de la caracterización de la mezcla de oligonucleótidos plantilla sintéticos, puede añadirse una dilución limitante de esta mezcla a una muestra biológica para determinar el número de células B o T presentes en dicha muestra biológica. Un factor de amplificación se determina en base al número de oligonucleótidos plantilla sintéticos en una muestra inicial de oligonucleótidos plantilla sintéticos que se ha añadido a una muestra biológica. El factor de amplificación se calcula mediante la comparación del número de lecturas de secuenciación totales de los oligonucleótidos plantilla sintéticos obtenidos a partir de una muestra con el número total de oligonucleótidos plantilla sintéticos iniciales en la muestra, y puede usarse para determinar el número de linfocitos totales (células T o células B) en una muestra biológica. Este factor de amplificación puede asumirse para aplicarlo a plantillas biológicas (por ejemplo, moléculas de ácidos nucleicos de TCR o IG reordenados) que se han amplificado con los mismos cebadores específicos de segmentos V y segmentos J usados para amplificar moléculas de oligonucleótidos plantilla sintéticos.

En un caso, el factor de amplificación (relación) del número de lecturas de secuenciación de los amplicones de oligonucleótidos plantilla sintéticos respecto al número de moléculas de oligonucleótidos plantilla sintéticos totales iniciales se compara con el número de lecturas de secuenciación totales de amplicones de moléculas biológicas para calcular el número inicial de moléculas biológicas iniciales. El número de moléculas de oligonucleótidos plantilla sintéticos al inicio del ensayo de PCR puede usarse después en los cálculos de la representación relativa de células de la inmunidad adaptativa en la muestra, como se describe en detalle a continuación.

Métodos para determinar representación absoluta de células de la inmunidad adaptativa en una muestra

Se proporcionan métodos para determinar la representación absoluta de secuencias reordenadas que codifican el receptor de inmunidad adaptativa en una muestra

Los métodos descritos en la presente descripción incluyen extraer moléculas biológicas de ácidos nucleicos (por ejemplo, moléculas de ADN de TCR o IG reordenadas) a partir de una muestra biológica que comprende células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa. Las moléculas biológicas de ácidos nucleicos en la muestra se "enriquecen" con una cantidad conocida de oligonucleótidos plantilla sintéticos (por ejemplo, como se describió anteriormente y se determina por dilución limitante). Los oligonucleótidos plantilla sintéticos comprenden las mismas secuencias oligonucleotídicas de segmentos V y segmentos J que las plantillas de moléculas biológicas de ácidos nucleicos.

En ciertos casos, el método para cuantificar el número absoluto de moléculas de ADN reordenadas que codifican una pluralidad de receptores de inmunidad adaptativa en una muestra biológica de un sujeto, comprende las siguientes etapas:

I. Amplificar, en un ensayo de PCR múltiple, un subconjunto de moléculas de oligonucleótidos plantilla sintéticos obtenidas a partir de una mezcla de oligonucleótidos plantilla sintéticos, el subconjunto de moléculas de oligonucleótidos plantilla sintéticos se diluye de manera que solo esté presente una copia única o un número pequeño de copias de una porción de oligonucleótidos plantilla sintéticos únicos. Los oligonucleótidos plantilla sintéticos amplificados se secuencian, y se determina el número de oligonucleótidos plantilla sintéticos únicos en base a las secuencias de código de barras únicas. El número de lecturas de secuenciación totales de los oligonucleótidos plantilla sintéticos también se determina a partir de los resultados de la secuenciación. A continuación, se hace referencia a los resultados de una simulación *in silico* en base a la caracterización previa de la mezcla de oligonucleótidos plantilla sintéticos (por PCR simple) para determinar a partir del número de secuencias oligonucleotídicas plantilla sintéticas únicas el número inicial total de moléculas de oligonucleótidos plantilla sintéticos (por ejemplo, en base a la relación que se muestra en la Figura 2). Se determina un factor de amplificación a partir de la relación del resultado total de lecturas de secuenciación de la muestra y el número total estimado de oligonucleótidos plantilla sintéticos iniciales. Este factor de amplificación puede usarse para estimar el número total de moléculas biológicas reordenadas, y por lo tanto, el número total de células linfoides, en una muestra dada. Esto puede realizarse mediante la adición ("enriquecimiento") de una pequeña porción de la mezcla de oligonucleótidos plantilla sintéticos diluidos a la PCR múltiple.

II. Amplificar las moléculas de ácidos nucleicos obtenidas a partir de una muestra dada, en una PCR múltiple mediante el uso de un conjunto de cebadores oligonucleotídicos de amplificación que comprende cebadores de segmentos V y segmentos J como se describe en la presente descripción capaces de amplificar sustancialmente todas las combinaciones de segmentos V y segmentos J de receptores de inmunidad adaptativa reordenados, la muestra comprende i) moléculas de ácidos nucleicos reordenados de receptores de inmunidad adaptativa TCR o Ig, que comprenden cada una una región V y una región J, y ii) una porción de oligonucleótidos plantilla sintéticos "de enriquecimiento" como se describió anteriormente que tienen una cantidad inicial conocida, para generar de este modo amplicones que comprenden una pluralidad de amplicones de receptores de inmunidad adaptativa TCR o Ig reordenados de manera única y una pluralidad de amplicones de plantillas sintéticas.

III. Secuenciar cuantitativamente la pluralidad de amplicones de receptores de inmunidad adaptativa TCR o Ig reordenados de manera única y una pluralidad de amplicones de plantillas sintéticas generados en (I) para determinar el número total de amplicones de receptores de inmunidad adaptativa TCR o Ig reordenados observados por secuenciación (denominado en la presente descripción A_i) y el número total de amplicones de plantilla sintética observados por secuenciación (denominado en la presente descripción A_{ii}). La información de la secuenciación incluye el número de productos de secuenciación de salida de la pluralidad de amplicones de receptores de inmunidad adaptativa TCR o Ig reordenados (A_i) y el número de productos de secuenciación de salida de los amplicones de plantillas sintéticas (A_{ii}).

IV. Determinar una representación absoluta de células de la inmunidad adaptativa en la muestra en base a la información de la secuenciación cuantitativa determinada a partir de la etapa II.

Para determinar la representación absoluta de células de la inmunidad adaptativa, primero se calcula un factor de amplificación. El factor de amplificación es la relación del número de productos de secuenciación de salida de los amplicones de plantillas sintéticas (A_{ii}) con el número conocido de oligonucleótidos plantilla sintéticos iniciales (denominado en la presente descripción A_{iii}). El número de oligonucleótidos plantilla sintéticos iniciales se determina en base a la simulación *in silico* realizada en (I) para determinar la relación entre el número de amplicones de oligonucleótidos plantilla sintéticos únicos y el número inicial total de oligonucleótidos plantilla sintéticos. Se asume que el factor de amplificación de un conjunto de cebadores particular para un oligonucleótido plantilla sintético es el mismo factor de amplificación para la plantilla biológica.

Factor de amplificación = A_{ii} / A_{iii} = número de productos de secuenciación de salida de los amplicones de plantillas sintéticas / número conocido de oligonucleótidos plantilla sintéticos iniciales.

En el cálculo de este factor de amplificación, se asume que la relación del número de lecturas de secuenciación de salida por molécula inicial es la misma para una molécula de oligonucleótido plantilla sintético y una molécula biológica de ácido nucleico reordenado de receptor de inmunidad adaptativa TCR o Ig.

Después de calcular el factor de amplificación, puede determinarse el número total de moléculas de receptor de inmunidad adaptativa TCR o Ig reordenado en la muestra, y en consecuencia, el número total de células linfocitarias.

En un caso, el número de moléculas biológicas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa se determina mediante lo siguiente:

Número de moléculas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa = $A_i / (A_{ii} / A_{iii})$ = (Número de productos de secuenciación de salida determinado a partir de la pluralidad de amplicones de receptores de inmunidad adaptativa TCR o Ig reordenados) / (Factor de amplificación)

El número total de moléculas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa es igual al número total de células de la inmunidad adaptativa (por ejemplo, células T o células B) en la muestra. En consecuencia,

puede determinarse el número total de células de la inmunidad adaptativa en la muestra.

Determinación de la representación relativa de células de la inmunidad adaptativa en una mezcla compleja de células

5 Los métodos descritos en la presente descripción incluyen determinar una representación relativa de células de la inmunidad adaptativa en una mezcla compleja de células que incluye células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa. En algunas modalidades, el número total de células de la inmunidad adaptativa se determina como se describió en la sección anterior y después se usa para calcular la representación relativa de células de la inmunidad adaptativa en la muestra total de células.

10 El número total de moléculas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa (o número total de células de la inmunidad adaptativa) se usa para determinar la representación relativa de células de la inmunidad adaptativa en la mezcla compleja. En una modalidad, la masa total de ADN en la muestra se usa para cuantificar el número total de células de la inmunidad adaptativa y células que no son de la inmunidad adaptativa en la mezcla compleja. Asumiendo que cada célula tiene aproximadamente 6,5 picogramos de ADN y dada una masa total conocida de ADN inicial para el ensayo de PCR, el número total de células de la inmunidad adaptativa y células que no son de la inmunidad adaptativa totales en la muestra se cuantifica mediante la división de la masa total conocida de ADN inicial por 6,5 picogramos. Esto da como resultado la representación relativa de células de la inmunidad adaptativa en la mezcla compleja de células que incluye células de la inmunidad adaptativa y células que no son células de la inmunidad adaptativa.

15 En otras palabras, la representación relativa de células de la inmunidad adaptativa = número total de moléculas de ácidos nucleicos reordenados que codifican receptores de inmunidad adaptativa / (masa total de ADN que representa células de la inmunidad adaptativa y células que no son de la inmunidad adaptativa).

20 Varios otros cálculos como conocen los expertos en la técnica pueden usarse para determinar la representación relativa de células de la inmunidad adaptativa en una mezcla compleja.

25 Métodos para diagnosticar, prevenir, o tratar la enfermedad en pacientes en base a la determinación de la representación relativa de células de la inmunidad adaptativa en la muestra de un paciente

30 Se proporcionan métodos para determinar un curso de tratamiento para un paciente que lo necesita, que comprende cuantificar la representación relativa de linfocitos infiltrantes de tumor o linfocitos que infiltran un tejido somático que es la diana de una reacción autoinmunitaria, mediante el uso de los métodos descritos en la presente descripción. En este sentido, el paciente que lo necesita puede ser un paciente con cáncer o un paciente que tiene una enfermedad autoinmunitaria. En ciertos casos, un paciente puede tener un cáncer que incluye, pero sin limitarse a, un cáncer colorrectal, hepatocelular, de vesícula biliar, pancreático, esofágico, de pulmón, de mama, de próstata, de piel (*por ejemplo*, melanoma), de cabeza y cuello, carcinoma de células renales, de ovario, endometrial, cervical, de vejiga y urotelial. En ciertos otros casos, un paciente puede tener un trasplante de 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/injerto de piel, un trasplante de intestino, y un trasplante de timo.

