

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 767 696**

51 Int. Cl.:

C12Q 1/04 (2006.01)

C12Q 1/68 (2008.01)

C12Q 1/6883 (2008.01)

C12Q 1/6881 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.03.2016 PCT/US2016/021430**

87 Fecha y número de publicación internacional: **15.09.2016 WO16144996**

96 Fecha de presentación y número de la solicitud europea: **09.03.2016 E 16762365 (1)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019 EP 3268488**

54 Título: **Método para identificar patrones de CDR3 asociados a enfermedades en un repertorio inmunitario**

30 Prioridad:

09.03.2015 US 201562130512 P

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

18.06.2020

73 Titular/es:

**IREPERTOIRE, INC. (100.0%)
601 Genome Way, Suite 3005
Huntsville, AL 35806, US**

72 Inventor/es:

HAN, JIAN

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 767 696 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

Método para identificar patrones de CDR3 asociados a enfermedades en un repertorio inmunitario

5 Campo de la invención

La invención se refiere a métodos para reconocer repertorios inmunitarios asociados a enfermedades en sujetos humanos y/o animales y para el uso de tales métodos para el diagnóstico de enfermedades y para el estudio de procesos de enfermedades.

10

Antecedentes de la invención

Los diversos receptores de antígenos de los linfocitos T y B se producen por recombinación somática de un número limitado pero grande de segmentos de genes. Estos segmentos de genes - V (variable), D (diversidad), J (unión) y C (constante) - determinan la especificidad de unión y las aplicaciones posteriores de inmunoglobulinas y receptores de linfocitos T (TCR). La porción reordenada V(D)J del receptor, denominada la región V, es de gran interés, ya que es responsable del reconocimiento de epítomos. Cuando el V(D)J se traduce en una secuencia de aminoácidos, la región V puede subdividirse en varias partes que consisten en la secuencia líder, la región marco conservada (FR) 1, la región determinante de complementariedad 1 (CDR1), FR2, CDR2, FR3, CDR3, FR4 y los dominios C.

15

20

La CDR3 es de particular interés porque los estudios han indicado que esta región está asociada con la especificidad antigénica. En comparación con los sujetos normales, los pacientes con diversas enfermedades pueden experimentar cambios cuantitativos y/o cualitativos en su repertorio inmunitario. Los cambios cuantitativos pueden ser evidentes a medida que aumenta y disminuye la diversidad del repertorio inmunitario. Los cambios cualitativos pueden presentarse como un mayor intercambio de CDR3 específicas de enfermedades en los linfocitos T o B.

25

El sistema inmunitario monta una respuesta a diversas afecciones, tales como cáncer, infecciones bacterianas, infecciones víricas e infecciones fúngicas. Adicionalmente, en algunos sujetos, en realidad puede producir una respuesta nociva a los tejidos del cuerpo, dando como resultado, por ejemplo, una enfermedad autoinmunitaria o el rechazo de injertos. Los tipos y grados de estas respuestas inmunitarias potencialmente podrían, si se accede con precisión, ser uno de los indicadores más importantes y precisos de la presencia o ausencia de una enfermedad en particular o de respuestas inmunitarias indeseables.

30

Sin embargo, se estima que los seres humanos tienen hasta 10^{15} - 10^{25} linfocitos T diferentes, debido al número de posibles reordenamientos de VDJ, adiciones de n y combinaciones de cadenas alfa-beta. Suponiendo que, para una enfermedad en particular, solo hay 10^3 CDR3 específicas de la enfermedad que pueden cambiar cuantitativamente (a través de regulación positiva o negativa) o cualitativamente (a través de ganancia o pérdida), la relación señal/ruido en estas circunstancias es muy débil. Por lo tanto, los métodos convencionales no son prácticos para evaluar esta información con fines de diagnóstico.

35

40

Lo que se necesita son métodos mejorados para la evaluación de la respuesta inmunitaria y para el desarrollo de pruebas de diagnóstico basadas en dicha respuesta inmunitaria y el repertorio inmunitario resultante.

Sumario de la invención

45

En una realización, la presente divulgación se refiere a un método para desarrollar una prueba de diagnóstico usando el repertorio inmunitario, comprendiendo el método las etapas de: (a) recoger una muestra de cada uno de los múltiples sujetos en un grupo de pacientes y un grupo de control, en donde el grupo de pacientes comprende sujetos que tienen la misma enfermedad y el grupo de control comprende sujetos que se clasifican como sanos; (b) amplificar y secuenciar el repertorio inmunitario de cada sujeto en cada uno de los dos grupos para identificar cada secuencia de CDR3 exclusiva presente en la muestra y determinar la frecuencia de aparición de cada secuencia de CDR3 exclusiva; (c) identificar secuencias de CDR3 que se comparten entre al menos dos sujetos en cada uno de los grupos de control y el grupo de pacientes; (d) clasificar las secuencias de CDR3 identificadas por orden de frecuencia de aparición; (e) identificar Linklets de cada grupo; y (f) identificar los Linklets que están asociados en un grado estadísticamente significativo con el grupo de pacientes para proporcionar una firma de la enfermedad.

50

55

En determinadas realizaciones, la muestra es sangre periférica. En otras realizaciones, la muestra es tejido. En determinadas realizaciones, se identifican menos de 1.000 secuencias de CDR3 como compartidas entre al menos dos sujetos. En otras realizaciones, se identifican al menos 1.000 secuencias de CDR3 como compartidas entre al menos dos sujetos. En determinadas realizaciones, se identifican al menos aproximadamente 10^6 Linklets de cada grupo. En otras realizaciones, se identifican menos de 10^6 Linklets de cada grupo.

