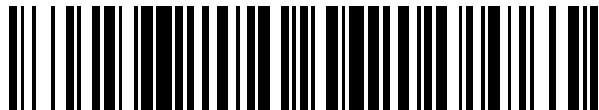


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 659 492**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.11.2014** E 14191967 (0)

97 Fecha y número de publicación de la concesión europea: **13.12.2017** EP 3018214

54 Título: **Método para cuantificar el nivel de enfermedad mínima residual en un sujeto**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
15.03.2018

73 Titular/es:

**FUNDACION DE INVESTIGACION HOSPITAL 12
DE OCTUBRE (100.0%)
Avda. de Cordoba s/n edificio CAA
28041 Madrid, ES**

72 Inventor/es:

**BARRIO GARCÍA, SANTIAGO;
MARTÍNEZ LÓPEZ, JOAQUÍN;
MARÍN SEBASTIÁN, CARLOS;
RAPADO MARTÍNEZ, MARÍA INMUACULADA;
AYALA DÍAZ, ROSA MARÍA y
SÁNCHEZ VEGA CARRIÓN, BEATRIZ**

74 Agente/Representante:

ILLESCAS TABOADA, Manuel

ES 2 659 492 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 cuantificar el nivel de enfermedad mínima residual en un sujeto

5 **Campo de la invención**

La presente invención puede incluirse en el campo de la medicina en general, más particularmente en el campo del diagnóstico de enfermedades.

10 En particular, la presente invención se centra en un método y un kit para cuantificar el nivel de enfermedad mínima residual en un sujeto. Además, la presente invención se centra en el uso del método y/o kit para cuantificar el nivel de enfermedad mínima residual en un sujeto.

15 **Antecedentes de la invención**

15 Los métodos actuales para la detección y tratamiento de una enfermedad indican que es posible controlar muchas enfermedades en un nivel clínico, borrando, de este modo, cualquier huella de la enfermedad. No obstante, es posible que algunas enfermedades no sean detectadas o que la recuperación del sujeto no sea completa después del tratamiento. En este último caso, la enfermedad puede desarrollar una resistencia a los fármacos bajo presión selectiva del tratamiento mediante un proceso de selección clonal, permitiendo, de este modo, la expresión y en última instancia la reincidencia o recaída en la enfermedad. Es, por lo tanto, muy importante controlar la cantidad de células muertas en un tejido dado. En particular, resulta especialmente beneficioso controlar el nivel de enfermedad mínima residual (EMR) en sujetos que han sido tratados para una enfermedad. La EMR es el nombre que se proporciona a la enfermedad que permanece en un sujeto o en un tejido en particular del mismo durante o después del tratamiento de dicha enfermedad. Típicamente, la EMR se refiere al cáncer y/o leucemia que permanece en un sujeto durante o después del tratamiento de dicho cáncer y/o leucemia mediante, por ejemplo, quimioterapia.

30 Actualmente el control a través de, por ejemplo, la determinación de la EMR puede realizarse mediante distintas técnicas. En primer lugar, la citometría de flujo (CMF) puede usar hasta 8 marcadores distintos para determinar el fenotipo de la enfermedad. En segundo lugar, otro método que se usa para este fin es la PCR de oligonucleótidos específicos de alelo (ASO-PCR) de genes de inmunoglobulina (Igs), que requiere el diseño de cebadores específicos para cada paciente o un marcador molecular específico y es aplicable únicamente al 40 % de los casos. A este respecto, el documento WO2004033728 A2 detalla cebadores para la amplificación de reordenamientos de inmunoglobulinas para el diagnóstico de síndromes linfoproliferativos. La amplificación mediante PCR de inmunoglobulinas reordenadas o de genes reordenados también se ha usado en el documento US 7785783 B2 para caracterizar regiones de ácido nucleico basándose en la identificación de regiones que flanquean una región de ácido nucleico marcadora. Esto proporciona medios de análisis de un marcador que es característico de una población clonal de células y para controlar la progresión de una afección.

40 Finalmente, la aparición de tecnologías de secuenciación de nueva generación (SNG) ha hecho posible analizar millones de secuencias a la vez, cada una proviniendo de distintos fragmentos de genoma, o el resultado de la amplificación de la misma región, representando una gran cantidad de células. En consecuencia, la tecnología de SNG puede usarse para identificar una enfermedad y/o cuantificar el nivel de una enfermedad en sujetos. A este respecto, una profunda secuenciación se ha revelado como una metodología alternativa a las técnicas anteriormente mencionadas, proporcionando una aplicabilidad superior que la ASO-PCR y una sensibilidad superior que la CMF [cf. Martínez-López J. y col. "Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma"; *Blood* (2014) 123(20):3073-3079]. Esta técnica también ha demostrado ser útil para la detección y cuantificación de células de mieloma tanto en médula ósea como en sangre periférica [cf. Vij R. y col. "Deep sequencing reveals myeloma cells in peripheral blood in majority of multiple myeloma patients"; *J. Clin. Lymphoma Myeloma Leuk.* (2013) 14(2): 131-139].

50 Además, la tecnología de SNG ha hecho posible la identificación de linfocitos T o B clonogénicos con alta sensibilidad y especificidad, permitiendo, de este modo, la detección de enfermedad mínima residual en condiciones tales como leucemia linfocítica aguda (LLA), linfoma de células del manto (LCM) y mieloma múltiple (MM) [cf. Ladetto M. et al. "Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders"; *Leukemia* (2014) 28(6):1299-1307]. La SNG también ha demostrado ser un método útil para la identificación de perfiles clonotípicos para detectar y controlar una enfermedad a partir de una muestra de linfocitos. El documento US 8628927 B2 demuestra que la EMR de LLA puede establecerse mediante la comparación de la secuencia de una muestra con perfiles de clonotipos previamente obtenidos y marcadores sobre células leucémicas. Además, otro enfoque para la detección de EMR es la determinación de un perfil de clonotipos mediante el uso de etiquetas de secuencia para producir perfiles basados en secuencias de poblaciones de ácido nucleico complejas, tal como se divulga en el documento WO 2013188471 A2. Faham y col. (*Blood* (2012) 120(26) 5173-5180) describieron un enfoque de secuenciación profunda para la detección de enfermedad mínima residual en LLA que se basa en la adición de una cantidad conocida de IgH de referencia. Aunque los documentos anteriormente mencionados describen varios enfoques distintos para detectar EMR en donde se logran mejoras mediante distintos métodos de secuenciación, ninguno de estos documentos desvela específicamente un método que logre una mejora mediante la estrategia de alineación usada. El método de alineación usado es importante para la precisión de