35 Las enfermedades autoinmunitarias incluyen, pero sin limitarse a, artritis (que incluye artritis reumatoide, artritis reactiva), lupus eritematoso sistémico (SLE), psoriasis, enfermedad inflamatoria intestinal (IBD) (que incluye colitis ulcerosa y enfermedad de Crohn), encefalomiелitis, uveítis, miastenia gravis, esclerosis múltiple, diabetes dependiente de insulina, enfermedad de Addison, enfermedad celíaca, síndrome de fatiga crónica, hepatitis autoinmunitaria, alopecia autoinmunitaria, espondilitis anquilosante, fibromialgia, pénfigo vulgar, síndrome de Sjogren, enfermedad de Kawasaki, hipertiroidismo/enfermedad de Graves, hipotiroidismo/enfermedad de Hashimoto, endometriosis, escleroderma, anemia perniciosa, síndrome de Goodpasture, síndrome de Guillain-Barré, enfermedad de Wegener, glomerulonefritis, anemia aplásica (que incluye pacientes de anemia aplásica con transfusiones múltiples), hemoglobinuria nocturna paroxística, púrpura trombocitopénica idiopática, anemia hemolítica autoinmunitaria, síndrome de Evan, síndrome por inhibidor de Factor VIII, vasculitis sistémica, dermatomiositis, polimiositis y fiebre reumática, síndrome linfoproliferativo autoinmunitario (ALPS), penfigoide ampolloso autoinmunitario, enfermedad de Parkinson, sarcoidosis, vitiligo, cirrosis biliar primaria, y miocarditis autoinmunitaria.

40 Los métodos descritos en la presente descripción pueden usarse para enumerar la presencia relativa de linfocitos infiltrantes de tumor, o de linfocitos que infiltran un tejido somático que es la diana de una reacción autoinmunitaria, en base a la cuantificación de la representación relativa del ADN de tales células de la inmunidad adaptativa en el ADN extraído de una muestra biológica, que comprende una mezcla de tipos celulares, que se ha obtenido a partir de tal tumor o tejido. Tales métodos son útiles para determinar el pronóstico y diagnóstico del cáncer o la enfermedad autoinmunitaria, para evaluar los efectos de un tratamiento terapéutico (*por ejemplo*, evaluar la eficacia de fármacos y/o las relaciones de respuesta a dosis), y para identificar los cursos terapéuticos para el tratamiento del cáncer, para el tratamiento de enfermedades autoinmunitarias, o para el tratamiento del rechazo al trasplante, y puede encontrar otros usos relacionados.

45 Para evaluar un tratamiento terapéutico, por ejemplo, ciertos casos contemplan un método en el que se evalúa un efecto

del tratamiento terapéutico sobre la representación relativa de células de la inmunidad adaptativa en al menos un tejido en un sujeto al que se ha administrado el tratamiento. A modo de ilustración y no de limitación, de acuerdo con ciertas modalidades de este tipo un tratamiento que altera (*por ejemplo*, aumenta o disminuye de una manera estadísticamente significativa) la representación relativa de células de la inmunidad adaptativa en un tejido o tejidos puede conferir ciertos beneficios al sujeto. Por ejemplo, ciertas inmunoterapias para el cáncer se diseñan para mejorar el número de linfocitos infiltrantes de tumor (TIL). Se ha demostrado que la presencia de TIL CD3+ en tumores de ovario se correlaciona fuertemente con el resultado del paciente (*ver, por ejemplo*, Hwang y otros, 2011 *Gynecol. Oncol.*, 124(2): 192). Datos adicionales aclararon que además de la presencia de TIL, las características de las poblaciones de TIL también eran significativas: Los TIL CD8+ y los TIL clonales se asociaron con supervivencia libre de enfermedad (DFS) más larga, y las células T reguladoras infiltrantes se asociaron con DFS más corta *ver*, Stumpf y otros, 2009 *Br. J. Cancer* 101:1513-21). Estos estudios indicaron que los TIL pueden ser un factor de pronóstico independiente (*ver*, Clarke y otros, 2009 *Mod. Pathol.* 22:393-402). Por lo tanto, la cuantificación de la representación relativa del ADN de células de la inmunidad adaptativa como se describe en la presente descripción, para propósitos de detección de posibles aumentos de TIL en muestras de tejido tumoral obtenidas en uno o una pluralidad de puntos de tiempo antes del tratamiento, durante el curso de 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.

Como otro ejemplo, ciertas inmunoterapias dirigidas a enfermedades autoinmunitarias se diseñan para reducir el número de linfocitos infiltrantes de tejido en uno o más tejidos afectados tales como tejidos u órganos que pueden ser dianas del ataque autoinmunitario clínicamente inapropiado, de manera que la cuantificación de la representación relativa del ADN de células de la inmunidad adaptativa como se describe en la presente descripción, para propósitos de detección de posibles disminuciones de células de la inmunidad adaptativa en muestras de tejido obtenidas en uno o una pluralidad de puntos de tiempo antes del tratamiento, durante el curso de 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.

Como un ejemplo adicional, ciertas inmunoterapias dirigidas al rechazo al trasplante se diseñan para reducir el número de linfocitos infiltrantes de tejido en órganos trasplantados, de manera que la cuantificación de la representación relativa del ADN de células de la inmunidad adaptativa como se describe en la presente descripción, para propósitos de detección de posibles disminuciones de células de la inmunidad adaptativa en muestras de tejido de órganos trasplantados obtenidas en uno o una pluralidad de puntos de tiempo antes del tratamiento, durante el curso de 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.

En estos casos y los relacionados, los métodos descritos en la presente descripción para cuantificar la representación relativa del ADN de células de la inmunidad adaptativa pueden llevarse a la práctica mediante el uso de muestras biológicas de prueba obtenidas a partir 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 pueden obtenerse a partir de tejidos iguales o diferentes, lo cual puede variar en función de la afección 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 a partir del sujeto antes y después del tratamiento pueden ser del mismo tejido, mientras que en el caso de un tumor que se extirpa quirúrgicamente de manera parcial, o que aparece en múltiples sitios en el sujeto, las muestras biológicas de prueba pueden obtenerse a partir de tejidos diferentes o de sitios del tejido diferentes antes y después de administrar el tratamiento terapéutico.

En la presente descripción también se contemplan casos en los que cualquiera de los métodos descritos en la presente descripción puede comprender además la determinación de la diversidad estructural relativa de receptores de inmunidad adaptativa (*por ejemplo*, la diversidad de secuencias entre los productos de genes de TCR y/o inmunoglobulina reordenados de manera productiva) en el componente de células de la inmunidad adaptativa de la mezcla de células que está presente en la muestra biológica de prueba. En ciertos casos de este tipo, las presentes metodologías de qPCR que usan los conjuntos de cebadores oligonucleotídicos específicos que codifican receptores de inmunidad adaptativa reordenados descritos en la presente descripción permite la identificación fácil de las combinaciones de cebadores particulares que generan la producción de moléculas de ADN reordenadas amplificadas. En consecuencia, por ejemplo, estos casos permiten la determinación del grado relativo de clonalidad de una población de células de la inmunidad adaptativa que está presente como parte de una población de células mixtas en una muestra biológica de prueba, lo que puede tener valor pronóstico.

Por ejemplo, en una muestra de tumor sólido en la que se detectan TIL mediante la cuantificación de la representación relativa del ADN de células de la inmunidad adaptativa en el ADN extraído de la muestra como se describe en la presente descripción, los presentes métodos contemplan la determinación de si solo una o algunas (*por ejemplo*, no más de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10) combinaciones de un cebador oligonucleotídico de segmento V particular y un cebador oligonucleotídico de segmento J particular son predominantemente responsables (*por ejemplo*, generan al menos 80, 85, 90, 95, 97 o 99 por ciento de los productos de amplificación) de la producción por PCR de moléculas de ADN reordenadas amplificadas de células de la inmunidad adaptativa. Una observación de este tipo de uno o algunos productos de amplificación predominantes que codifican genes de receptor de inmunidad adaptativa indicaría, de acuerdo con la teoría no limitante, un grado bajo de heterogeneidad de TIL. Por el contrario, la determinación de un grado alto de

heterogeneidad en la diversidad estructural de receptores de inmunidad adaptativa mediante la caracterización del ADN de TIL indicaría que no está presente un clon de TIL predominante.

5 En consecuencia, en la presente descripción se describen métodos para medir el número de células de la inmunidad adaptativa (por ejemplo células T) en una mezcla compleja de células. Los presentes métodos tienen utilidad particular en la cuantificación de linfocitos infiltrantes de tumor o linfocitos infiltrantes de tejido somático que es la diana de una respuesta autoinmunitaria. Los métodos existentes para la cuantificación de células T y B se basan en la separación física de tales células de la mezcla. Sin embargo, en muchos casos, las células T y B no pueden separarse de la muestra inicial, tal como muestras de tejido fijadas con formalina o congeladas. Además, los métodos anteriores para la cuantificación de células de la inmunidad adaptativa (*por ejemplo*, inmunocitometría de flujo, clasificación de células activadas por fluorescencia (FACS), inmunohistoquímica (IHC)) se basan en la expresión de proteínas específicas de células T o células B, tales como receptores de superficie celular. Dado que las células inmunitarias expresan cantidades variables de estos receptores específicos de linaje, la cuantificación del número de células a partir de una medida altamente variable como esta requiere una estandarización costosa, equipos especializados y personal altamente calificado. Los métodos descritos en la presente descripción, a diferencia de esto, no dependen de una plataforma y pueden realizarse en cualquier instrumento de PCR e instrumento de secuenciación de alta productividad, y los reactivos pueden sintetizarse y proporcionarse en forma de kit. Los métodos descritos en la presente descripción son además altamente sensibles y pueden aplicarse en configuraciones de alta productividad que anteriormente no podían lograrse. Como se describe en la presente descripción, la cuantificación de células de la inmunidad adaptativa puede lograrse mediante una preparación simple del ADN de una mezcla compleja de células, conjuntamente con la cuantificación de la proporción relativa de células de la inmunidad adaptativa presentes por amplificación de los genes reordenados que codifican CDR3 de células de la inmunidad adaptativa. _

25 En la presente descripción se describen métodos para comparar cantidades de ADN de células de la inmunidad adaptativa con el ADN de células totales (*por ejemplo*, de células de la inmunidad adaptativa más células que no son de la inmunidad adaptativa en la mezcla de células). Opcionalmente los métodos incluyen además 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 anteriormente han demostrado no experimentar una alteración estadísticamente significativa del estado fisiológico, tal como un control de inyección simulada, de solución salina, DMSO u otro vehículo o tampón, enantiómeros inactivos, péptidos o nucleótidos mezclados, *etc.*, y/o antes, durante o después de la administración de controles positivos que anteriormente han demostrado provocar una alteración estadísticamente significativa del estado fisiológico, tal como un compuesto terapéutico aprobado por la FDA.