60

Breve descripción de las figuras

65

La divulgación se puede entender mejor con referencia a las siguientes figuras.

La figura 1 es una viñeta que ilustra el concepto básico del método divulgado, donde los datos de la secuencia de polinucleótidos de las células del sistema inmunitario se procesan utilizando un software diseñado para clasificar y contar las secuencias, clasificarlas por números de frecuencia, generar valores de p y otros criterios, para producir una firma de diagnóstico de los Linklets designados como "Linklets significativos".

La figura 2 es una tabla que ilustra la clasificación de las CDR3 por número de clones presentes en una muestra. Por lo general, un resultado de secuenciación de una muestra generará hasta 400.000 CDR3. Cada CDR3 está asociada con un recuento de lectura; debido a la naturaleza semicuantitativa del método de amplificación (arm-PCR), el recuento de lectura también refleja la abundancia relativa del clon. El análisis de software elimina los errores y clasifica las CDR3 para producir un archivo de salida como se muestra en la tabla.

La figura 3 es una tabla que ilustra cómo se detectan los Linklets en las secuencias obtenidas de las muestras de sangre. Las secuencias de CDR3 se contaron para proporcionar una lista de las secuencias de CDR3 que están presentes en mayor número en una muestra de sangre. Entre esas secuencias de CDR3, los Linklets representan pares de CDR3 que están presentes dentro de la misma muestra, a un nivel que es más alto que un nivel de corte designado.

La figura 4 es una lista representativa de algunos de los Linklets identificados durante un estudio de cáncer de mama (comparación entre Linklets detectados en dos grupos de estudio: un primer grupo de sujetos que se habían diagnosticado con cáncer de mama y un segundo grupo de sujetos que se designaron como controles sanos).

La figura 5 enumera los Linklets públicos detectados en el estudio de cáncer de mama, en función de sus valores de p. El Linklet mejor clasificado, por ejemplo, es el par de CDR3 'ASSYSR-GEEF' y 'ASSLGRTHQPQH', y este Linklet se compartió en 32 de las 98 muestras de pacientes con cáncer de mama, mientras que solo 1 de las 106 muestras de control lo tenía. El valor de p fue 0. Solo aquellos Linklets con un valor de p <0,05 se incluyen en la lista final que representa una firma de diagnóstico de cáncer de mama. Se identificaron un total de 101.902 Linklets significativos.

La Fig. 6 es un diagrama de dispersión que ilustra los resultados para sujetos en tres grupos: control, cáncer de mama y CMV (citomegalovirus). Un análisis de la curva de característica operativa del receptor (ROC, por sus siglas en inglés) sugirió un valor de corte de 600 DSL. De los 103 pacientes con cáncer de mama, 98 tenían más de 600 Linklets significativos; de los 110 controles, solo 7 tenían más de 600 Linklets significativos. La sensibilidad y especificidad diagnóstica fueron, por lo tanto, del 95 % y del 93 %, respectivamente. Cuando se estudiaron 188 muestras de cáncer no de mama (de pacientes inscritos en un estudio de CMV) contra la firma de la enfermedad de cáncer de mama, solo 3 muestras fueron positivos falsos con más de 600 Linklets significativos, dando una especificidad del 98,4 %.

Descripción detallada

La presente divulgación generalmente se refiere a un método para desarrollar pruebas de diagnóstico que se basan en la respuesta inmunitaria y el repertorio inmunitario resultante. El método divulgado actualmente aumenta la señal y reduce el fondo para permitir la identificación de CDR3 compartidas que pueden usarse para producir una firma de enfermedad. El método divulgado actualmente puede usarse para desarrollar una prueba de diagnóstico para diferentes enfermedades que incluyen, pero sin limitación, cáncer, enfermedad autoinmunitaria, enfermedad inflamatoria y enfermedad infecciosa.

Como se usa en el presente documento, "enfermedad" incluye enfermedad diagnosticada y otras alteraciones, diagnosticadas y no diagnosticadas, de la salud normal de un sujeto.

Como se usa en el presente documento, "sano" significa que actualmente no presenta síntomas de, y actualmente no está diagnosticado con, una enfermedad.

Como se usa en el presente documento, un "repertorio inmunitario" comprende los linfocitos T y B funcionalmente diversos de un sujeto.

Como se usa en el presente documento, un "Linklet" es un par de CDR3 exclusivas que están presentes en la misma muestra. Cuando dos o más personas comparten un Linklet en particular, es un "Linklet Público". Si un Linklet solo se detecta en un sujeto, es un "Linklet privado". Los Linklets públicos provienen de CDR3 públicas (es decir, CDR3 que se detectan en más de un sujeto). Hablando en general, el repertorio de cada sujeto es en gran parte "privado" y solo un pequeño porcentaje del repertorio inmunitario de ese sujeto representa CDR3 compartidas. Por lo tanto, los Linklets Públicos están presentes en un nivel mucho más bajo que los Linklets Privados, un hecho que dificulta la identificación de las firmas de enfermedad. Por lo tanto, es importante utilizar un enfoque que reduzca el fondo para permitir la identificación del repertorio significativo de CDR3 que constituye una o más firmas de enfermedad.