5 cualquier método que se base en la comparación de secuencias nucleotídicas puesto que la tasa de fallos de
 secuenciadores que usan una lógica binaria clásica - en las que las secuencias solo pueden ser iguales o distintas -
 es tan alta que no resulta útil. De hecho, una elevada proporción (casi todas) de secuencias de nucleótidos que
 10 evalúan como distintas, son iguales pero aparecen como distintas debido a un error en el secuenciador. A este
 respecto, el documento US 8628927 B2 menciona que algunas secuencias son más complicadas de alinear que
 otras debido a mutaciones somáticas y diversas regiones [tales como la región NDN entre la variable (V) y la unión
 (J) de los segmentos génicos en linfocitos]. Sin embargo, a pesar de describir de forma general que la alineación
 puede lograrse usando secuencias de referencias tales como secuencias de unión de cebadores o secuencias sin
 referencias, este documento no describe un método específico de alineación que sea capaz de determinar el nivel
 de enfermedad en un sujeto independientemente de las características genéticas del nucleótido o de la enfermedad.

15 El problema de la presente invención es proporcionar un método para la cuantificación del nivel de enfermedad en
 un paciente, en el que dicho método muestra una sensibilidad mejorada, una reproducibilidad superior y una
 determinación más precisa de dichos niveles. Es un problema adicional de la presente invención proporcionar un
 método universal que sea capaz de determinar el nivel de enfermedad en un sujeto independientemente de las
 características genéticas del nucleótido o la enfermedad. Además, es un problema de la presente invención
 proporcionar un método que sea específico para un sujeto y no requiera acceder a bases de datos externas que
 comprendan datos obtenidos a partir de poblaciones de sujetos. Por otra parte, es un problema de la presente
 20 invención proporcionar un método para la cuantificación de secuencias clonotípicas tumorales para reordenamientos
 génicos de inmunoglobulina, así como carga alélica, mutaciones puntuales (SNV), mutaciones múltiples (MNV),
 indeles, inserciones largas y translocaciones.

Breve descripción de la invención

25 La presente invención describe un método para cuantificar la enfermedad mínima residual (EMR) en un sujeto que
 ha sido tratado para dicha enfermedad, que comprende:

(a)

- 30 - amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de
 nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de
 dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso
 promedio, k, por célula diploide de dicha muestra biológica; y
 35 - secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de
 caracteres de lectura de izquierda a derecha;

(b)

- 40 - amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al
 menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del
 tratamiento de dicha enfermedad; y
 - secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de
 caracteres de lectura de izquierda a derecha;

45 (c) determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada
 segunda lista de caracteres obtenida en la etapa (b), en el que un grado de similitud, DS, de una primera lista de
 caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina
 bien mediante:

- 50 (i) selección del carácter o la secuencia continua más larga de caracteres que son los mismos en la primera
 lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más
 caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter
 o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;
 55 (ii) exclusión del carácter o la secuencia continua más larga de caracteres seleccionados en la etapa (i) de
 etapas posteriores de selección un carácter o secuencia continua más larga de caracteres que son los
 mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- 60 - selección del carácter o la secuencia continua más larga de caracteres que se encuentra a la izquierda del
 carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la
 primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o
 más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el
 carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de
 caracteres; y
 65 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del
 carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la

primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

5 (iv) exclusión cada carácter y/o cada secuencia continua más larga de caracteres seleccionado en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(v)

- 10 - selección del carácter o la secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 15 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 20

25 (vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionado en la etapa (v) de etapas posteriores de selección de un carácter o de secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que es la misma en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(viii) sumar

- 30 - el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
- 35 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son las mismas que en la segunda y en primera lista de caracteres, respectivamente;

(ix) sumar

- 40 - C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
- 45 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y en segunda lista de caracteres; y

(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

o mediante:

55 (xi) selección del carácter o de la secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

60 (xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y la segunda lista de caracteres;

(xiii)

- 65 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en

la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- 5 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 10

(xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

- 15 (xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;
- (xvi) sumar

- 20 - el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

25 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

(xvii) sumar

- 30 - C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
- 35 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y

(xviii) calcular el DS según la siguiente fórmula:

40
$$DS = C_c / C_t$$

- (d) seleccionar, para cada primera lista de caracteres obtenida en la etapa (a), el DS de valor más alto, DS_{HV} ;
- (e) sumar el número de primeras listas de caracteres que tienen un DS_{HV} que es superior que un valor de umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;
- 45 (f) sumar

- 50 - L_c ; y
- el número de primeras listas de caracteres que no tienen un DS_{HV} que es superior que T ,

para obtener el número total de primeras listas de caracteres, L_t ; y

(g) calcular el nivel de enfermedad mínima residual (EMRERM) según la siguiente fórmula:

55
$$EMR = L_c \times (D / k) / L_t^2.$$

La invención se refiere adicionalmente a un kit y su uso que se define en las reivindicaciones adjuntas 8-15. Además, la presente divulgación contiene un kit para cuantificar el nivel de enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad que comprende:

- 60 (a)
- medios para amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D , de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k , por célula diploide de dicha muestra biológica; y
- 65 - medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera

lista de caracteres de lectura de izquierda a derecha;

(b)

- 5 - medios para amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- 10 - medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de caracteres de lectura de izquierda a derecha;

(c) medios para determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), en el que un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

- 15 (i) selección del carácter o secuencia continua más larga de caracteres que son los mismos que en la primera lista de caracteres y en la segunda lista de caracteres donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;
- 20 (ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- (iii)
- 25 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 30 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 35 (iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- 40 (v)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 45 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 50 (vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- 55 (vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;
- 60 (viii) sumar
- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
- 65 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas

(i) a (vii)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

5 (ix) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

10

15

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

20

o mediante:

(xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos que en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

25

(xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

30

(xiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

35

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

40

(xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionado en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

45

(xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

50

(xvi) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

55

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

60

(xvii) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

65

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y (xviii) calcular el DS según la siguiente fórmula:

5
$$DS = C_c / C_t$$

(d) medios para seleccionar, para cada primera lista de caracteres obtenida en la etapa (a), el DS del valor más alto, DS_{HV} ;

10 (e) medios para sumar el número de primeras listas de caracteres que tienen un DS_{HV} superior a un valor de umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;

(f) medios para sumar

- 15 - L_c ; y
- el número de primeras listas de caracteres que no tienen un DS_{HV} superior a T ,

para obtener el número total de primeras listas de caracteres, L_t ; y

(g) medios para calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

20
$$EMR = L_c \times (D / k) / L_t^2.$$

Además, la presente invención describe un uso del método descrito en el presente documento o del kit descrito en el presente documento en la cuantificación de la enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad.