35 La práctica de ciertas modalidades de la presente invención empleará, a menos que se indique específicamente lo contrario, métodos convencionales en técnicas de microbiología, biología molecular, bioquímica, genética molecular, biología celular, virología e inmunología que están dentro de la experiencia de la técnica, y a varias de las cuales se hace referencia a continuación con propósitos de ilustración. Tales técnicas se explican completamente en la bibliografía. Ver, *por ejemplo*, Sambrook, y otros, *Molecular Cloning: A Laboratory Manual* (3^{ra} Edición, 2001); Sambrook, y otros, *Molecular Cloning: A Laboratory Manual* (2^{da} Edición, 1989); Maniatis y otros, *Molecular Cloning: A Laboratory Manual* (1982); Ausubel y otros, *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 & II (IRL Press, Oxford Univ. Press Estados Unidos, 1985); *Current Protocols in Immunology* (Editado por: John E. Coligan, Ada M. Krusbeek, 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, Reino Unido; 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 & S. Higgins, Eds., 1985); *Transcription and Translation* (B. Hames & S. Higgins, Eds., 1984); *Animal Cell Culture* (R. Freshney, Ed., 1986); Perbal, *A Practical Guide to Molecular Cloning* (1984); *Next-Generation Genome Sequencing* (Janitz, 2008 Wiley-VCH); *PCR Protocols (Methods in Molecular Biology)* (Park, Ed., 3^{ra} Edición, 2010 Humana Press); *Immobilized Cells and Enzymes* (IRL Press, 1986); el tratado, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller y M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Harlow y Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998); *Immunochemical Methods In Cell and Molecular Biology* (Mayer y Walker, eds., Academic Press, Londres, 1987); *Handbook Of Experimental Immunology*, Volúmenes I-IV (D. M. Weir y CC Blackwell, eds., 1986); Riott, *Essential Immunology*, 6ta Edición, (Blackwell Scientific Publications, Oxford, 1988); *Embryonic Stem Cells: Methods and Protocols* (Methods in Molecular Biology) (Kurstad Turksen, Ed., 2002); *Embryonic Stem Cell Protocols: Volumen I: Isolation and Characterization* (Methods in Molecular Biology) (Kurstad Turksen, Ed., 2006); *Embryonic Stem Cell Protocols: Volumen 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 en relación con, y los procedimientos y técnicas de laboratorio de, biología molecular, química analítica, química de síntesis orgánica, y química de medicamentos y farmacéutica descritos en la presente descripción son aquellos bien conocidos y usados comúnmente en la técnica. Pueden usarse técnicas estándar de tecnología recombinante, biología molecular, microbiología, síntesis química, análisis químico, preparación, formulación, y suministro farmacéutico, y tratamiento de los pacientes.

A menos que el contexto lo requiera de cualquier otra manera, a lo largo de la presente descripción y las reivindicaciones, la palabra "comprender" y variaciones de esta, tales como, "comprende" y "que comprende" deben interpretarse en un sentido abierto e inclusivo, es decir, como "que incluye, pero sin limitarse a". Por "que consiste en" se entiende que incluye, y típicamente se limita a, lo que sigue a continuación de la frase "que consiste en". Por "que consiste esencialmente en" se entiende que incluye cualquiera de los elementos enumerados después de la frase, y se limita a otros elementos que no interfieren con o contribuyen a la actividad o acción especificada en la descripción para los elementos enumerados. Por lo tanto, la frase "que consiste esencialmente en" indica que los elementos enumerados se requieren o son obligatorios, pero que ningún otro elemento se requiere y puede o no estar presente en dependencia de si afecta o no la actividad o acción de los elementos enumerados.

En esta descripción y las reivindicaciones adjuntas, las formas singulares "un", "una" y "el/la" incluyen las referencias en plural a menos que el contenido lo indique claramente de cualquier otra manera. Como se usa en la presente descripción, en modalidades particulares, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 5 %, 6 %, 7 %, 8 % o 9 %. En otras modalidades, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 10 %, 11 %, 12 %, 13 % o 14 %. Aún en otras modalidades, los términos "alrededor de" o "aproximadamente" cuando preceden a un valor numérico indican el valor más o menos un intervalo de 15 %, 16 %, 17 %, 18 %, 19 % o 20 %.

La referencia a lo largo de esta descripción a "una modalidad" o "la modalidad" o "un aspecto" significa que un elemento, estructura o característica particular descrita en relación con la modalidad está incluida en al menos una modalidad de la presente invención. Por lo tanto, la aparición de las frases "en una modalidad" o "en la modalidad" en varios lugares a lo largo de esta descripción no se refiere necesariamente en todos los casos a la misma modalidad. Además, los elementos, estructuras, materiales o características particulares pueden combinarse de cualquier manera adecuada en una o más modalidades.

En algunas modalidades, las moléculas plantilla sintéticas pueden incluir una secuencia adaptadora universal directa, al menos una secuencia de código de barras única, una secuencia complementaria a un segmento génico V, una secuencia de marcador interno de plantilla, una secuencia oligonucleotídica aleatoria de longitud N, y una secuencia adaptadora universal inversa. Las moléculas plantilla sintéticas con la secuencia oligonucleotídica aleatoria se nombran "vBloques." En ciertas modalidades, las moléculas plantilla sintéticas no incluyen la secuencia oligonucleotídica aleatoria de longitud N (se llaman "gBloques"). En otras modalidades, la secuencia de marcador interno de plantilla se usa para distinguir las moléculas plantilla sintéticas de las moléculas biológicas.

En una modalidad, la molécula plantilla sintética puede tener un intervalo de longitud de 100 a unos miles de pares de bases de longitud. En ciertas modalidades, la molécula plantilla sintética es de 100-2.500 pb de longitud. En un ejemplo, la molécula plantilla sintética puede sintetizarse como oligonucleótidos de 495 pares de bases con la siguiente estructura (5' a 3'): (1) una secuencia adaptadora universal (UA), (2) un código de barras de 16 pares de bases que identifica segmentos V y J, (3) un gen V (de alrededor de 300 pares de bases), (4) un marcador interno de plantilla sintética (IM) de 9 pares de bases, (5) una repetición del código de barras de 16 pares de bases, (6) una cadena de 12 oligonucleótidos aleatorios (N12), (7) un gen J (de alrededor de 100 pares de bases), (8) una repetición del código de barras de 16 pares de bases, y (9) una secuencia adaptadora universal (UB). Las secuencias de código de barras pueden variar en longitud de 2-100 pares de bases. La secuencia de nucleótidos aleatorios (N) puede variar en longitud de 2-100 pares de bases, por ejemplo. En una modalidad, la secuencia oligonucleotídica aleatoria es de 8 pb de longitud (N8). Los ejemplos de moléculas plantilla sintéticas pueden encontrarse en las SEQ ID NO:3004-3159.

La Figura 1 ilustra una molécula plantilla sintética ilustrativa, de acuerdo con una modalidad de la invención.

Los ejemplos de moléculas plantilla sintéticas pueden encontrarse en las SEQ ID NO:3004-3159.

En una modalidad, se usan adaptadores universales para caracterizar moléculas sin usar PCR múltiple. Los adaptadores universales pueden estar presentes en todas las plantillas sintéticas. Cuando a los cebadores se les anexan los adaptadores universales y de Illumina y se secuencian con adaptadores de Illumina (ver la Figura 1A, anterior), estas plantillas se comportan de la misma manera que las plantillas sintéticas típicas. Cuando se amplifican mediante el uso de cebadores de PCR múltiple VF y JR y se secuencian con cebadores JR (ver la Figura 1B, anterior), estas moléculas producen una lectura de secuenciación con la siguiente estructura, por ejemplo, (5' a 3'): (1) secuencia génica J (de alrededor de 15 pares de bases), (2) una cadena de 12 nucleótidos aleatorios (N12), (3) un código de barras (BC) de V-J de 16 pares de bases, (4) un marcador interno de plantilla sintética (IM) de 9 pares de bases, y (5) un gen V (de alrededor de 15 pares de bases). En una modalidad, la secuencia oligonucleotídica aleatoria es de 8 pb de longitud (N8).

El propósito principal de incluir N bases de secuencia aleatoria es garantizar que cada molécula que se use como material inicial para la PCR múltiple tenga una cadena de nucleótidos aleatorios esencialmente única (es decir, la gran mayoría de las moléculas usadas en la PCR contendrá una secuencia oligonucleotídica aleatoria que es distinta a cualquier otra molécula usada en la PCR). Por ejemplo, una secuencia oligonucleotídica aleatoria que es de 12 pares de bases de longitud garantiza que dos moléculas cualesquiera tengan una probabilidad de alrededor de 1 en 17 millones de contener la misma región N12. Esto significa que decenas o cientos de miles de estas moléculas sintéticas pueden incluirse en la reacción de PCR con colisiones mínimas entre regiones N12. En una modalidad preferida, la secuencia oligonucleotídica aleatoria es de 8 pb de longitud.

La garantía de que la gran mayoría de las combinaciones N-V-J únicas con oligonucleótidos aleatorios representen moléculas individuales de material inicial para PCR permite el cálculo exacto de lo siguiente:

(1) El número promedio de lecturas de secuenciación obtenidas a partir de cada plantilla que contiene un segmento V y J dado.

(2) Mediante el uso de lo anterior, se puede calcular el sesgo de amplificación (en cada reacción de PCR individual) asociado con cada segmento V y J y su cebador correspondiente. Una vez calculado el sesgo de amplificación, puede determinarse un factor de normalización para corregir el sesgo de amplificación exhibido por cada reacción de amplificación.

(3)

Alternativamente, asumiendo un volumen suficientemente grande de plantilla sintética inicial, es posible generar una distribución de lecturas de secuenciación por molécula de plantilla sintética inicial para cada combinación V. J única por separado, lo que elimina la necesidad de estimar los sesgos de amplificación de V/J globales y en lugar de esto permite la comparación directa de cada secuencia biológica con su plantilla sintética asociada en base al uso de V/J y por consiguiente la estimación del número de moléculas iniciales para una plantilla inicial de V/J específica.

En una modalidad, se producen 150 tipos diferentes de plantillas sintéticas, que cubren diferentes combinaciones de genes V y J. Cada combinación de V y J específica se indica mediante una secuencia de código de barras específica. Para determinar el número preciso de moléculas de cada tipo de combinación en una muestra, pueden usarse las secuencias de nucleótidos aleatorios. Si se determinan 100 lecturas de secuenciación para un código de barras de 16 pares de bases dado (por ejemplo, para la combinación V3-2 y J1-7), no puede establecerse inmediatamente cuántas moléculas con el código de barras particular se usaron inicialmente como material inicial para PCR. Sin embargo, si se cuenta la totalidad de los segmentos de nucleótidos aleatorios asociados con ese código de barras, y hay 5 segmentos de nucleótidos aleatorios únicos, puede determinarse que las 100 lecturas de secuenciación corresponden a 5 moléculas iniciales para PCR. Esta relación de 20:1 es comparable con otras combinaciones de genes V/J para determinar el sesgo de cebador que estaba presente en la reacción de PCR, y puede usarse para contar las moléculas biológicas iniciales para PCR, asumiendo una molécula de material biológico inicial por cada veinte lecturas de secuenciación biológicas.