Como se usa en el presente documento, "muestra" comprende sangre y tejido. En determinadas realizaciones, la sangre es sangre periférica recogida de un sujeto. En determinadas realizaciones, el tejido es una biopsia obtenida a partir de un sujeto.

Como se usa en el presente documento, "sujeto" significa un ser humano o animal.

En determinadas realizaciones, el método divulgado actualmente reduce el fondo mediante el uso de "Linklets

Positivos". Cuando se secuencian dos secuencias de CDR3, A y B, también se obtiene información cuantitativa, representada por los recuentos de lecturas. Si el método de amplificación del repertorio inmunitario utilizado es semicuantitativo (arm-PCR), y si CDR3-A se expresa en una muestra a un nivel más alto que CDR3-B, se obtendrán más recuentos de lecturas de secuencia para A que para B. En tal escenario, el par A-B se designa como un Linklet Positivo, mientras que el par B-A se designaría como un Linklet Negativo. En determinadas realizaciones del método divulgado actualmente, solo se usan los Linklets Positivos para análisis adicional. El uso de Linklets Positivos enriquece la señal de diagnóstico, ya que ayuda a filtrar el ruido experimental.

Biológicamente, más de un antígeno o epítipo está asociado, generalmente, con una enfermedad en particular. Por lo tanto, los receptores de linfocitos T y B relevantes generalmente aparecerían en las muestras de los pacientes como grupos. La información cuantitativa proporcionada por los Linklets Positivos (y los Linklets Negativos) puede reflejar el perfil de expresión de antígenos específicos de la enfermedad.

Los procedimientos experimentales pueden introducir "ruido" adicional en los datos generados. Por ejemplo, es una práctica común agrupar muchas muestras de diferentes sujetos en una ejecución de secuenciación para reducir el coste (el repertorio inmunitario se amplifica por separado usando cebadores con código de barras). Sin embargo, si una CDR3 es muy dominante en una muestra, aparecerá en el chip de secuenciación varias veces. Ese clon dominante tendrá una probabilidad de 1/8.000 de ser asignado al código de barras incorrecto y ser "compartido" por todas las otras muestras en la misma ejecución. Si se usa la CDR3 como la unidad analítica básica, estas secuencias "contaminadas" se considerarían biológicamente compartidas y se usarían como señales de diagnóstico. El uso de Linklets permite que estos ruidos se filtren, ya que las CDR3 asignadas de forma incorrecta generalmente tienen una frecuencia muy baja, y la probabilidad de que formen parte de uno o más Linklets Positivos se reduce (con una mayor probabilidad de que formen Linklets negativos). Al considerar solo los Linklets Positivos, el ruido se puede filtrar. Además, si solo se usan las CDR3 mejor clasificadas de una muestra (tal como, por ejemplo, 5.000 o entre 1.000 y 50.000 de las CDR3 mejor clasificadas), no se considerarán, generalmente, las CDR3 asignadas de forma incorrecta, debido a sus bajas frecuencias.

Cuando se descubre que un grupo de Linklets Públicos está asociado en mayor grado con un grupo de sujetos que tienen una enfermedad en particular en común, esos Linklets Públicos se pueden tratar como Linklets específicos de la enfermedad, o "Linklets significativos". Por lo tanto, un grupo de Linklets Significativos asociados con una enfermedad en particular puede constituir una "firma de la enfermedad". Por lo tanto, si se encuentra que la muestra de un sujeto tiene una superposición estadísticamente significativa con la firma de la enfermedad, se puede hacer un diagnóstico de dicha enfermedad para ese sujeto.

El método actualmente divulgado comprende las siguientes etapas: (1) recoger muestras de sujetos asignados a un grupo de pacientes y un grupo de control; (2) amplificar y secuenciar un repertorio inmunitario para cada muestra; (3) identificar las secuencias de CDR3 exclusivas del repertorio inmunitario de cada muestra; (4) contar el número de veces que se detecta una secuencia de CDR3 individual (exclusiva) en el repertorio inmunitario, identificando de ese modo aquellos clones que son dominantes (determinado clasificándolos en orden de mayor frecuencia de aparición a menor frecuencia de aparición); (5) comparar los repertorios inmunitarios de los sujetos para identificar CDR3 que se comparten entre al menos dos sujetos ("CDR3 públicas"); (6) clasificar las CDR3 públicas en función de sus frecuencias de aparición; (7) generar una lista de Linklets Positivos a partir de las CDR3 mejor clasificadas; (8) filtrar los Linklets Privados y conservar los Linklets Públicos; y (9) identificar Linklets Públicos que están asociados con pacientes en el grupo de enfermedad diana, pero no con el grupo de control.

En determinadas realizaciones, se asignan al menos aproximadamente 100 sujetos a cada grupo. En determinadas realizaciones, el repertorio inmunitario se amplifica utilizando el método de arm-PCR (descrito en el documento WO2009/124293). En determinadas realizaciones, los 5.000 clones principales se identifican como dominantes. En determinadas realizaciones, la lista de Linklets Positivos incluye las 1.000 a 20.000 CDR3 mejor clasificadas. En determinadas realizaciones, el valor de confianza para los Linklets Públicos asociados con el paciente en el grupo de enfermedad diana, pero no con el grupo de control, es $p < 0,05$.