25 Por otra parte, la presente invención se refiere a un método para cuantificar la enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad, que comprende:

(a)

- 30 - amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D , de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k , por célula diploide de dicha muestra biológica; y
35 - secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres de lectura de izquierda a derecha;

(b)

- 40 - amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de caracteres de lectura de izquierda a derecha;

45 (c) determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), en el que un grado de similitud, DS , de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

50 (i) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

55 (ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- 60 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- 65 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la

primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

5 (iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

10 (v)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

15 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

20 (vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que es igual en la primera lista de caracteres y en la segunda lista de caracteres;

25 (vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que es igual en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

30 (viii) sumar
- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
35 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

40 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos que en la segunda y primera lista de caracteres, respectivamente;

(ix) sumar
45 - C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

50 para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

o mediante:

55 (xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos que en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

60 (xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

65 (xiii)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en

la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

5 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde, cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas
10 de caracteres;

(xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

15 (xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(xvi) sumar

20 - el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

25 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son iguales en la segunda y primera lista de caracteres, respectivamente;
(xvii) sumar

30 - C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que
35 no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(xviii) calcular el DS según la siguiente fórmula:

40
$$DS = C_c / C_t$$

(d)

45 - amplificar mediante reacción en cadena de polimerasa usando cebadores, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (a) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa para obtener al menos una primera lista complementaria inversa de caracteres que se leen de izquierda a derecha; y

50 - amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa previa, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (b) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa para obtener al menos una segunda lista complementaria inversa de caracteres que se leen de izquierda a derecha; y

55 - determinar, para cada primera lista complementaria inversa de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista complementaria inversa de caracteres obtenida en la etapa (b), en el que un grado de similitud, DS_{rcs} , de una primera lista complementaria inversa de caracteres obtenida en la etapa (a) con una segunda lista complementaria inversa de caracteres obtenida en la etapa (b) se determina bien mediante:

60 (xix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;

65 (xx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los

mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxi)

- 5 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
- 10 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;
- 15

(xxii) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

20

(xxiii)

- 25 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
- 30 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;
- 35

(xxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

40

(xxv) repetición de las etapas (xxii) y (xxiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;

45

(xxvi) sumar

- 50 - el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv) para obtener el número total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera lista complementarias inversas de caracteres, respectivamente;

55 (xxvii) sumar

- C_c ; y
- 60 - el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c),
- 65

ES 2 659 492 T3

para obtener el número total de caracteres, C_t , en la primera y segunda lista complementarias inversas de caracteres; y

(xxviii) calcular el DS según la siguiente fórmula:

$$5 \quad \quad \quad DS_{rcs} = C_c / C_t$$

o mediante:

10 (xxix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;

15 (xxx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xxix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxxi)

20 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y

25 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;

30 (xxxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

35 (xxxiii) repetición de las etapas (xxxii) y (xxii) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;

40 (xxxiv) sumar

- el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xxix) a (xxxiii); y

45 - el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xxix) a (xxxiii) para obtener el número total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera lista complementarias inversas de caracteres, respectivamente;

(xxxv) sumar

50 - C_c ; y

- el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xxix) a (xxxiii) de la etapa (c); y

55 - el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xxix) a (xxxiii) de la etapa (c),

60 para obtener el número total de caracteres, C_t , en la primera y segunda lista complementarias inversas de caracteres; y

(xxxvi) calcular el DS según la siguiente fórmula:

$$65 \quad \quad \quad DS_{rcs} = C_c / C_t$$

en el que el DS se determina para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (i) a (x), el DS_{rcs} se determina para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xix) a (xxviii) y cuando el DS está determinada para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (xi) a (xvii), el DS_{rcs} se determinada para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xxix) a (xxxvi); y seleccionando, para cada primera lista de caracteres obtenida en la etapa (a) y su correspondiente primera lista complementaria inversa de caracteres, el DS o DS_{rcs} de valor más alto, DS_{HV} :

(e) sumar el número de primeras listas de caracteres que tienen un DS_{HV} que es superior que un valor de umbral, T, para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;

(f) sumar

- L_c ; y

- el número de primeras listas de caracteres que no tienen un DS_{HV} superior a T,

para obtener el número total de primeras listas de caracteres, L_c ; y

(g) calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_c^2;$$

donde en las sub-etapas (iii) y (xiii) de la etapa (c) y las sub-etapas (xxi) y (xxxi) de la etapa (d) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres que se encuentra a la izquierda del carácter o de la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii) de la etapa (c) y sub-etapas (xx) y (xxx) de la etapa (d), respectivamente, y para la secuencia continua de caracteres que se encuentra a la derecha del carácter o de la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii) de la etapa y sub-etapas (xx) y (xxx) de la etapa (d), respectivamente; y

donde en la sub-etapa (v) de la etapa (c) y la sub-etapa (xxiii) de la etapa (d) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa, y para la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa.

Breve descripción de las figuras

Figura 1. Diagrama esquemático que representa las sub-etapas (i) a (vii) de la etapa (c) según la invención, en el que la estrecha línea gris (—) representa una primera lista de caracteres y la estrecha línea negra (-) represente una segunda lista de caracteres. La selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y la segunda lista de caracteres se representa mediante líneas más anchas () y (), respectivamente, que se excluyen posteriormente de las anteriormente mencionadas listas de caracteres.

Figura 2. Diagrama esquemático que representa las sub-etapas (xi) a (xv) de la etapa (c) según la invención, en el que la estrecha línea gris (—) representa una primera lista de caracteres y la estrecha línea negra (-) represente una segunda lista de caracteres. La selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y la segunda lista de caracteres se representa mediante líneas más anchas () y (), respectivamente, que se excluyen posteriormente de las anteriormente mencionadas listas de caracteres.

Figura 3. Sensibilidad del método de la presente invención en la caracterización de reordenamientos génicos de inmunoglobulinas en muestras a partir de pacientes con mieloma múltiple. El porcentaje de lecturas clonotípicas se mide mediante dilución en serie de reordenamientos clonales sobre un origen policlonal obtenido a partir de pacientes sanos o a partir de líneas no de linfocitos B, en la que la sensibilidad depende de la cantidad de variantes encontradas en una lectura clonotípica.

Figura 4. Correlación ($R^2 = 0,51$) entre los datos de la EMR medidos con citometría de flujo (CMF de EMR Log, eje y) y secuenciación paralela masiva (SNG de EMR Log, eje x) para reordenamientos principales en muestras de seguimiento de pacientes diagnosticados y tratados para mieloma múltiple.