En una modalidad, la región de ADN aleatorizada puede situarse en cualquier parte en el amplicón previsto (es decir, en cualquier parte incluida en una región cuya amplificación se espera en una reacción de PCR). En otra modalidad, el marcador interno de plantilla sintética puede situarse en cualquier parte en el amplicón previsto, o estar ausente del amplicón. En otra modalidad, las secuencias de ADN específicas para secuencias génicas reordenadas de receptor de inmunidad adaptativa pueden reemplazarse con otros cebadores específicos de secuencia de ADN, lo que permite que este método sea útil para corregir el sesgo de amplificación y calcular la cuantificación absoluta de plantillas en cualquier configuración en la que se vayan a realizar la PCR múltiple y la secuenciación del ADN. En otra modalidad, el segmento aleatorizado de ADN puede contener cualquier cadena suficientemente grande de N nucleótidos aleatorios.

En una modalidad, no hay necesidad de un cálculo para cada combinación V/J independientemente. En lugar de esto, puede usarse una regresión lineal para todos los códigos de barras mediante el uso de cada gen V y todos los códigos de barras para cada gen J, lo que permite los cálculos necesarios sin tener que preparar una molécula sintética para cada combinación V/J posible. En lugar de esto, todo lo que se necesita es tener moléculas suficientes con cada V o J dado para medir el sesgo de amplificación.

Métodos para calcular y corregir el sesgo de amplificación con la utilización de plantillas sintéticas

Los presentes métodos superan las inexactitudes que pueden surgir en los métodos actuales que cuantifican la diversidad de los genes de TCR y BCR mediante la secuenciación de los productos de amplificación múltiple de ácidos nucleicos. Para dar cabida a la gran diversidad de secuencias plantilla de los genes de TCR y BCR que pueden estar presentes en una muestra biológica, los conjuntos de cebadores oligonucleotídicos usados en las reacciones de amplificación múltiple típicamente comprenden una amplia variedad de longitudes de secuencia y composiciones de nucleótidos (*por ejemplo*, contenido de GC). Por consiguiente, en un conjunto dado de condiciones de reacción de amplificación, las eficiencias a las que se aparean los diferentes cebadores y soportan la amplificación de sus secuencias plantilla afines pueden diferir notablemente, lo que da como resultado la utilización no uniforme de los diferentes cebadores, lo que conduce a sesgos de artefactos en la representación cuantitativa relativa de distintos productos de amplificación.

Para superar el problema de tal utilización sesgada de subpoblaciones de cebadores de amplificación, la presente descripción proporciona una composición de plantillas y un método para estandarizar las eficiencias de amplificación de los miembros de un conjunto de cebadores oligonucleotídicos, donde el conjunto de cebadores es capaz de amplificar el ADN reordenado que codifica una pluralidad de receptores de inmunidad adaptativa (TCR o Ig) en una muestra biológica que comprende ADN de células linfoides. El conjunto de cebadores es capaz de amplificar la molécula plantilla sintética y la plantilla biológica con la misma eficiencia de amplificación. Por lo tanto, la eficiencia de amplificación de un conjunto de cebadores en una plantilla sintética es la misma para la plantilla biológica correspondiente con las mismas secuencias V y J que la plantilla sintética.

Las plantillas sintéticas se usan como controles en línea para medir las eficiencias de amplificación de los pares de cebadores en un ensayo de PCR múltiple. Los amplicones resultantes de moléculas plantilla sintéticas y plantillas biológicas se secuencian mediante el uso de técnicas de secuenciación de alta productividad conocidas, tales como Illumina®. Los métodos y composiciones para minimizar el sesgo de amplificación se describen en el documento WO2013/169957.

I. Extracción y agrupamiento de vBloques

Los métodos de la invención incluyen la identificación de todas las secuencias de vBloques a partir de los resultados de la secuenciación que incluyen los vBloques amplificados y las secuencias biológicas amplificadas. Los vBloques se identifican a través de métodos estadísticos que identifican la presencia de secuencias de nucleótidos aleatorios frente a la ausencia de secuencias de nucleótidos aleatorios en las secuencias biológicas amplificadas. La combinación única de las ID de cada gen V y gen J también se identifica, lo que permite por lo tanto la identificación y segregación de todos los vBloques y permite además la identificación de cada combinación V/J mostrada en cada vBloque.

Para analizar los datos de los vBloques, las lecturas de secuencias de vBloques se extraen del archivo de secuenciación y se agrupan.

En una modalidad, se usa un algoritmo para separar secuencias de vBloques de un archivo de datos que incluye secuencias biológicas.

Los siguientes parámetros pueden definirse:

-dist_máx = distancia de edición máxima a la cual puede declararse que una lectura coincide con una secuencia de vBloque (con exclusión del segmento de nucleótidos aleatorios de N pb, por ejemplo donde N es 8 o 12).
-mapeo = nombre del archivo de salida para lecturas derivadas de error agrupadas en consenso.

-hilos = # de procesos a usar para cálculos multihilo.

Comparar cada secuencia leída contra todos los vBloques.

Cuando se compara una secuencia leída contra un vBloque dado, eliminar las bases en la ubicación esperada de las secuencias oligonucleotídicas aleatorias de ambas secuencias.

Primer pase: Comparar las secuencias leídas vs. vBloques por métrica de Hamming. La distancia de Hamming entre dos cadenas de igual longitud es el número de posiciones en las que los símbolos correspondientes son diferentes. Cualquier lectura con una distancia de Hamming \leq el parámetro -dist_máx se considera coincidente con un vBloque. En este caso, la secuencia oligonucleotídica aleatoria se identifica mediante el registro de la secuencia leída en la ubicación esperada de las secuencias oligonucleotídicas aleatorias en la secuencia de vBloque de mejor coincidencia.

Segundo pase: Para las lecturas que no encontraron una buena coincidencia mediante distancia de Hamming, repetir la comparación de lectura vs. vBloque con una métrica de Levenshtein. La distancia de Levenshtein es una métrica de cadena para medir la diferencia entre dos secuencias. Cualquier lectura con una distancia de Levenshtein \leq el parámetro de -dist_máx se considera coincidente con un vBloque con indel (inserción/deleciones). En este caso, la secuencia oligonucleotídica aleatoria se identifica mediante el recuento de la ubicaciones de los indel en el alineamiento de secuencias.

Agrupar las lecturas identificadas como vBloques mediante el colapso de las secuencias que comparten una secuencia oligonucleotídica aleatoria (coincidencia exacta de cadena).

Designar la secuencia de vBloque esperada como el consenso fuera de la secuencia oligonucleotídica aleatoria.

Imprimir la secuencia consenso de cada vBloque con una secuencia oligonucleotídica aleatoria única, el número correspondiente de lecturas, y las ID de los genes V y J.

II. Métodos de normalización de vBloques

Los métodos de la invención incluyen el cálculo de los factores de normalización para todas las combinaciones posibles de genes V/J. Un factor de normalización es un número que, cuando se multiplica por el recuento de lecturas de una secuencia, cambia el recuento de lecturas al valor que podría esperarse si no hubiera sesgo de amplificación por PCR. Por ejemplo, los genes que tienden a la subamplificación tendrán números mayores que uno de modo que el producto (recuento de lecturas x factor de normalización) es mayor que el recuento de lecturas original. El factor de normalización es el recíproco del factor de amplificación, como se describe a continuación.

En una modalidad, el método de normalización incluye las siguientes etapas:

1) Leer un archivo de secuencias de vBloques esperadas y las ID de los genes V y J que estas modelan.

2) Leer un archivo de secuencias de vBloques observadas, sus recuentos de lecturas, e ID de genes V y J (como se identifica por el algoritmo de extracción de vBloques)". Usar las ID de los vBloques para ubicar cada recuento de lectura en un vBloque esperado. Para cada vBloque esperado, construir una lista de recuentos de lecturas a partir de los datos observados.

3) Calcular los factores de normalización para los genes V:

A) Para cada gen V único en los vBloques esperados, y para cada gen J de "referencia" (por ejemplo, TRCBJ1-2, TRCBJ2-6) asociado con el gen V único, calcular el recuento de lecturas promedio para las secuencias de vBloques observadas con esta combinación V/J. Añadir el recuento promedio a una lista de recuentos de lecturas promedio para este gen V.

B) Calcular la media general de los recuentos de lecturas promedio de todos los genes V únicos y los genes J de referencia.

C) Para cada gen V único en los vBloques esperados, y para cada gen J de "referencia", dividir el recuento de lecturas promedio para esta combinación V/J por la media general de recuentos de lecturas promedio. Esto produce un factor de amplificación para el gen V actual y cada J de referencia.

D) Tomar la media de los factores de amplificación para el gen V actual en diferentes genes J de referencia, después tomar el recíproco; esto produce un factor de normalización para el gen V.

4) Calcular los factores de normalización para los genes J mediante el uso del mismo enfoque anterior, con los genes V y J en papeles invertidos. Los genes V de referencia son TCRBV03-1 y TCRBV21.

Este método produce un factor de normalización para cada gen V y gen J.

Métodos para calcular el número de plantillas iniciales en ensayos de inmunosecuenciación

En algunos casos, las moléculas plantilla sintéticas pueden usarse para cuantificar el número de plantillas iniciales en un experimento de inmunosecuenciación. Los métodos de amplificación por PCR múltiple y secuenciación de alta productividad ("inmunosecuenciación") se describen en detalle, al menos en los documentos US2010/0330571 y US2012/0058902.

En algunos casos, se usa un ensayo de PCR para seleccionar una región CDR3 de cadenas de TRB reordenadas que amplifica un fragmento de 110 pares de bases (pb). Dado que la región de interés (ROI) es de aproximadamente 110 pb, los pares de cebadores usados para estimar los números totales de genomas iniciales también se requieren para amplificar regiones de aproximadamente 110 pb del genoma. Para usar la secuenciación por síntesis, los pares de cebadores también necesitan secuencias adaptadoras 5'. Estas secuencias adaptadoras pueden ser adaptadores de secuenciación por síntesis o cebadores universales que pueden usarse después para aplicar los adaptadores de secuenciación por síntesis. En esta modalidad, los cebadores incluyen segmentos colgantes de cebadores universales de pGEX 5'. Los adaptadores de secuenciación por síntesis pueden añadirse después con una segunda reacción de PCR mediante el uso de estos cebadores universales (SEQ ID NO:765-786). Los métodos para el uso de cebadores universales en la secuenciación de alta productividad para amplificar receptores TCR e IG reordenados se describen en los documentos US2010/0330571 y US2012/0058902.