La lista de Linklets Públicos que están asociados con pacientes (Linklets asociados con enfermedad o DSL), pero no con el grupo de control, constituye un grupo designado como "Linklets de Firma". En determinadas realizaciones, se puede obtener una firma analizando alrededor de 100 pacientes y un número igual de controles. Después se determina un valor de corte de Linklet de Firma de Enfermedad (DSL, por sus siglas en inglés). Las muestras desconocidas pueden analizarse mediante secuenciación seguida por recuento de los DSL. Si el número de DSL cumple o excede el valor de corte, se realiza un diagnóstico para una enfermedad en particular.

Por ejemplo, en determinadas realizaciones para obtener los datos de secuencia para el análisis, la sangre completa de un sujeto (p. ej., ser humano o animal, grupo de enfermedad o control (sano)) se trata con Ficoll® para extraer células mononucleares de sangre periférica o PBMC, para obtener la mayor concentración de linfocitos. Cada tipo de linfocito tiene un identificador específico llamado marcador CD, o grupo de marcadores de diferenciación, que está numerado. Por ejemplo, los linfocitos T citotóxicos tienen un marcador CD8, y los linfocitos T auxiliares tienen un marcador CD4. Las perlas magnéticas, que se han marcado con un marcador anti-CD específico, se pueden añadir a la suspensión celular. Después de aplicar la columna a un campo magnético, las células unidas a las perlas quedarán

atrapadas, o se seleccionarán positivamente, mientras permiten que fluyan los otros tipos de células. El flujo continuo, o la suspensión celular seleccionada negativamente, se puede usar para aislar adicionalmente otras poblaciones celulares en aplicaciones posteriores. En otras realizaciones, la muestra puede ser un tejido, que puede procesarse, usando métodos conocidos en la técnica, para aislar linfocitos.

Debido a que hay subpoblaciones dentro de determinadas poblaciones de linfocitos T (p. ej., los linfocitos T reguladores son una subpoblación de los linfocitos T auxiliares), si es necesario separar las subpoblaciones, se pueden añadir reactivos de liberación a las células unidas a las perlas CD4 para liberar la perla, de modo que otra perla magnética pueda unirse a la célula. En el caso de los linfocitos T reguladores, por ejemplo, se puede añadir una microperla de selección CD25+ a la suspensión celular para extraer la población de linfocitos T reguladores de la población de linfocitos T auxiliares.

El aislamiento de polinucleótidos (ARN o ADN) se puede realizar por medios conocidos por los expertos en la materia (véase, por ejemplo, Murray, BMC Res Notes. 1 de Noviembre de 2013;6:440). La amplificación de secuencias puede realizarse usando el método descrito en el documento WO2009/124293 (arm-PCR), que proporciona la sensibilidad y especificidad que es necesaria para lograr resultados superiores en el método divulgado actualmente.

La secuenciación también puede realizarse usando métodos conocidos en la técnica. Dado el número de secuencias que deben determinarse, generalmente se emplean métodos de secuenciación de alto rendimiento, tal como, por ejemplo, la Secuenciación de Nueva Generación de Illumina, usando cebadores de secuenciación Illumina.

Se generan grandes cantidades de datos que deben analizarse y manipularse como resultado de la secuenciación, contando el número de veces que aparece una secuencia particular (que representa un clon individual), clasificando los clones en orden según la frecuencia de aparición y otros análisis descritos en el presente documento. Esto se realiza de manera más conveniente y eficaz utilizando programas de análisis de datos de secuencias. Uno de estos programas es Álgebra CDR3, que realiza la ordenación, clasificación y emparejamiento del investigador. El análisis estadístico, como el cálculo de los valores de p, también se puede realizar utilizando dichos programas.

El método actualmente divulgado se presta al desarrollo de pruebas de diagnóstico para una variedad de enfermedades que incluyen, pero sin limitación, cáncer, enfermedad autoinmunitaria, infecciones bacterianas, infecciones víricas e infecciones fúngicas, brindando, así, a los investigadores y médicos una herramienta valiosa para el diagnóstico y estudio de una enfermedad de interés.

El método actualmente divulgado puede describirse adicionalmente por medio de los siguientes ejemplos no limitantes.

Ejemplos

Aislamiento de Células Mononucleares de Sangre Periférica (PB-MC) de Sangre Completa

La sangre completa de sujetos sanos (grupo de control) y de pacientes diagnosticados previamente con cáncer de mama (grupo de pacientes) se diluyó con tampón PBS a 2-4 veces el volumen original. Se transfirieron 10 ml de sangre completa recogida en heparina de sodio a un tubo cónico de 50 ml y se diluyó con tampón a la línea de 35 ml. La suspensión celular diluida (35 ml) se colocó cuidadosamente sobre 15 ml de Ficoll-Paque® en un tubo cónico separado de 50 ml. El tubo se centrifugó a 400 x g durante 30 minutos a 20 grados Celsius en un rotor de cubeta oscilante sin freno.

La capa superior que contenía tampón PBS y plasma se aspiró cuidadosamente para eliminarla. La capa de células mononucleares turbia se transfirió cuidadosamente a un tubo cónico nuevo de 50 ml. Después se llenó el tubo con tampón hasta la marca de 50 ml y se centrifugó a 300 x g durante 20 minutos a 20 grados Celsius. Se retiró el sobrenadante transparente y se volvió a suspender el sedimento celular en 8 ml de tampón.

Aislamiento de Monocitos a partir de PBMC aisladas

Las células se contaron usando un hemocitómetro y la muestra se centrifugó a 300 x g durante 10 minutos a temperatura ambiente. El sobrenadante se eliminó por aspiración. Las células se resuspendieron en 80 µl de tampón por 10^7 células.