Figura 5. Cuantificación de sensibilidad del método de la presente invención en la caracterización de mutaciones puntuales (DNM3A, IDH2, FLT3 y NMP1) en pacientes con leucemia mieloide aguda (LMA). El porcentaje de lecturas clonotípicas se mide en diluciones en serie de genes mutados con LMA.

Descripción detallada de la invención

La presente invención se refiere a un método para cuantificar el nivel de enfermedad. En particular, la presente invención se refiere a un método para cuantificar el nivel de enfermedad mínima residual (EMR). La EMR es el nombre que se proporciona a la enfermedad que permanece en un sujeto después del tratamiento de dicha enfermedad. Por lo tanto, cuantificar el nivel de EMR significa cuantificar la cantidad de células muertas en un sujeto o cuantificar la cantidad de material genético que está asociado con la enfermedad en un sujeto tras el tratamiento de dicha enfermedad. Preferentemente, cuantificar el nivel de EMR significa cuantificar el número de células muertas en una muestra biológica o tejido de un sujeto después del tratamiento de dicha enfermedad o cuantificar el nivel de EMR significa cuantificar la cantidad de material genético que está asociado con la enfermedad en una muestra o tejido biológico de un sujeto, después del tratamiento de dicha enfermedad. Una célula muerta puede identificarse basándose en la expresión o carencia de expresión de un marcador biológico sobre la superficie de la célula muerta y/o dentro de dicha célula muerta o basándose en la presencia de al menos una molécula extraña sobre la superficie de la célula muerta y/o dentro de dicha célula muerta.

En la presente invención, la enfermedad es una enfermedad genética. Dicha enfermedad genética se caracteriza por al menos una variante o la ausencia de dicha al menos una variante en una secuencia de nucleótidos, en la que dicha variante es preferentemente una secuencia de nucleótidos clonotípicos para los reordenamientos génicos de inmunoglobulinas, carga alélica alta, una mutación puntual (SNV), una mutación múltiple (MNV), un indel, una inserción larga, una deleción larga y/o una translocación. Más preferentemente dicha enfermedad se caracteriza por una carga alélica alta y/o al menos una secuencia de nucleótidos clonotípica tumoral para al menos un reordenamiento génico de inmunoglobulinas, al menos una mutación puntual (SNV), al menos una mutación múltiple (MNV), al menos un indel, al menos una inserción larga, al menos una deleción larga y/o al menos una translocación. En una realización preferente de la invención, la enfermedad se selecciona entre cáncer o leucemia. En una realización preferente de la invención, dicha enfermedad se selecciona entre leucemia linfoblástica aguda, leucemia mieloide aguda, leucemia linfocítica crónica, leucemia mielógena (o mieloide) crónica, linfoma folicular, linfoma de células del manto, mieloma múltiple, cáncer de mama o neuroblastoma. Además, preferentemente, dicha enfermedad se caracteriza por:

- una mutación puntual (SNV), una mutación múltiple (MNV) y/o indel seleccionado entre una duplicación interna del tándem de FLT3 (FLT3-ITD) o una mutación de nucleofosmina1 (NMP1) en leucemia mieloide aguda;
- una inserción larga y/o translocación seleccionada de t(9;22) BCR-Abl o t(12;21) ETV6-RUNX1 (TEL-AML1) en leucemia linfoblástica aguda; t(15;17) PML-RARa, t(8;21) AML1-RUNX1T1 (AML-ETO) o inv(16) CBFb/MYH11 en leucemia mieloide aguda; t(9;22) BCR-Abl en leucemia mieloide, t(14;18) IgH/BCL2 en linfoma folicular; t(11;14) IgH/CCND1 (IgH/ BCL1) en linfoma de células del manto o t(4;14) en mieloma múltiple; o
- reordenamiento de inmunoglobulinas específico para el paciente en leucemia linfoblástica aguda, mieloma múltiple, linfoma de células del manto, linfoma folicular, leucemia linfocítica crónica o leucemia linfoblástica aguda. Además, el método de la presente invención puede aplicarse a cualquiera de las neoplasias hematológicas desveladas en Hauwel M., Matthes T. "Minimal residual disease monitoring: the new standard for treatment evaluation of haematological malignancies?"; Swiss Med Wkly. (2014) 144:w13907].

En una realización alternativa, el método de la invención es, por lo tanto, un método para cuantificar el nivel de carga alélica y/o al menos una secuencia nucleotídica clonotípica para al menos un reordenamiento génico de inmunoglobulinas, al menos una mutación puntual (SNV), al menos una mutación múltiple (MNV), al menos un indel, al menos una inserción larga, al menos una deleción larga y/o al menos una translocación en un sujeto que ha sido tratado para una enfermedad.

En la presente invención, el nivel de EMR se cuantifica en un sujeto que ha sido tratado para dicha enfermedad mediante un método que comprende siete etapas, (a) a (g). La etapa (a) comprende las etapas secuenciales de:

- amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k, por célula diploide de dicha muestra biológica; y
- secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres de lectura de izquierda a derecha.

De un modo análogo, la etapa (b) comprende las etapas secuenciales de:

- amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de caracteres que se leen de izquierda a derecha. La etapa (a) puede realizarse de forma simultánea con la etapa (b), o la etapa (a) puede realizarse antes o después de la etapa (b). Preferentemente, la etapa (b) se realiza antes de la etapa (a).

La muestra biológica en las etapas (a) y (b) comprende una muestra de materia biológica tomada de un sujeto. Dicha muestra biológica comprende al menos una secuencia de nucleótidos en al menos una célula. Preferentemente dicha muestra biológica comprende al menos una secuencia de nucleótidos en el ADN genómico de al menos una célula en un tejido, sangre, orina, heces, saliva, moco, esperma, hueso, cabello y/o uñas. La muestra biológica en la etapa (a) es un diagnóstico de muestras de ensayo (o de seguimiento) para una enfermedad mínima residual. La muestra biológica en la etapa (b) es un diagnóstico de muestras de diagnóstico (o calibración o control) para la enfermedad antes de un tratamiento. Preferentemente, la muestra biológica en la etapa (b) es una muestra con alta carga alélica o clonal que se ha tomado antes del tratamiento de la enfermedad del mismo sujeto que la muestra biológica en la etapa (a). La muestra biológica en la etapa (a) tiene un peso promedio, k , de ADN genómico por célula diploide.