Una vez diseñados los pares de cebadores, el método incluye diseñar las plantillas sintéticas de control. Las plantillas sintéticas se diseñan para garantizar que los pares de cebadores amplifican las plantillas sintéticas con la misma eficiencia que las regiones genómicas. Para hacer esto, se requiere que las plantillas sintéticas incluyan los mismos sitios de cebado que el genoma, y los pares de cebadores deben amplificar la región del mismo tamaño. Adicionalmente, las plantillas sintéticas deben incluir además una cadena interna de secuencias de nucleótidos que diferencia las secuencias derivadas de las secuencias sintéticas de las derivadas del genoma. Aunque el número de nucleótidos requeridos puede ser de un par de bases, en una modalidad, las secuencias amplificadas de las secuencias sintéticas pueden diferir en 26 pares de

bases de las secuencias amplificadas derivadas del genoma. Dieciséis pb de estas secuencias diferentes son un código de barras único de 16 pb que identifica de manera única cada plantilla sintética. En algunas modalidades, las plantillas sintéticas pueden diseñarse como ADN bicatenario (por ejemplo, adquirido de una compañía como Integrated DNA Technologies) y no requieren procesamiento, o se diseñan como una serie de cebadores o un cebador largo (por ejemplo, adquirido de una compañía de síntesis de cebadores como IDT o Invitrogen), y pueden amplificarse para obtener ADN bicatenario. En una modalidad, se diseñaron cebadores monocatenarios extremadamente largos (y se adquirieron de Integrated DNA Technologies). En una modalidad, estas plantillas sintéticas incluyen regiones de cebado 5' y 3' que permitieron la amplificación para generar ADN bicatenario.

Para estimar el número de plantillas genómicas totales añadidas a una reacción de PCR, la reacción de PCR incluye enzima, plantilla y cebadores, que incluyen cebadores de ROI. En una modalidad, el ensayo incluye un conjunto múltiple de cebadores para amplificar 110 pares de bases de la región CDR3 del locus de TRB reordenado, el(los) par(es) de cebadores de control genómico, y un número de plantilla(s) sintética(s) conocido o que puede conocerse. En este ejemplo, se usa una segunda reacción de PCR para añadir los adaptadores de secuenciación por síntesis. Después la genoteca se secuencía, por ejemplo, mediante el uso de un método de secuenciación por síntesis.

Para estimar el número de genomas iniciales, se cuenta el número total de secuencias sintéticas para cada plantilla sintética. Debido a que el número de plantillas sintéticas añadidas a la PCR se conoce, ya sea por diseño de plantillas o por técnica de biología molecular cuidadosa, la cobertura, que es el número de copias secuenciadas de cada plantilla sintética, puede calcularse. La cobertura puede usarse después para estimar el número de genomas iniciales. Debido a que los pares de cebadores amplifican las plantillas sintéticas y biológicas con la misma afinidad, la cobertura de las plantillas sintéticas también representa la cobertura de las plantillas biológicas. Dado esto, para calcular el número total de genomas iniciales, se puede dividir el número total de secuencias biológicas por la cobertura. Esto se repite para cada región única del genoma muestreado.

El método incluye los siguientes cálculos:

- A) Recuento del número total de secuencias para cada plantilla sintética
- B) Recuento del número único de moléculas sintéticas de control de ADN añadidas a la reacción de PCR, que representa el número inicial de plantillas sintéticas.
- C) Cobertura = número total de secuencias sintéticas / número inicial de plantillas sintéticas.
- D) Recuento del número total de secuencias para cada tipo de fragmento biológico amplificado por los pares de cebadores de control de ADN,
- E) Número de genomas iniciales = número total de secuencias biológicas / cobertura.

Para mejorar la confianza en los datos, este proceso podría repetirse para cada región única del genoma muestreado. Por ejemplo, si hay 7 pares de cebadores únicos que amplifican 7 regiones del genoma, pueden obtenerse 7 mediciones idénticas pero independientes del número de genomas iniciales.

Pueden hacerse modificaciones para mejorar la exactitud y la eficiencia del método. Una modalidad de la invención incluye cebadores y plantillas sintéticas no solo para una región genómica, sino pares de cebadores y plantillas sintéticas para muchas regiones genómicas. Aunque esta modalidad requiere más lecturas de secuenciación para cada muestra, permite mediciones independientes de los genomas iniciales.

Por ejemplo, si se diseñan e implementan múltiples pares de cebadores que amplifican y miden el número total de genomas iniciales para diferentes cromosomas, podría usarse el promedio de cada par de cebadores y/o identificar si los pares de cebadores específicos están proporcionando recuentos anómalos. Las células cancerosas tienen una incidencia más alta de aneuploidía que las células normales, lo que aumenta la probabilidad de que este método estime de manera inexacta el número de genomas iniciales. Mediante el uso de múltiples pares de cebadores, especialmente múltiples pares de cebadores que tienen como diana regiones del genoma en cromosomas diferentes, se pueden desechar los valores atípicos.

Otro caso descrito en la presente descripción que aumenta la exactitud del método es aumentar el número de plantillas sintéticas únicas subyacentes a cada marcador (cada par de cebadores tiene muchas plantillas sintéticas). Un caso incluye el aumento del número de plantillas sintéticas para incluir una cadena de nucleótidos aleatorios entre los pares de cebadores. En este caso, el método incluye el uso de 10 nucleótidos aleatorios, lo que aumenta el número de plantillas sintéticas únicas para cada par de cebadores de 1 a 1.048.576 plantillas únicas. En este caso, se añaden menos de 10.000 plantillas sintéticas únicas a la reacción de PCR, lo que garantiza que cada plantilla sintética sea una secuencia única. Esto proporciona un recuento exquisitamente exacto del número de plantillas sintéticas iniciales.

En un caso, la ecuación para calcular la cobertura es:

Cobertura = número de secuencias sintéticas totales / número de secuencias sintéticas únicas.

Estas dos modificaciones diferentes del método, por ejemplo, muestreo de múltiples regiones genómicas y/o aumento del número de especies de cada plantilla sintética, aumentan la exactitud del estimado del número de genomas iniciales. Sin embargo, ambos métodos aumentan la cantidad de lecturas de secuenciación requeridas para lograr un estimado exacto. Las dos modificaciones pueden retener la exactitud del método, a la vez que disminuyen la cantidad requerida de espacio de secuenciación.

Otro método incluye el aumento del número de reacciones por muestra de una a dos. En este caso, una reacción incluye un volumen grande de plantilla y los cebadores para la ROI. La otra reacción incluye un volumen menor pero consistente de plantilla (1/2, 1/4, 1/8, o 1/16 del volumen de plantilla usado para la ROI) y los cebadores y plantillas sintéticas para estimar el número total de genomas. En esta modalidad, la ecuación para calcular el número de genomas iniciales añadidos a la reacción de ROI se modifica a:

Número de genomas iniciales = (número total de secuencias biológicas / cobertura) * (Volumen de plantilla de ROI / Volumen de plantillas de genomas)

En otro caso, el método mantiene el uso de una reacción de PCR para amplificar tanto la ROI como las regiones para estimar el número total de genomas. En este caso, la reducción del número de secuencias genómicas muestreadas se logra mediante la modificación de las secuencias adaptadoras 5' en el(los) par(es) de cebadores usado(s) para amplificar las regiones genómicas. En este caso, los pares de cebadores usan dos o más secuencias adaptadoras 5'. Una de las secuencias adaptadoras 5' es idéntica a la secuencia adaptadora 5' usada en los cebadores de ROI. El resto de las secuencias adaptadoras 5' usan una secuencia diferente (una a muchas). Cuando los adaptadores de secuenciación por síntesis se añaden a los amplicones de PCR con una segunda PCR, solo se incluye la secuencia adaptadora usada por la ROI. En esta modalidad, durante la segunda reacción de PCR, solo los amplicones con la secuencia adaptadora 5' que coincide con la ROI recibirán adaptador de secuenciación por síntesis. Esto tomará una submuestra de las plantillas genómicas de secuenciación.

Número de genomas iniciales = (número total de secuencias biológicas / cobertura) *
(secuencia adaptadora 5' de interés / cebadores de secuenciación de genomas totales).

Las SEQ ID NO:3254-3268 demuestran secuencias ilustrativas de segmentos génicos amplificados, cebadores directos e inversos de genes específicos, y gen constitutivo como control sintético, como se usa en la invención.

Las SEQ ID NO:3269-3274 demuestran ejemplos de cebadores con adaptador (pGEX_GAPDPH_108bp_F y pGEXR_GAPDPH_108bp_R), secuencias adaptadoras ilustrativas (Cebador con adaptador SEQ pGEXF, Cebador con adaptador SEQ pGEXR), y adaptadores de secuenciación por síntesis (OligoC_PERead2Seq_N6_WD565_pGEXr y OligoD_PERead1Seq_WD565_N6_pGEXf).

EJEMPLOS

Ejemplo 1: Validación de la eficacia de secuencias sintéticas para determinar el factor de normalización en la corrección del sesgo de amplificación

Las moléculas plantilla sintéticas que incluyen secuencias oligonucleotídicas aleatorias (llamadas "vBloques" en la presente descripción) se probaron en métodos para minimizar el sesgo de amplificación y se compararon con moléculas plantilla sintéticas que no incluyen secuencias oligonucleotídicas aleatorias (llamadas "gBloques" en la presente descripción). Como se muestra en el documento WO2013/169957, las moléculas plantilla sintéticas llamadas gBloques proporcionan una referencia para las moléculas plantilla sintéticas usadas para medir y minimizar el sesgo de amplificación de cebadores de PCR múltiple.

Los vBloques se diseñaron y se construyeron como se muestra en las Figuras 1A y 1B. Cada plantilla sintética puede sintetizarse como oligonucleótidos de 495 pares de bases con la siguiente estructura (5' a 3'): (1) una secuencia adaptadora universal (UA), (2) un código de barras de 16 pares de bases que identifica segmentos V y J, (3) un gen V (de alrededor de 300 pares de bases), (4) un marcador interno de plantilla sintética (IM) de 9 pares de bases, (5) una repetición del código de barras de 16 pares de bases, (6) una cadena de 12 oligonucleótidos aleatorios (N12), (7) un gen J (de alrededor de 100 pares de bases), (8) una repetición del código de barras de 16 pares de bases, y (9) una secuencia adaptadora universal (UB).

Los vBloques y gBloques se evaluaron para determinar si los mismos factores de amplificación y factores de normalización podrían determinarse en ausencia de la secuencia de nucleótidos aleatorios. Se utilizaron dos corridas de PCR: la primera corrida usó 5.812 moléculas de vBloques únicas, y la segunda corrida utilizó 1.245 moléculas de vBloques únicas. Se determinó que el sesgo de amplificación para cada uno de los vBloques y gBloques en ambas corridas era similar, como se representa en la Figura 2. Cada punto representa el sesgo de amplificación promedio observado para plantillas

sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. La leyenda de cada gráfica muestra la correlación de Pearson al cuadrado (R^2) entre las mediciones del sesgo de amplificación de vBloques y gBloques. La correlación es más fuerte en la gráfica a la izquierda debido a que la Corrida 1 de PCR incluyó un número mayor de vBloques.

5 Se encontró que las mediciones del sesgo de amplificación son consistentes en diferentes genes V y J de referencia en ambos ensayos de PCR, como se representa en la Figura 3. Al igual que antes, cada punto representa el sesgo de amplificación promedio observado para plantillas sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. Aquí, se calcularon las correlaciones de Pearson al cuadrado (R^2) entre las mediciones del sesgo de amplificación de diferentes genes V y J de referencia en una corrida de PCR dada. La correlación es más fuerte en la gráfica a la izquierda debido a que la Corrida 1 de PCR incluyó un número mayor de vBloques.