Se añadieron veinte microlitros de microperlas CD14 por 1×10^7 células, y la mezcla se realizó pipeteando suavemente hacia arriba y hacia abajo. La mezcla microperla/célula se incubó a 4 °C durante 15 minutos. Las células se lavaron añadiendo 2 ml de tampón por 1×10^7 células y después se centrifugaron a 300 x g durante 10 minutos.

El sobrenadante se aspiró completamente y se resuspendió en tampón (10^8 células en 500 µl de tampón). Se colocó una columna magnética LS sobre el imán y se lavó con 3 ml de tampón. Se descartó el tampón de flujo continuo. La suspensión celular se aplicó a la columna y las células no marcadas que atraviesan se recogieron en un tubo cónico de 15 ml marcado. La columna se lavó 3 veces con 3 ml de tampón, y se añadió un nuevo tampón solo cuando el depósito de la columna estaba vacío.

Se colocó un nuevo tubo cónico limpio de 15 ml etiquetado como "Monocito" debajo de la columna y la columna se retiró del imán. El tampón (5 ml) se pipeteó en la columna y las células marcadas magnéticamente se expulsaron inmediatamente empujando firmemente el émbolo hacia la columna.

5 Ambos tubos se centrifugaron durante 10 minutos a 300 x g, y el sobrenadante se aspiró por completo. Para el tubo etiquetado como "Monocito", las células se resuspendieron en 2 ml de tampón. Se pipetearon veinte microlitros para utilizarlos en el protocolo de recuento celular, y el tubo se centrifugó a 300 x g durante 10 minutos. Las células se resuspendieron en 500 µl de RNAprotect® y se almacenaron a 4 °C para su posterior extracción de ARN. Para el tubo etiquetado como "CD14 -", las células se resuspendieron en 80 µl de tampón por 10⁷ células.

Extracción de ARN

15 Las células se centrifugaron durante 3 minutos a 3.000 rpm a 20 °C, se retiró el sobrenadante y se soltó el sedimento celular moviendo el tubo. Se añadió tampón BME (350 µl) a la muestra, y el sedimento celular se disolvió completamente mediante agitación con formación de vórtice.

20 La muestra se transfirió a una columna QIAshredder y se homogeneizó por centrifugación durante 2 minutos a 10.000 rpm. Se descartó la columna y se guardó el flujo continuo. Se añadió etanol (70 %, 350 µl) al flujo continuo y la muestra se mezcló por pipeteo. La muestra (700 µl) se transfirió a una columna de centrifugación RNeasy® y se colocó en un tubo de recogida de 2 ml. La muestra se centrifugó durante 15 segundos a 10.000 rpm. Se descartó el flujo continuo. En los casos en que había más de 700 µl de muestra, esta etapa se repitió usando la misma columna.

25 Se añadieron 700 µl de tampón RW1 a la columna de centrifugación y la muestra se centrifugó a 10.000 rpm durante 15 segundos, descartando el flujo continuo. Se añadieron 500 µl de tampón RPE a la columna de centrifugación y la muestra se centrifugó a 10.000 rpm durante 15 segundos, descartando el flujo continuo.

30 Se añadieron 500 µl de tampón RPE a la columna de centrifugación y la muestra se centrifugó durante 2 minutos a 10.000 rpm. La columna de centrifugación se colocó en un nuevo tubo de recogida de 2 ml y se centrifugó durante 1 minuto a 10.000 rpm para secar la membrana de la columna. La columna de centrifugación se colocó en un nuevo tubo de recogida de 1,5 ml. Se añadieron 25 µl de agua sin RNasa a todas las muestras, excepto a las muestras que contienen linfocitos T reguladores aislados. A las muestras con linfocitos T reguladores, se añadieron 20 µl de agua sin RNasa. La muestra se dejó reposar a temperatura ambiente durante 1 minuto. La muestra se centrifugó durante 1 minuto a 10.000 rpm y la columna se descartó.

35 *Amplificación de secuencias de CDR3 Utilizando la Reacción en Cadena de la Polimerasa*

40 La amplificación por PCR de las secuencias de CDR3 se realizó usando el método de arm-PCR divulgado en el documento WO2009/124293 (Han). Generalmente se recomienda un mínimo de 100 ng de ARN o ADN_g (dependiendo del sistema de reactivo seleccionado) con un 260/280 de 1,8 o mayor como material de partida para obtener la mejor diversidad de la biblioteca del repertorio inmunitario por arm-PCR. Durante la primera ronda de PCR, se usaron cebadores específicos de genes anidados dirigidos a cada uno de los genes V y J (o C). Los cebadores directos, F_o (directo-hacia fuera) y F_i (directo-hacia dentro), se dirigieron a los genes V. Los cebadores inversos, R_o (inverso-hacia fuera) y R_i (inverso-hacia dentro), se dirigieron a cada uno de los genes J o C. Los cebadores F_i y R_i también incluyeron adaptadores de secuenciación B y A, respectivamente, para las plataformas Illumina® (HiSeq, MiSeq y GAllx) para la secuenciación de extremos emparejados. La segunda ronda de PCR se llevó a cabo utilizando los cebadores comunales (comunes) B y A. Después de la purificación en gel, el producto resultante estaba listo para la secuenciación de alto rendimiento con las plataformas Illumina®. La primera ronda de PCR introdujo códigos de barras y cebadores de secuenciación en los productos de PCR.