Cada secuencia de nucleótidos en el ADN genómico se amplifica mediante PCR usando cebadores, donde dichos cebadores comprenden un cebador directo y un cebador inverso que se unen a secuencias complementarias distintas sobre las cadenas de Watson y Crick adyacentes a dicha secuencia nucleotídica, identificando, de este modo, los límites 5' y 3' de dicha secuencia nucleotídica. En particular, el extremo 3' de la secuencia nucleotídica de la cadena de Watson empieza con el nucleótido que es adyacente al nucleótido en el extremo 5' de la secuencia que está hibridada con el cebador directo. Por el contrario, el extremo 5' de la secuencia nucleotídica de la cadena de Watson empieza con el nucleótido complementario al nucleótido que es adyacente al nucleótido en el extremo 5' de la secuencia que está hibridada con el cebador inverso. De forma análoga, el extremo 3' de la secuencia nucleotídica de la cadena de Crick empieza con el nucleótido que es adyacente al nucleótido en el extremo 5' de la secuencia que está hibridada con el cebador inverso. Por el contrario, el extremo 5' de la secuencia nucleotídica de la cadena de Crick empieza con el nucleótido complementario al nucleótido que es adyacente al nucleótido en el extremo 5' de la secuencia que está hibridada con el cebador directo. En consecuencia, una polimerasa de ADN se une al extremo 5' de los cebadores anteriormente mencionados y replica la secuencia de nucleótidos múltiples veces.

Preferentemente, los cebadores son cebadores específicos de locus, escogidos para identificar al menos una variante específica de una secuencia nucleotídica presente en la muestra biológica en la etapa (b), donde dicha al menos una variante o la ausencia de dicha al menos una variante es indicativo de enfermedad. En particular, dicha al menos una variante comprende al menos una secuencia de nucleótidos clonotípicos para al menos un reordenamiento génico de inmunoglobulinas, al menos una mutación puntual (SNV), al menos una mutación múltiple (MNV), al menos un indel, al menos una inserción larga y/o al menos una translocación. La al menos una variante específica de una secuencia nucleotídica presente en la muestra biológica obtenida del sujeto antes del tratamiento para la enfermedad es la al menos una variante específica de una secuencia nucleotídica que es indicativa de dicha enfermedad e identificada en dicha muestra biológica en la mayor proporción. Por lo tanto, cada variante específica indicativa de dicha enfermedad se identifica y la proporción de cada variante específica se compara, con la variante específica o variantes específicas que está o están clasificadas en la mayor proporción, respectivamente, siendo esa o esas cuyos cebadores específicos de locus han sido elegidos para identificar. Cabe señalar que en una realización el proceso de clasificación es una etapa que se lleva a cabo mediante el FrequencyRank.java script.

Como consecuencia del hecho de que al menos una variante específica de una secuencia nucleotídica está, de este modo, identificada en las etapas (a) y (b), y/o más de un tipo de cebador puede usarse en las etapas (a) y (b), las etapas (a) y (b) implican la identificación, amplificación y secuenciación de al menos una secuencia nucleotídica (es decir, una o más secuencias de nucleótidos) en una muestra biológica, proporcionando, de este modo, al menos una lista de caracteres (es decir, una o más listas de caracteres) que corresponden a la misma. Sin embargo, en una realización más preferente, los cebadores usados en las etapas (a) y (b) son un cebador directo específico de locus y un cebador inverso específico de locus escogidos para identificar una variante específica de una secuencia de nucleótidos presente en las muestras biológicas, en la que dicha una variante o la ausencia de dicha una variante es indicativo de enfermedad. Por lo tanto, es esta realización preferente, la etapa (b) comprende las etapas secuenciales de:

- amplificar mediante reacción en cadena de polimerasa usando el mismo cebador directo específico de locus y el cebador inverso específico de locus usado en la etapa (a), una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad de dicha secuencia nucleotídica; y
- secuenciar dicha secuencia de nucleótidos para obtener una segunda lista de caracteres que se leen de izquierda a derecha.

Por lo tanto, la amplificación de al menos una secuencia nucleotídica presente en cada muestra biológica se realiza con cebadores específicos que identifican al menos una región de interés (es decir, al menos una variante específica indicativa de la enfermedad para la cual el sujeto ha sido tratado), antes de procesar cada una sobre una interfaz de secuenciación paralela masiva. En consecuencia, la muestra de ensayo sobre esta al menos una región de interés se amplificó y secuenció con una cobertura de sensibilidad superior, o igual a, la esperada. Para la amplificación de la muestra de ensayo, una cantidad, D , de ADN genómico (ADNg) de dicha muestra de ensayo (seguimiento) se usa en la PCR y la amplificación se repite preferentemente hasta que se obtiene una cantidad suficiente para la secuenciación con una sensibilidad deseada. Preferentemente, una cantidad, D , de ADNg de dicha muestra de

ensayo se usa en la PCR para asegurar que se obtiene una sensibilidad equivalente a la obtenible a partir de muestrear una cantidad de células. La sensibilidad se determinada en cada caso para la aplicación del estudio de la enfermedad y de células tumorales circulantes residuales.

- 5 La cantidad, D, del ADN genómico de la muestra de ensayo (seguimiento) que se usa en la PCR para secuenciar con una sensibilidad (S) deseada se establece, en primer lugar, midiendo la concentración de ADN ([ADN], pg/μl) en la muestra biológica obtenida de un sujeto después del tratamiento de dicha enfermedad (muestra de ensayo). Este valor, a continuación, se usa para determinar el número de células equivalentes por microlitro (N) de la muestra de ensayo según la siguiente fórmula:

10

$$N = [ADN] / k$$

en el que N y [ADN] son tal como se definen anteriormente, y k es el peso promedio del AND genómico por célula diploide de la muestra de ensayo, en el cual k preferentemente asume un valor de 6,49 picogramos por célula. El número de células equivalentes por microlitro (N) de la muestra de ensayo permite, posteriormente, el cálculo del volumen de la muestra (V, μl) que es necesario para usar en la PCR para alcanzar una sensibilidad (S) deseada según la siguiente fórmula:

15

$$V = 1 / (NxS)$$

20

Una sensibilidad de 10^{-5} equivale al que se puede lograr a partir del uso de ADN genómico a partir de al menos 100.000 células equivalentes. El volumen de la muestra de ensayo (V) determina la cantidad de experimentos de PCR necesarios para obtener una cantidad suficiente (D, pg) de ADN genómico para la secuenciación y, además, se usa para calcular la cantidad (D) del ADN genómico de la muestra de ensayo que se usa en la PCR según la siguiente fórmula:

25

$$D = [ADN] \times V$$

30

La amplificación puede llevarse a cabo mediante una cualquiera de las siguientes técnicas de PCR seleccionadas de PCR-multiplex y PCR único usando un par de cebadores. Preferentemente la amplificación se realiza mediante RCP-multiplex.