15 Tanto los vBloques como los gBloques demostraron mediciones estables del sesgo de amplificación en las diferentes corridas de PCR, como se representa en la Figura 4. Al igual que antes, cada punto representa el sesgo de amplificación promedio observado para plantillas sintéticas con un gen V (tono más oscuro) o gen J (tono más claro) dado. Aquí, se calcularon las correlaciones de Pearson al cuadrado (R^2) entre las mediciones del sesgo de amplificación de gBloques (izquierda) y vBloques (derecha). La correlación es más fuerte en la gráfica a la izquierda debido a que se usaron números mayores de gBloques que de vBloques en las dos corridas.

20 En conjunto, los vBloques produjeron mediciones del sesgo de amplificación de cebador que eran consistentes con los estimados de gBloques, consistentes en diferentes genes V y J de referencia, y consistentes en ambas corridas de PCR.

Ejemplo 2: Determinación de los factores de normalización en la corrección del sesgo de amplificación en la amplificación de polinucleótidos en células de la inmunidad adaptativa

25 Los vBloques y las moléculas biológicas reordenadas de regiones CDR3 de células T se amplificaron en reacciones de PCR múltiple y se secuenciaron mediante el uso de los métodos descritos anteriormente. Los datos de las lecturas de secuencias de vBloques se extraen del archivo de secuenciación y se agrupan si se determina que comparten la misma secuencia de nucleótidos aleatorios. Se usaron dos pasos estadísticos para identificar los vBloques. El primer pase usó la métrica de Hamming para comparar las secuencias leídas frente a las secuencias de vBloques, en donde la secuencia de nucleótidos aleatorios se identificó mediante el registro de la secuencia leída en la ubicación esperada de las secuencias de nucleótidos aleatorios en la secuencia de vBloque de mejor coincidencia. El segundo pase se utilizó para lecturas que no encontraron una buena coincidencia por la métrica de Hamming. La métrica de Levenshtein se usó en este caso, en donde la secuencia de nucleótidos aleatorios se identificó mediante la determinación de las ubicaciones de los indel en el alineamiento de secuencias. Tras completar los dos pasos estadísticos, las lecturas identificadas como vBloques se agruparon mediante el colapso de las secuencias que compartían la misma secuencia de nucleótidos aleatorios.

40 El sesgo de amplificación se determinó mediante la determinación del recuento de lecturas de cada secuencia de vBloque colapsada que comprende una combinación V/J única, y el mapeo del recuento de lecturas de cada secuencia biológica de salida en el vBloque correspondiente que contiene la misma combinación V/J.

45 Los factores de normalización para los genes V se calcularon mediante el cálculo del recuento de lecturas promedio de cada vBloque observado con un gen V único, que estaba acompañado por un gen J de referencia. Por lo tanto, los recuentos de lecturas promedio para cada combinación V/J única se determinaron y se compilaron en una lista de recuentos de lecturas promedio para vBloques que comprenden el gen V específico. A partir de cada una de estas listas compiladas, se calculó la media general de los recuentos de lecturas promedio de todos los genes V únicos y genes J de referencia.

50 Para cada uno de los genes V únicos en los vBloques esperados, y para cada gen J de referencia, el recuento de lecturas promedio para esta combinación V/J se dividió por la media general de los recuentos de lecturas promedio, para llegar así al factor de amplificación para un gen V único y cada gen J de referencia. Se calculó la media de los factores de amplificación para cada combinación de un gen V único con diferentes genes J de referencia y después se tomó el recíproco; para producir así el factor de normalización para el gen V único.

55 Los factores de normalización para cada uno de los genes J únicos también se calcularon mediante el uso del mismo enfoque anterior, con los genes V y J en papeles invertidos. El número de lecturas de cada combinación V/J única se multiplicó después por los factores de normalización específicos, para llegar así a una lectura exacta que se ha corregido para el sesgo de amplificación.

60 Ejemplo 3: Determinación del número de genomas iniciales de una muestra de células de la inmunidad adaptativa

En este ejemplo, se usaron plantillas sintéticas y genes de control genómico para calcular con exactitud la representación relativa de células de la inmunidad adaptativa en una muestra que contiene células de la inmunidad adaptativa y células que no son de la inmunidad adaptativa.

65

Fuente de la muestra:

5 Las células T se aislaron a partir de sangre total mediante el uso de técnicas de biología celular estándar. Se extrajo el ADN de la población de células T purificadas. El ADN se normalizó, asumiendo 6,4 pg de ADN/genoma humano bicatenario de manera que aproximadamente 5 genomas, 250 genomas, 1.250 genomas, o 6.250 genomas de ADN de células T se añadieron a una reacción de PCR de TCRB estándar.

Reacción de PCR múltiple:

10 Ensayo de TCRB: Los genes de TCRB reordenados se amplificaron mediante el uso de una PCR múltiple. Los cebadores de segmentos V y segmentos J se diseñaron para amplificar fragmentos reordenados de ~ 110 pb. Las plantillas sintéticas se añadieron a cada reacción de PCR y se amplificaron con los mismos cebadores, y las plantillas sintéticas incluían un código de barras para diferenciarlas de las plantillas biológicas. El volumen de ADN necesario para añadir 5.250, 1.250, y 6.250 genomas se añadió a cada reacción de PCR.

15 Una segunda reacción de anexión por PCR se realizó mediante el uso de cebadores de anexión que comprenden códigos de barras específicos de pocillo y adaptadores de secuenciación de Illumina. La reacción de anexión por PCR añadió códigos de barras específicos de pocillo y adaptadores de secuenciación Illumina a cada producto de PCR.

20 Ensayo de control genómico:

Además del ensayo de TCRB, cinco loci autosómicos de copia única se amplificaron mediante el uso de un ensayo de PCR múltiple. Cada locus autosómico de copia única está presente en cada célula y sirve como un control genómico. 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.

30 La reacción de PCR múltiple incluyó la coamplificación de plantillas sintéticas que incluyen secuencias oligonucleotídicas de cada uno de los cinco genes autosómicos. Las plantillas sintéticas incluían códigos de barras únicos que identifican las moléculas como plantillas sintéticas y una secuencia aleatoria de 6 pb. Se usó la misma concentración de ADN para los controles genómicos que para los genes de TCRB, pero a un octavo del volumen, de manera que menos de 1 genoma, 31, 156, 781 genomas bicatenarios se añadieron a cada reacción de PCR. Los códigos de barras específicos de pocillo y los adaptadores de secuenciación de Illumina se añadieron a cada producto de PCR en un segundo ensayo de PCR de anexión, como se describió anteriormente.

35 Secuenciación:

Las muestras se mezclaron, se normalizaron, y se cargaron en un instrumento MiSEQ de Illumina. Los datos de secuencia de salida se procesaron, y las lecturas de secuencias de las plantillas sintéticas se usaron para medir la cobertura de la secuenciación. La cobertura de la secuenciación es un estimado del número de grupos de secuenciación derivados de una molécula individual añadida a la reacción de PCR.

Análisis:

45 El número de moléculas de TCRB en la muestra se estimó mediante el uso de los métodos descritos anteriormente. El número de genomas añadidos al ensayo de TCRB se determinó mediante la estimación del número de genomas en el ensayo de control genómico como se describió anteriormente (Sección III). El número de genomas calculado a partir del ensayo de control genómico se incrementó en 4 para tener en cuenta 1) que había 2 loci/genomas y 2) la reducción en ocho veces del material inicial (Figura 6).

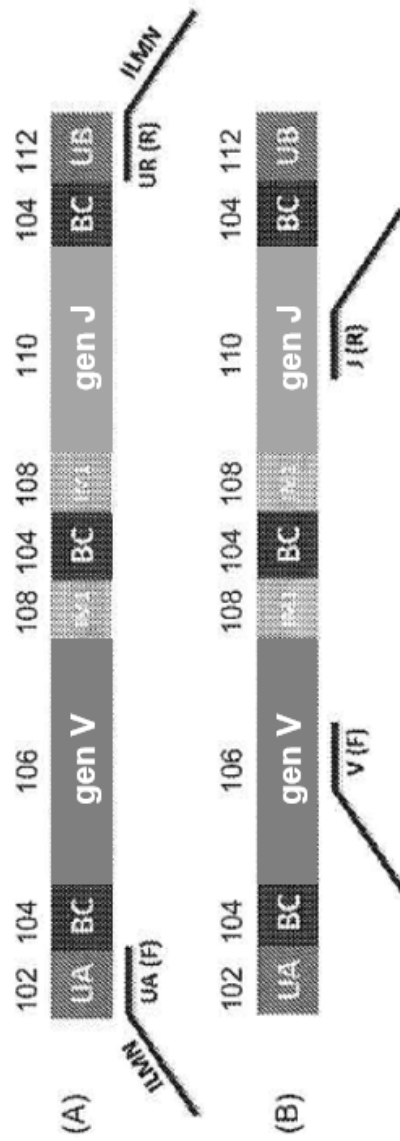
50 Debe señalarse que, como se usa en la descripción, las formas singulares “un”, “una” y “el/la” incluyen referencias plurales a menos que el contexto lo indique claramente de cualquier otra manera.