50 La fase exponencial de la amplificación se logró mediante los cebadores comunales en la segunda ronda de PCR; por lo tanto, el repertorio inmunitario diana se amplificó de manera uniforme y semicuantitativa, sin introducir sesgos de amplificación adicionales.

55 *Identificación de la Firma del Cáncer de Mama*

Se recogieron un total de 213 muestras, incluidas 103 de pacientes con cáncer de mama y 110 de controles. Se identificaron un total de 14.666.172 CDR3 de las 213 muestras, con un promedio de 68.855 CDR3 de cada muestra. Se encontraron 8.301.648 CDR3 exclusivas de las 213 muestras. Después de eliminar las CDR3 privadas, se identificaron un total de 98.076 CDR3 públicas (es decir, compartidas por al menos dos sujetos) y dominantes (es decir, clasificadas entre las primeras 5.000 CDR3 en cada muestra) de las 213 muestras, utilizando el software iRepertoire (Huntsville, Alabama, EE. UU.) disponible a través del sitio web de la empresa (por ejemplo, Álgebra CDR3). Se generaron un total de 287.198.206 Linklets Positivos a partir de las 213 muestras, con un promedio de 1.003.236 Linklets de cada muestra. Después de eliminar los Linklets Privados, quedaron 16.921.605 Linklets que se compartieron con al menos otra persona. Para cada Linklet compartido, se obtuvo un valor de p para identificar aquellos compartidos preferentemente entre los pacientes. Se identificó un total de 117.069 Linklets como Linklets

5 Significativos con $p < 0,05$, proporcionando una "firma" para las enfermedades. Un total de 6.171 CDR3 contribuyeron a 117.069 Linklets de firma de la enfermedad. Usando un valor de corte de 600 Linklets Significativos, se pudo diagnosticar el 95 % de los cánceres de mama, con un 93,6 % de especificidad. Cuando se estudiaron 188 muestras de cáncer no de mama, solo tres muestras fueron positivos falsos (con más de 600 DSL), dando una especificidad del 98,4 %.

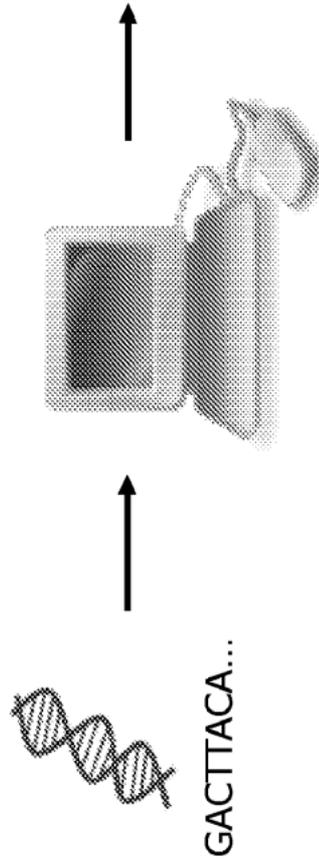
10 El método divulgado actualmente aumenta la señal y reduce el fondo para permitir la identificación de CDR3 compartidas que pueden usarse para producir una firma de enfermedad, que no es posible, de otra manera, utilizando métodos convencionales. Como resultado, el método actualmente divulgado tiene el beneficio de permitir el desarrollo de una prueba de diagnóstico para diferentes enfermedades que incluyen, pero sin limitación, cáncer, enfermedad autoinmunitaria, enfermedad inflamatoria y enfermedad infecciosa.

Las metodologías y las diversas realizaciones de las mismas descritas en el presente documento son ejemplares. Son posibles otras realizaciones diferentes de las metodologías descritas en el presente documento.