35

Opcionalmente, las etapas (a) y (b) pueden comprender una etapa adicional de aislamiento de dicha una secuencia nucleotídica amplificada antes de la etapa de secuenciación usando métodos habituales en la técnica. Por lo tanto, la primera etapa de las etapas (a) y (b) comprende la amplificación de al menos una secuencia de nucleótidos obtenida a partir de al menos una secuencia de nucleótidos larga mediante amplificación selectiva de dicha al menos una secuencia de nucleótidos sobre dicha al menos una secuencia de nucleótidos más larga, en la que cada secuencia de nucleótidos más larga comprende un polinucleótido, donde dicho polinucleótido se selecciona preferentemente dentree un ADN o ARN de cadena simple o doble, más preferentemente ADN de cadena doble, además preferentemente, ADN genómico de cadena doble. Cuando dicho polinucleótido es un ADN de cadena simple, se sintetiza una secuencia complementaria a partir de este antes de llevar a cabo las etapas (a) o (b) para proporcionar ADN de cadena doble. Cuando dicho polinucleótido es ARN, se sintetiza un ADN de cadena doble complementario (retrotranscrito) a partir de este antes de llevar a cabo las etapas (a) o (b).

40

45

La al menos una secuencia nucleotídica de cada una de las etapas (a) y (b), amplificada de este modo y opcionalmente aislada, se secuencia posteriormente. La secuenciación de una secuencia nucleotídica de la etapa (a) proporciona una primera lista de caracteres que se leen de izquierda a derecha que se corresponden a la misma, en la que cada primera lista de caracteres tiene un número total de caracteres, C_i. Por otra parte, el número total de primeras listas de caracteres (L_i) se corresponde con el número total de secuencias de nucleótidos distintas en la etapa (a). La secuenciación de una secuencia nucleotídica de la etapa (b) del mismo modo proporciona una segunda lista de caracteres que se leen de derecha a izquierda, que se corresponden a la misma.

50

La secuenciación es una técnica de secuenciación de nucleótidos multiplex y/o de alto rendimiento. Preferentemente, la secuenciación se lleva a cabo mediante una técnica de nueva generación, más preferentemente secuenciación paralela masiva [por ejemplo, secuenciación masiva paralela de firmas (MPSS)]. En una realización de la presente invención, cuando se usan múltiples cebadores en la secuenciación, las etapas de secuenciación en las etapas (a) y (b) se llevan a cabo usando códigos de barras para identificar entre los distintos cebadores usados. En una realización especialmente preferente de la presente invención, la secuenciación se lleva a cabo mediante secuenciación masiva paralela usando PCR de emulsión.

60

Cada una de las distintas etapas de amplificación y secuenciación de dicha al menos una secuencia nucleotídica en las etapas (a) y (b) puede llevarse a cabo con medios separados (es decir, con instrumentos separados). Como alternativa, dos o más de estas distintas etapas puede llevarse a cabo con el mismo instrumento.

65

La secuenciación de una secuencia nucleotídica en las etapas (a) y (b) proporciona una lista correspondiente de caracteres, mediante la cual cada carácter en cada lista de caracteres comprende una letra. En una realización de la

presente invención, la secuenciación de una secuencia nucleotídica en las etapas (a) y (b) proporciona una lista correspondiente de caracteres, mediante la cual cada carácter en cada lista de caracteres comprende una letra asociada con un número (o símbolo). Más preferentemente, cada letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica que tiene la calidad más alta (Q) dentro de los límites del método de secuenciación, y el número o símbolo asociado con la misma es la calidad (Q), en la que Q es un número entero que representa la probabilidad de que la letra que representa un nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica sea incorrecta. Por lo tanto, cada una de las listas de caracteres obtenidas en las etapas (a) y (b) de la presente invención está preferentemente comprendida en un archivo de formato de secuencias, más preferentemente un archivo FASTQ.

Como alternativa, cada carácter representa, más preferentemente, el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica en mayor proporción. En una realización aún más preferente de esta alternativa, la letra asociada con dicho carácter representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica en la mayor proporción y el número o símbolo asociado con la misma es la proporción (por ejemplo, como porcentaje, fracción o relación) de dicho nucleótido que se identifica en la misma.

Una secuencia continua de caracteres es una lista que no está rota por otro carácter o ausencia de un carácter, en la que dicha secuencia continua de caracteres representa una secuencia continua de nucleótidos sin romper. De forma análoga con lo anteriormente descrito, cada carácter en la secuencia continua de caracteres comprende una o más letras, preferentemente una o más letras asociadas con un número o símbolo, más preferentemente en el que cada letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica que tiene la calidad más alta (Q) dentro de los límites del método de secuenciación, y el número o símbolo asociado con la misma es la calidad (Q), donde Q es un número entero que representa la probabilidad de que la letra que representa un nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica sea incorrecta. Como tal, en esta realización más preferente de la invención, dicha secuencia continua de caracteres comprende una secuencia continua de letras que representa una secuencia continua de nucleótidos, cuando cada carácter en la secuencia continua de caracteres representa el nucleótido que se identifica en la posición correspondiente que tiene la mayor calidad (Q) dentro de los límites del método de secuenciación.

Como alternativa, cada carácter en la secuencia continua de caracteres comprende preferentemente una letra asociada con un número o símbolo, más preferentemente donde cada letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica en la mayor proporción. En una realización de esta más preferente alternativa, la letra asociada con dicho carácter en la secuencia continua de caracteres representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica en la mayor proporción y el número o símbolo asociado con la misma es la proporción (por ejemplo, como porcentaje, fracción o relación) de dicho nucleótido que se identifica en la misma. Como tal, en esta realización alternativa más preferente, dicha secuencia continua de caracteres comprende una secuencia continua de letras que representa una secuencia continua de nucleótidos, en la que cada carácter en la secuencia continua de caracteres representa el nucleótido que se identifica en la posición correspondiente en la secuencia continua de nucleótidos en la mayor proporción.

Cada carácter en cada lista de caracteres se corresponde con un nucleótido en dicha secuencia de nucleótidos y el orden de los caracteres en dicha lista se corresponde con el orden de nucleótidos en dicha secuencia de nucleótidos. Por lo tanto, el carácter en el extremo izquierdo de dicha lista se corresponde con el nucleótido o proporción de nucleótidos en el extremo 3' de la cadena de Watson de dicha secuencia de nucleótidos y el carácter en el extremo derecho de dicha lista se corresponde con el nucleótido o proporción de nucleótidos en el extremo 5' de la cadena de Watson de dicha secuencia de nucleótidos. De forma análoga, una lista complementaria (o parcialmente complementaria) de caracteres se obtiene representando cada cadena de Crick de dicha secuencia nucleotídica, en la cual el carácter en el extremo izquierdo de dicha lista se corresponde con el nucleótido o proporción de nucleótidos en el extremo 3' de la cadena de Crick de dicha secuencia de nucleótidos y el carácter en el extremo derecho de dicha lista se corresponde con el nucleótido o proporción de nucleótidos en el extremo 5' de la cadena de Crick de dicha secuencia de nucleótidos.