REIVINDICACIONES

1. Un método para corregir el sesgo de amplificación en una reacción de PCR de una muestra, el método comprende:
- 5
- A) amplificar por PCR múltiple, secuenciar, y cuantificar las lecturas de salida de:
- (i) moléculas plantilla biológicas que comprenden secuencias oligonucleotídicas de CDR3 reordenadas de loci de receptor de células T (TCR) de células T o loci de inmunoglobulina (Ig) de células B, cada secuencia comprende un segmento V de TCR o IG y un segmento J de TCR o IG, para obtener un número total de lecturas de secuencias biológicas de salida; y
- 10
- (ii) moléculas plantilla sintéticas que comprenden cada una un segmento V de TCR o Ig y un segmento J de TCR o IG, secuencias adaptadoras universales de cebadores directos y/o inversos, uno o más códigos de barras que identifican las moléculas plantilla como sintéticas, una secuencia oligonucleotídica de marcador interno, y una secuencia oligonucleotídica aleatoria, en donde cada secuencia oligonucleotídica aleatoria comprende una secuencia de nucleótidos única, y en donde cada molécula plantilla sintética comprende una combinación única de un segmento V y un segmento J, para obtener un número total de lecturas de secuencias sintéticas de salida;
- 15
- B) agrupar las lecturas de secuencias al:
- (i) extraer dichas lecturas de secuencias;
- (ii) identificar si una lectura de secuencia es una lectura de secuencia biológica o una lectura de secuencia sintética al:
- 20
- (a) comparar las lecturas de secuencias contra las secuencias plantilla sintéticas conocidas mediante el uso de una primera métrica para identificar secuencias plantilla sintéticas, mientras se ignora la porción de la lectura de secuencia que se espera corresponda a la secuencia oligonucleotídica aleatoria en secuencias sintéticas;
- 25
- (b) comparar las lecturas de secuencias sin coincidencia restantes contra las secuencias plantilla sintéticas conocidas mediante el uso de una segunda métrica, mientras se ignora la porción de la lectura de secuencia que se espera corresponda a la secuencia oligonucleotídica aleatoria en secuencias sintéticas;
- (iii) agrupar las lecturas de secuencias sintéticas mediante el colapso de las lecturas de secuencias que coinciden con la misma secuencia oligonucleotídica sintética esperada y comparten la misma secuencia oligonucleotídica aleatoria;
- 30
- (iv) asignar a cada grupo de lecturas de secuencias sintéticas, en base a la secuencia oligonucleotídica sintética esperada con la que coincidieron, una secuencia consenso que comprende la secuencia esperada de la secuencia oligonucleotídica sintética con la que coincidieron, que incluye un segmento V y un segmento J; y
- (v) determinar un número total de lecturas de secuencias sintéticas observadas en cada grupo;
- 35
- C) calcular uno o más factores de normalización para los segmentos V y los segmentos J en las lecturas de secuencias sintéticas al:
- (i) calcular un recuento de lecturas promedio entre todos los grupos de secuencias que coincidieron con cada secuencia oligonucleotídica sintética;
- 40
- (ii) calcular una media general de los recuentos de lecturas promedio para cada segmento V y J único, entre las secuencias oligonucleotídicas sintéticas que contienen un V dado y cualquier J o viceversa;
- (iii) calcular un sesgo de amplificación promedio mediante la división del recuento de lecturas promedio de cada segmento V y segmento J calculado en la etapa C(ii) por la media general de los recuentos de lecturas promedio de los segmentos V o de los recuentos de lecturas promedio de los segmentos J calculada en la etapa C(ii) para llegar a un factor de amplificación para cada segmento V y J; y
- 45
- (iv) producir el factor de normalización para un segmento V o J dado mediante el cálculo del recíproco del sesgo de amplificación promedio producido en la etapa C(iii); y
- D) multiplicar el recuento de lecturas de secuencias observadas de cada secuencia biológica única dada por el factor de normalización calculado en la etapa C(iv) correspondiente al segmento V presente en esa secuencia biológica única y por el factor de normalización calculado en la etapa C(iv) correspondiente al segmento J presente en esa secuencia biológica única, para corregir de este modo el sesgo de amplificación en la reacción de PCR múltiple de la muestra.
- 50
2. El método de conformidad con la reivindicación 1, en donde la etapa de comparar las lecturas de secuencias contra las secuencias plantilla sintéticas conocidas se realiza con una métrica de Hamming.
- 55
3. El método de conformidad con la reivindicación 1 o la reivindicación 2, en donde la etapa de comparar las lecturas de secuencias sin coincidencia restantes contra las secuencias plantilla sintéticas conocidas se realiza con una métrica de Levenshtein.
- 60
4. El método de conformidad con cualquiera de las reivindicaciones 1 a 3, en donde la muestra se obtiene a partir de un sujeto mamífero.
5. El método de conformidad con cualquiera de las reivindicaciones 1 a 4, en donde la muestra comprende una mezcla de células T y/o células B, y células que no son células T o células B.
- 65

6. El método de conformidad con cualquiera de las reivindicaciones 1 a 5, en donde la muestra biológica es una biopsia de tumor.
- 5 7. El método de conformidad con cualquiera de las reivindicaciones 1 a 6, en donde las moléculas plantilla sintéticas comprenden la secuencia de la fórmula I:
5'-U1-B1-V-I-B2-N-J-B3-U2-3', en donde
- 10 (i) V es una secuencia oligonucleotídica que comprende al menos 20 y no más de 1.000 nucleótidos contiguos de una secuencia génica que codifica la región variable (V) de TCR o Ig, o el complemento de esta, y cada plantilla sintética comprende una secuencia oligonucleotídica de región V única;
- (ii) J es una secuencia oligonucleotídica que comprende al menos 15 y no más de 600 nucleótidos contiguos de una secuencia génica que codifica la región de unión (J) de TCR o Ig, o el complemento de esta, y cada plantilla sintética comprende una secuencia oligonucleotídica de región V única;
- 15 (iii) U1 comprende una secuencia oligonucleotídica que se selecciona de: (a) una primera secuencia oligonucleotídica adaptadora universal, y (b) una primera secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 5' a una primera secuencia oligonucleotídica adaptadora universal;
- (iv) U2 comprende una secuencia oligonucleotídica que se selecciona de: (a) una segunda secuencia oligonucleotídica adaptadora universal, y (b) una segunda secuencia oligonucleotídica de plataforma de secuenciación unida a y posicionada 3' a una segunda secuencia oligonucleotídica adaptadora universal;
- 20 (v) I es una secuencia oligonucleotídica de marcador interno que comprende al menos 2 y no más de 100 nucleótidos;
- (vi) N es una secuencia oligonucleotídica aleatoria que comprende al menos 2 y no más de 100 nucleótidos; y
- (vii) B1, B2, y B3 comprenden cada uno independientemente nada o secuencias oligonucleotídicas de código de barras de al menos 2 y no más de 100 nucleótidos que identifican de manera única, como una combinación pareada, (a) dichas secuencias oligonucleotídicas de región V únicas; y (b) dichas secuencias oligonucleotídicas de región J únicas, en donde al menos uno de B1, B2, y B3 están presentes en cada plantilla sintética.
- 25 8. El método de conformidad con la reivindicación 7, en donde N comprende al menos 4 y no más de 15 nucleótidos, o en donde N comprende 8 nucleótidos.
- 30 9. El método de conformidad con la reivindicación 1 en donde la amplificación por PCR múltiple en la etapa A se realiza mediante el uso de una pluralidad de conjuntos de cebadores oligonucleotídicos que comprende:
- (a) una pluralidad de cebadores oligonucleotídicos de segmentos V que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región V de TCR o Ig o al complemento de este, en donde cada cebador de segmento V comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región V de TCR o Ig y en donde dicha pluralidad de cebadores de segmentos V se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región V de TCR o Ig que están presentes en la composición, y
- 35 (b) una pluralidad de cebadores oligonucleotídicos de segmentos J que son cada uno independientemente capaces de hibridarse específicamente al menos a un polinucleótido que codifica un polipéptido de región J de TCR o Ig o al complemento de este, en donde cada cebador de segmento J comprende una secuencia de nucleótidos de al menos 15 nucleótidos contiguos que es complementaria al menos a un segmento génico funcional que codifica la región J de TCR o Ig y en donde dicha pluralidad de cebadores de segmentos J se hibrida específicamente a sustancialmente todos los segmentos génicos funcionales que codifican la región J de TCR o Ig que están presentes en la composición.
- 40 10. El método de conformidad con la reivindicación 9, en donde dicha pluralidad de cebadores oligonucleotídicos de segmentos V y dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprenden las secuencias que se exponen en las SEQ ID NO:1-764.
- 45 11. El método de conformidad con la reivindicación 9, en donde cualquiera o ambas de:
- (i) dicha pluralidad de cebadores oligonucleotídicos de segmentos V comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:1-120, 147-158, 167-276, 407-578, y 593-740, y
- 55 (ii) dicha pluralidad de cebadores oligonucleotídicos de segmentos J comprende secuencias que tienen al menos 90 % de identidad de secuencia con las secuencias de nucleótidos que se exponen en las SEQ ID NO:121-146, 159-166, 277-406, 579-592, y 741-764.
- 60 12. El método de conformidad con cualquiera de las reivindicaciones 1 a 11, en donde dicho 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.
- 65 13. El método de conformidad con cualquiera de las reivindicaciones 1 a 12, en donde dicho 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.

14. El método de conformidad con cualquiera de las reivindicaciones 1 a 13, en donde dicho segmento V de Ig comprende un segmento génico V de IGH, un segmento génico V de IGL, o un segmento génico V de IGK.
- 5 15. El método de conformidad con cualquiera de las reivindicaciones 1 a 14, en donde dicho segmento de región J de Ig comprende un segmento génico J de IGH, un segmento génico J de IGL, o un segmento génico V de IGK.