REIVINDICACIONES

- 5 1. Un método para desarrollar una prueba de diagnóstico para una enfermedad en particular, comprendiendo el método las etapas de:
- 10 en una muestra recogida de cada uno de los múltiples sujetos en un grupo de pacientes y un grupo de control, en donde el grupo de pacientes comprende sujetos que tienen la misma enfermedad y el grupo de control comprende sujetos que están sanos;
- 15 amplificar y secuenciar el repertorio inmunitario de cada sujeto en cada uno de los dos grupos para identificar cada secuencia de CDR3 exclusiva presente en las muestras y determinar la frecuencia de aparición de cada secuencia de CDR3 exclusiva;
- identificar secuencias de CDR3 que se comparten entre al menos dos sujetos en cada uno de los grupos de control y el grupo de pacientes;
- clasificar las secuencias de CDR3 por orden de frecuencia de aparición;
- identificar los Linklets de cada grupo; e
- identificar los Linklets que están asociados en un grado estadísticamente significativo con el grupo de pacientes para proporcionar una firma de la enfermedad.
- 20 2. El método de la reivindicación 1, en donde la muestra es sangre.
3. El método de la reivindicación 1, en donde la muestra es tejido.
4. El método de la reivindicación 1, en donde se identifican al menos aproximadamente 1.000 secuencias de CDR3 que se comparten entre al menos dos sujetos en cada uno del grupo de control y del grupo de pacientes.
- 25 5. El método de la reivindicación 1, en donde se identifican al menos aproximadamente 10^6 Linklets de cada grupo.
6. El método de la reivindicación 1, en donde se establece un número mínimo de Linklets asociados con la firma de la enfermedad como un número de diagnóstico, de modo que cuando al menos se identifica ese número de Linklets en la muestra de sangre de un sujeto, se diagnostica que el sujeto tiene la enfermedad.
- 30 7. El método de la reivindicación 6, en donde el número de diagnóstico es al menos aproximadamente 500.
- 35 8. Un método para desarrollar una prueba de diagnóstico para una enfermedad en particular, comprendiendo el método las etapas de:
- 40 en una muestra de sangre recogida de cada uno de los múltiples sujetos en un grupo de pacientes y un grupo de control, en donde el grupo de pacientes comprende sujetos que tienen la misma enfermedad y el grupo de control comprende sujetos que se clasifican como sanos;
- 45 amplificar y secuenciar el repertorio inmunitario de cada sujeto en cada uno de los dos grupos para identificar cada secuencia de CDR3 exclusiva presente en las muestras de sangre y determinar la frecuencia de aparición de cada secuencia de CDR3 exclusiva;
- identificar al menos aproximadamente 1.000 secuencias de CDR3 que se comparten entre al menos dos sujetos en cada uno del grupo de control y del grupo de pacientes;
- clasificar las al menos aproximadamente 1.000 secuencias de CDR3 por orden de frecuencia de aparición;
- identificar al menos aproximadamente 10^6 Linklets de cada grupo; e
- identificar los Linklets que están asociados en un grado estadísticamente significativo con el grupo de pacientes para proporcionar una firma de la enfermedad;
- 50 en donde se establece un número mínimo de Linklets asociados con la firma de la enfermedad como un número de diagnóstico, de modo que cuando al menos se identifica ese número de Linklets en la muestra de sangre de un sujeto, se diagnostica que el sujeto tiene la enfermedad.

Fig 1



Estudio BC	Linkets sig	Estudio BC	Linkets sig
6852_	33189	7472_	4246
6845_	30523	7478_	4219
6245_	30187	8217_	3966
6244_	29807	7479_	3900
6853_	28703	7464_	3867
6246_	27629	8202_	3774
6242_	26762	7461_	3583
6241_	26642	6804_	3544
6849_	25160	6859_	3510
6844_	24202	8194_	3236
6851_	17490	7467_	2994
6847_	16968	8186_	2726
6846_	15854	6815_	2623
6240_	15493	7463_	2604
6861_	13208	8176_	2272
6848_	10937	6843_	2159
7441_	10048	6814_	1890
7448_	9581	8209_	1860
7454_	9049	8198_	1797
7451_	8749	8200_	1785
7449_	8738	6809_	1660
7456_	8602	6821_	1310
6858_	8404	6819_	1268

Fig 2

Clasificación	CDR3	Recuentos de lecturas
1	ASSEDRDQETQY	40303
2	ASSEAGSTNQPH	33353
3	ATSRDRTSDSYNEQF	16597
4	ASSLFQGGSGKLF	15063
5	ASSSTGTARRLNTEAF	14275
6	ASTEPSPNQRGVGETQY	12193
7	SARDGSGDADTQY	11974
8	ASSQDQVSNYGYT	6939
9	SASANPGQGAGNTEAF	6674
10	ASSSGPNTEAF	5633
11	SARDRDTEAF	5585
12	ASSPSGGASDTQY	5270
13	ASSPAVDSGANVLT	4954
14	ASSFTGQVYNEQF	4698
15	ASSPGPEAF	4693
16	SASLSDKGGQNQPQH	4638
17	ASSDLAIETQY	4482
18	ASSLSSGFRPVETQY	4265
19	ASSLGGREY	4145
20	ASRSRLEETQY	3895
21	ASSIWSDTDQY	3774
22	ASSLSAGVDTIY	3772
23	SARDLSGGDTEAF	3736
24	SARQGDTEAF	3702
25	ASSEAPFNRRVSSNQPH	3660

Fig 3

Clasificación	CDR3	Recuentos de lecturas
1	ASSEDRDQETQY	40303
2	ASSEAGSTNQPQH	33353
3	ATSRDRTSDSYNEQF	16597
4	ASSLFQGGSGKLF	15063
5	ASSSTGTARRLNTEAF	14275

Fig 4

c9968_C

(‘ASSYGLAGYEY’, ‘ASSDSSGSYNEQF’) 1
 (‘ASSYGLAGYEY’, ‘ASSPQGARTDTQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSDGASGNTIY’) 1
 (‘ASSYGLAGYEY’, ‘ASSSTSYEQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSLSGGADTQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSLWGATEAF’) 1
 (‘ASSYGLAGYEY’, ‘ASSQTGGTEAF’) 1
 (‘ASSYGLAGYEY’, ‘ASSYSYEQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSYPPSGIYNEQF’) 1
 (‘ASSYGLAGYEY’, ‘SARAGQVNPQH’) 1
 (‘ASSYGLAGYEY’, ‘ASSPTDTQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSSPVLNTEAF’) 1
 (‘ASSYGLAGYEY’, ‘ASSYSGDEQF’) 1
 (‘ASSYGLAGYEY’, ‘ASSQDSSGANVLT’) 1
 (‘ASSYGLAGYEY’, ‘ASSWRAGGTDQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSQYSNPQH’) 1
 (‘ASSYGLAGYEY’, ‘ASSYRLYNEQF’) 1
 (‘ASSYGLAGYEY’, ‘ASWAGGQPQH’) 1
 (‘ASSYGLAGYEY’, ‘ASSYTDYNEQF’) 1
 (‘ASSYGLAGYEY’, ‘ASSYREGQPQH’) 1
 (‘ASSYGLAGYEY’, ‘ASSLGRRRTEAF’) 1
 (‘ASSYGLAGYEY’, ‘ASSPRTSGSTGELF’) 1
 (‘ASSYGLAGYEY’, ‘ASSPGNYEQY’) 1
 (‘ASSYGLAGYEY’, ‘SASRQGVNSPLH’) 1
 (‘ASSYGLAGYEY’, ‘ASSTGGSGNPQH’) 1
 (‘ASSYGLAGYEY’, ‘ASSFRGGYEY’) 1
 (‘ASSYGLAGYEY’, ‘ASSEAGGTGELF’) 1
 (‘ASSYGLAGYEY’, ‘ASSDGSNTEAF’) 1
 (‘ASSYGLAGYEY’, ‘SAWDSSTDTQY’) 1
 (‘ASSYGLAGYEY’, ‘ASSYATNTEAF’) 1