Posteriormente, se realiza una comparación de cada primera lista de caracteres obtenida en la etapa (a) con cada segunda lista de caracteres obtenida en la etapa (b). Dicha comparación se realiza para determinar finalmente el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres. En otras palabras, la comparación se realiza para determinar las L_c que son idénticas a (es decir, coinciden) con una segunda lista de caracteres. Para determinar L_c , es necesario determinar el grado de similitud de cada primera lista de caracteres obtenida en la etapa (a) con cada segunda lista de caracteres obtenida en la etapa (b), donde un grado de similitud, DS, se determina para una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b). Aunque se conocen métodos adaptados a la bioinformática que acceden a datos externos (por ejemplo, bases de datos genéticas de poblaciones) para llevar a cabo la etapa de comparación y de algún modo implementar "conocimiento biológico", el método de la presente invención funciona sin la necesidad de acceder a datos externos. Con este objetivo, la primera característica que se considera esencial a implementar es una lógica difusa. La tasa de error de secuenciadores que usan una lógica clásica binaria - en la cual las secuencias solo pueden ser iguales o distintas - es tan alta que no resulta útil. Una elevada proporción (casi

toda) de secuencias de nucleótidos que evalúan como distintas, son iguales pero aparecen como distintas debido a un error en el secuenciador. Por lo tanto, se implementa un proceso de comparación para evaluar el grado de similitud entre dos listas cualesquiera de caracteres.

5 En una realización de la invención, cada carácter en una lista de caracteres comprende una letra, de modo que un carácter en la primera lista de caracteres se determina como el mismo que el carácter en la segunda lista de caracteres, cuando la letra es la misma en la primera y segunda lista de caracteres (es decir, un carácter en una lista de caracteres se determina como el mismo que el carácter en otra lista de caracteres cuando las letras son las mismas en cada lista). En una realización preferente de la invención, cada carácter en una lista de caracteres comprende una letra asociada con un número o símbolo, más preferentemente en el que cada letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica en la calidad más alta (Q) dentro de los límites del método de secuenciación y donde cada número o símbolo representa la calidad (Q). Por lo tanto, en dicha realización más preferente del método de la invención, donde cada carácter en una primera lista de caracteres y cada carácter en una segunda lista de caracteres comprende una letra asociada con un número o símbolo, donde dicho número o símbolo representa la calidad (Q) donde dicha letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica que tiene la mayor calidad (Q), un carácter en la primera lista de caracteres se determina como el mismo que el carácter en la segunda lista de caracteres, cuando la letra que tiene la mayor calidad es la misma en la primera y segunda lista de caracteres (es decir, un carácter en una lista de caracteres se determina como el mismo que el carácter en otra lista de caracteres cuando las letras son las mismas en cada lista). Además, no solo las letras sino que también los número o símbolos asociados con las mismas pueden compararse entre las listas, preferentemente comparando la letra y el número o símbolo que representa la calidad (Q) asociada a la misma para cada carácter en cada lista. Por lo tanto, en una realización más preferente del método de la invención, un carácter en una lista de caracteres que comprende una letra asociada con una calidad (Q) se determina como el mismo que el carácter en otra lista de caracteres que comprende una lista asociada con una calidad (Q'), cuando la letra que tiene la mayor calidad es la misma en cada lista, y la calidad (Q) de las letras es la misma en cada lista dentro de un límite de punto de corte o un error, más preferente un límite de punto de corte. Por ejemplo, un carácter en una posición dada que se asigna como T con una calidad de 1,00 (es decir, 100 %) puede considerarse la misma que un carácter en una posición dada que se asigna como T con una calidad de 0,99 (es decir, 99 %), cuando el límite de punto de corte se establece en 0,99 (es decir, el error se establece al 1 %). Por lo tanto, en la siguiente etapa (c), cada etapa de selección del carácter o secuencia continua más larga de caracteres que son los mismos, dentro de un límite de punto de corte, en la primera y segunda listas de caracteres o partes de la misma, comprende, en primer lugar, realizar la anteriormente mencionada comparación entre la primera lista de caracteres y la segunda lista de caracteres o partes de la misma y, en segundo lugar, escoger el carácter o secuencia continua más larga de caracteres basándose en los criterios que se proporcionan a continuación, cuando uno o más caracteres o una o más secuencias continuas de caracteres se identifican como la más larga de dicha comparación. En este método, el límite de punto de corte se establece a una calidad (Q) de 0,99, más preferentemente de 0,999, más preferentemente de 0,9999, lo más preferentemente de 0,99999 y/o el error se establece a un máximo del 1 %, más preferentemente un 0,1%, más preferentemente un 0,01%, más preferentemente un 0,001%. En una realización más preferente del método de la invención, un carácter en una primera lista de caracteres se determina como el mismo que el carácter en una segunda lista de caracteres, cuando la letra que tiene la calidad más alta (Q) es la misma en la primera y segunda lista de caracteres, y la calidad de la letra de la primera y segunda lista de caracteres se encuentra dentro de 0,01 (1%) de la calidad de la letra en la segunda lista de caracteres, además preferentemente dentro de 0,001 (0,1%), aún más preferentemente dentro de 0,0001 (0,01%), lo más preferentemente dentro de 0,00001 (0,001%).

45 Como alternativa, la comparación se realiza comparando la letra que comprende cada carácter que está presente en la mejor calidad (Q) o en la mejor proporción en cada lista de caracteres. Por lo tanto, en el método de la invención, un carácter en una lista de caracteres se determina como el mismo que un carácter en otra lista de caracteres preferentemente cuando las listas son iguales.