Figuras 1A y 1B

Figura 2

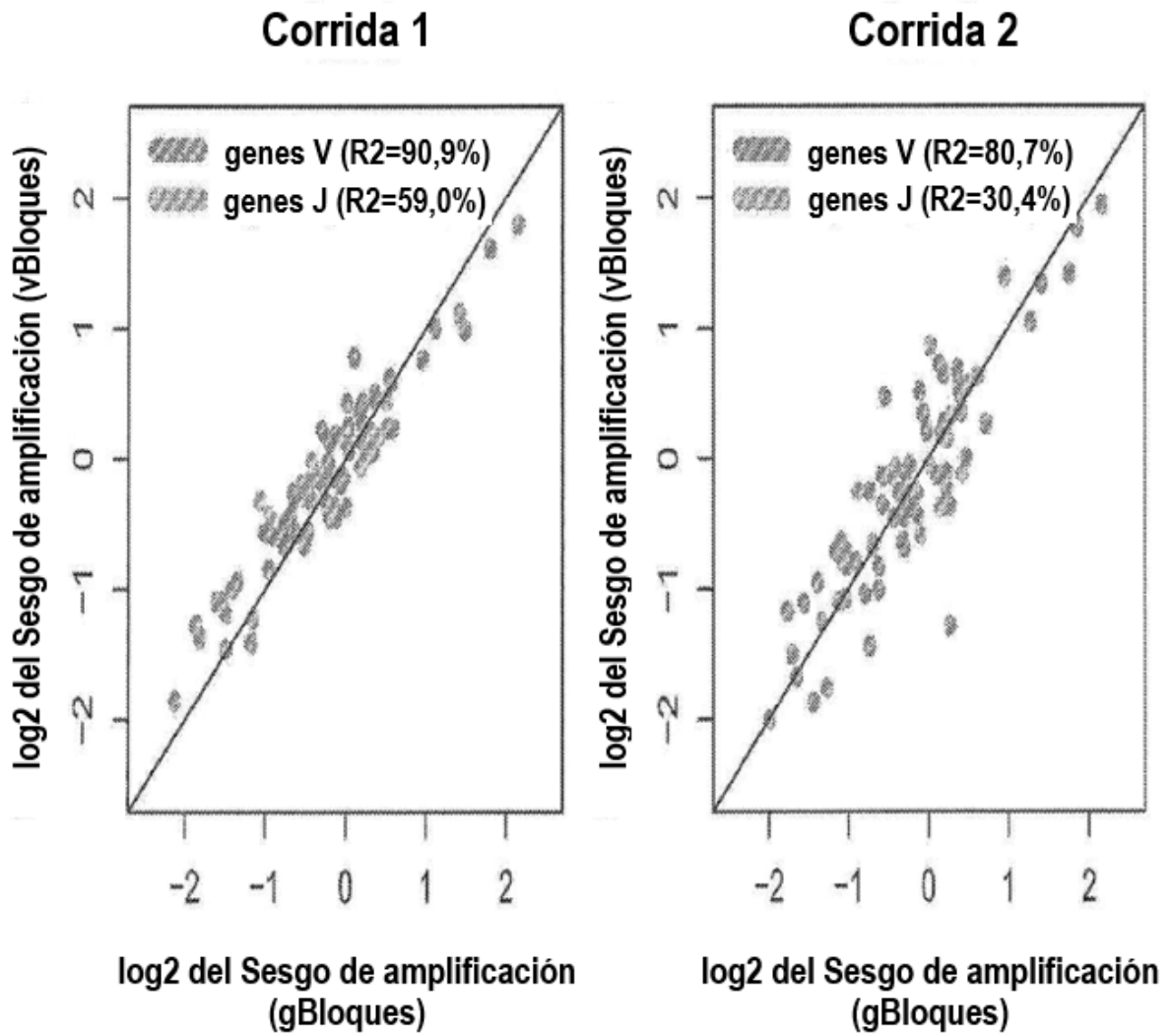


Figura 3

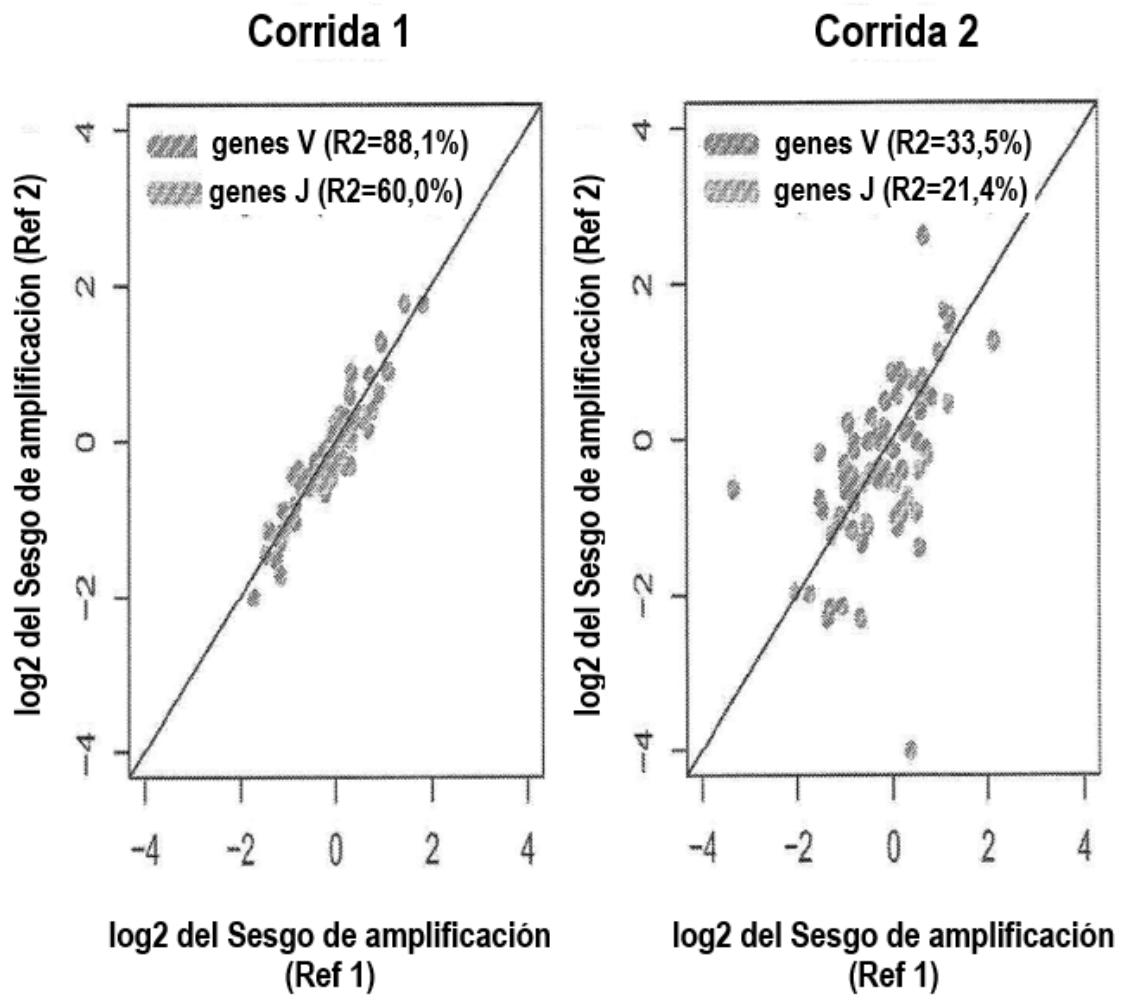


Figura 4

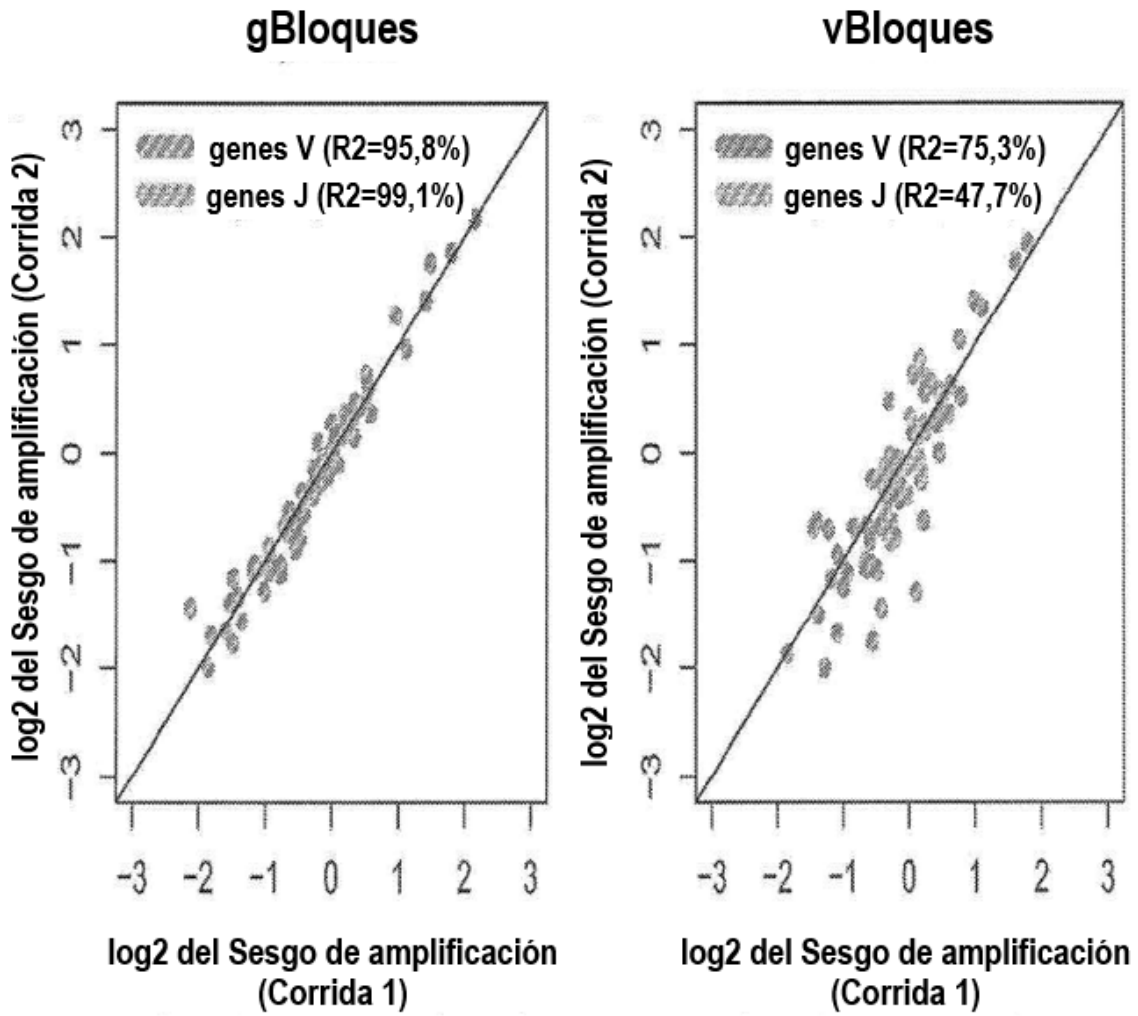


Figura 5

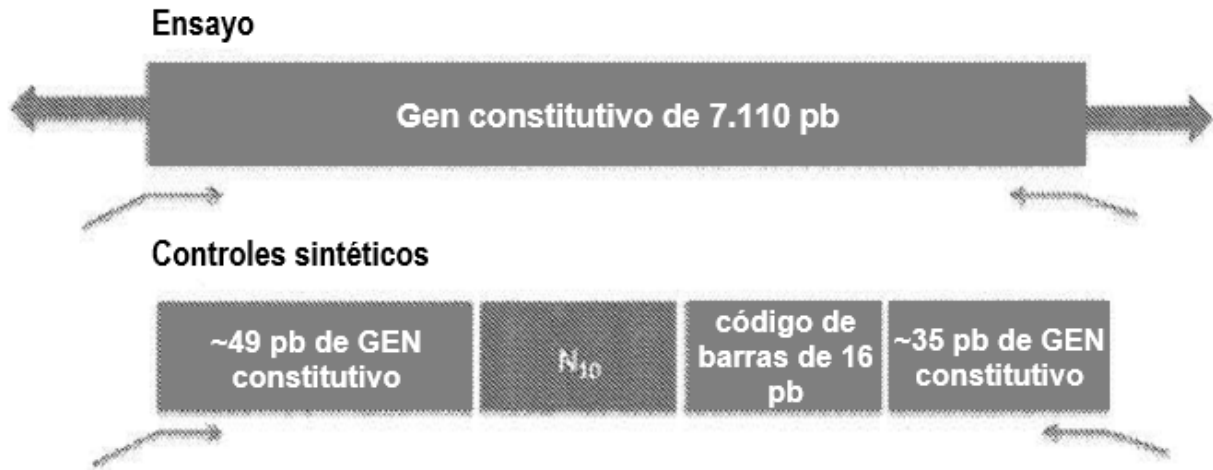
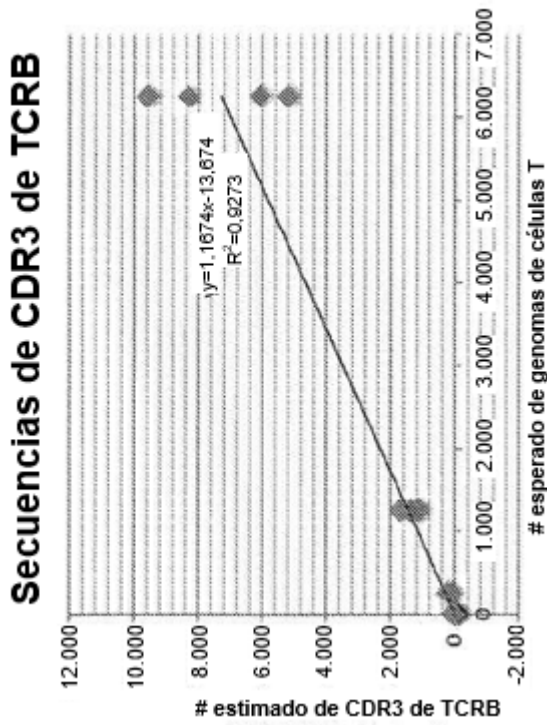


Figura 6



Loci de control genómico