Fig 5

	Valores de p	Valores de p
('ASSYSRGEEF', 'ASSLGRTHQPQH')	0	('ASSLQVNTEAF', 'ASSLGGQPQH')
('ASSLGTSAEQY', 'ASSLTGETQY')	0	('ASSLSRGLLNEQF', 'ASSLDRGNQPQH')
('ATSSHGGETQY', 'ASSLSGGSPDTQY')	0	('ASSLEALMNTTEAF', 'ASSYSTTGVGGELF')
('ASSYSRGEEF', 'SARDLAGDRWNEQF')	0	('ASSFFRGYEQY', 'ASSLSVNTEAF')
('ASSLEAGTNTTEAF', 'ATRQDITDTQY')	0	('ASSLEALMNTTEAF', 'ASSLGNQPQH')
('ASSLTGETQY', 'ASSLGRTHHQPQH')	0	('ASSYRENTTEAF', 'ASSSTTDTQY')
('ASNRQGSTTEAF', 'ASSYKTRRNEQF')	0	('ATSRGNQPQH', 'ASRGSTDTQY')
('ASSPTSGANNEQF', 'ASTDNYGYT')	0	('ASSNTDTQY', 'ASSYSRGEEF')
('ASSLSRPDTQY', 'ASSLSGGSPDTQY')	0	('ASSYSGGDEQY', 'ASSLAGSTDTQY')
('ASSFQGTTEAF', 'ASSFGETQY')	0	('ASSSYTDTQY', 'ASSLAGDTQY')
('ASNRQGSTTEAF', 'ATSREANTTEAF')	0	('ASSYSRGEEF', 'ASSYSADTQY')
('ASSDDSYNSPLH', 'ASSLTGETQY')	0	('ASSFLAGTDTQY', 'ASSPQGGTEAF')
('ASCRDRAYNEQY', 'ASSQARSSGGAQDTQY')	0	('ASSYSRGEEF', 'ASSELENTEAF')
('ASSPRDRTYEQY', 'ASSLGRTHQPQH')	0	('ASGTREPTDTQY', 'ASSLGLTDTQY')
('ATSSGTSGIQY', 'ASSLGDYQY')	0	('ASSYKTRRNEQF', 'SARDISGGNQPQH')
('ASSLSRPDTQY', 'ASSPRDRTYEQY')	0	('ATSRLAGETQY', 'ASSYSSNQPQH')
('ASSYSRGEEF', 'ATSSGTSGIQY')	0	('ASSYSTGYEQY', 'ASSPSYEQY')
('ASSLSRGLLNEQF', 'ASSLGETQY')	0	('ASGRNTQY', 'ASSLGTTEAF')
('ASMDIAETQY', 'ASSSTDTQY')	0	('ASCRDRAYNEQF', 'ASSYSETQY')
('ASSLGTSAEQY', 'ASSQSSGAYNEQF')	0	('ASSLTDYQY', 'SARGQNEQF')
('ASSYGRVTEAF', 'ASTQWTSGLGEQF')	0	('ASSPRDRTYEQY', 'ASSLVNQPQH')
('ASCRDRAYNEQF', 'ASSSTDTQY')	0	('ASSLGGNQPQH', 'ASRTGGQPQH')
('ASSLGGDNYGYT', 'ATSREANTTEAF')	0	('ASSLYNEQF', 'ASSSDLTDTQY')
('ASSPTSGANNEQF', 'ASGRMKKTQY')	0	('ASSFQGTTEAF', 'ASSFGTSYEQY')
('ASSLGTSAEQY', 'ASSLGYEQY')	0	('ASSFSGSTDTQY', 'ASSLADTQY')
('ATSSHGGETQY', 'ANGDRLYTEAF')	0	('ASSLGGNQPQH', 'ASSPNTTEAF')
('ASCRDRAYNEQF', 'ASSLATSDTQY')	0	('ASSDDSYNSPLH', 'ASSLGNTEAF')
('ATSSGTSGIQY', 'ASSFLIETGMDNEQY')	0	('ASSFNSDRDITDTQY', 'ASSLGGTEAF')
('ASTLGLLEQY', 'ASSPRDRTYEQY')	0	('ASSFNSDRDITDTQY', 'ASSLGRNTEAF')
('ATSSTAGETQY', 'SARDLAGDRWNEQF')	0	('ASSDDSYNSPLH', 'ASSLSRNTTEAF')

Fig 6