50 Como alternativa, la comparación se realiza comparando la proporción de cada una o más letras que comprenden cada carácter. Por lo tanto, la comparación se realiza comparando la proporción de cada uno o más nucleótidos que se identifica en cada posición en la secuencia nucleotídica. En este método, un carácter en una lista de caracteres se determina como el mismo que un carácter en otra lista de caracteres cuando la proporción de letras es la misma dentro del error. Por ejemplo, un carácter para el cual la proporción de A en una posición dada es 0,11 y la proporción de T en dicha posición dada es de 0,89 (es decir, la relación de A:T es 0,11:0,89) puede considerarse la misma que un carácter para el cual la proporción de A en una posición dada es 0,1 y la proporción de T en dicha posición dada es 0,9 (es decir, la relación de A:T es de 0,1:0,9, cuando el error se establece al 5 % de error). Por lo tanto, en la siguiente etapa (c), cada etapa de selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y la segunda lista de caracteres o partes de las mismas comprende, en primer lugar, realizar la anteriormente mencionada comparación entre la primera lista de caracteres y la segunda lista de caracteres y partes de las mismas y, en segundo lugar, escoger el carácter o secuencia continua más larga de caracteres basándose en los criterios proporcionados a continuación, cuando uno o más caracteres o una o más secuencias continuas de caracteres se identifican como la más larga de dicha comparación. En este método, el error se establece en un máximo del 1%, más preferentemente un 0,1%, más preferentemente un 0,01%, más preferentemente un 0,001%.

Por lo tanto, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b) se determina posteriormente en la etapa (c), donde un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante sub-etapas (i) a (x) [sub-etapas (i) a (vii) de las cuales se representan esquemáticamente en la **Figura 1**] o (xi) a (xvii) [sub-etapas (xi) a (xv) de las cuales se representan esquemáticamente en la **Figura 2**]. En particular, un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

(i) a selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(iv) exclusión de cada carácter y/o de cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(v)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(vi) exclusión de cada carácter y/o de cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(viii) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y

- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos que en la segunda y primera lista de caracteres, respectivamente;

(ix) sumar

- C_c ; y

- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o

ES 2 659 492 T3

secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y

- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

o mediante:

(xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(xiv) exclusión de cada carácter y/o de cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(xvi) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;
(xvii) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(xviii) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

En la Etapa (c), las sub-etapas (i) y (xi) de selección del carácter que es el mismo en la primera y segunda listas de caracteres implica la comparación de caracteres individuales en la primera y segunda lista de caracteres según los

anteriores criterios de comparación. Por otra parte, las sub-etapas (i) y (xi) de selección de la secuencia continua más larga de caracteres que es la misma en la primera y segunda listas de caracteres implica la comparación consecutiva de caracteres individuales en la primera y segunda lista de caracteres según los anteriores criterios de comparación. Cabe señalar que las sub-etapas (i) a (iv) y (xi) a (xiv) son idénticas.

5 En la Etapa (c), después de cada etapa de selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres o partes de las mismas, una etapa de exclusión de dicho carácter o secuencia continua más larga de caracteres, seleccionada de este modo, tiene lugar, donde cada etapa de exclusión comprende retirar el carácter o secuencia continua más larga de caracteres, seleccionada de este modo, a partir de la consideración en las etapas posteriores de selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres. Cabe señalar que cada etapa de exclusión de resultados en una secuencia no continua de caracteres que está rota en el punto entre cada carácter que flanquea el carácter o secuencia continua más larga de caracteres, excluida de este modo. Como tal, cualquier etapa posterior de selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda listas de caracteres o partes de las mismas en las sub-etapas (v) a (vii) de la etapa (c) no considera una secuencia que se extiende más allá de un carácter o secuencia continua más larga de caracteres previamente excluida, sino que considera la secuencia continua de caracteres ubicados adyacentes a cada carácter o a cada secuencia continua más larga de caracteres excluida en la etapa previa. Por otra parte, cualquier etapa posterior de selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda listas de caracteres o partes de las mismas en las sub-etapas (xiii) a (xv) no considera una secuencia que une los caracteres soenbre cualquier lado del carácter o secuencia continua más larga de caracteres excluida.

25 Cada ciclo de selección y exclusión de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres se repite hasta que no se selecciona ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres. En las sub-etapas (iii) y (xiii) del a etapa (c) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres que se encuentra a la izquierda del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente, y para la secuencia continua de caracteres que se encuentra a la derecha del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente. Como alternativa, esto puede repetirse, en primer lugar, para la secuencia continua de caracteres que se encuentra a la izquierda del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente y, a continuación, para la secuencia continua de caracteres que se encuentra a la derecha del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente. Como alternativa, esto puede repetirse, en primer lugar, para la secuencia continua de caracteres que se encuentra a la derecha del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente y, a continuación, para la secuencia continua de caracteres que se encuentra a la izquierda del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii), respectivamente.

40 De forma análoga, en la sub-etapa (v) del a etapa (c) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa, y para la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa. Como alternativa, esto puede repetirse, en primer lugar, para la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa y, a continuación, para la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa. Como alternativa, esto puede repetirse, en primer lugar, para la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa y, a continuación, para la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa.

55 Cada uno de estos enfoques es novedoso en comparación con las ejecuciones estándar y genera un resultado de comparación ligeramente más compacto.

60 Por otra parte, en una realización preferida de la etapa (c), cada etapa de selección del carácter o secuencia continua más larga de caracteres es una etapa de selección de la secuencia continua más larga de caracteres, en la que dicha secuencia continua más larga de caracteres comprende un mínimo de dos caracteres. Como tal, cada ciclo de selección y exclusión de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres se repite en la etapa (c) hasta que no se seleccione ninguna secuencia continua más larga de caracteres que tiene un mínimo de dos caracteres que son los mismos en la primera y segunda lista de caracteres. Más preferentemente, dicha secuencia continua más larga de caracteres comprende un mínimo de 3 caracteres, además preferentemente un mínimo de 4 caracteres.

65 Una vez que no sea posible seleccionar un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y segunda lista de caracteres (porque todos los caracteres o secuencias

continuas más largas de caracteres que son los mismos en la primera lista de caracteres y segunda lista de caracteres hayan sido excluidos), el número total de caracteres, C_c , que han sido excluidos de la primera lista de caracteres y excluidos en la segunda lista de caracteres se obtiene mediante las sub-etapas (viii) y/o (xvi) de la etapa (c), en la que el número de caracteres en la primera lista de caracteres que han sido respectivamente
 5 excluidos en cualquiera de las sub-etapas (i) a (vii) y (xi) a (xv) y el número de caracteres en la segunda lista de caracteres que se excluyeron respectivamente en cualquiera de las sub-etapas (i) a (vii) y (xi) a (xv) se suma. El número total de caracteres, C_c , que se han excluido en la primera y segunda lista de caracteres también puede considerarse como $2 \times$ (el número de caracteres que se han excluido en la primera lista de caracteres) o como $2 \times$ (el número de caracteres que se han excluido en la segunda lista de caracteres). De forma análoga, el número total
 10 de caracteres, C_t , en la primera lista de caracteres se obtiene mediante las sub-etapas (ix) y/o (xvii) de la etapa (c), donde C_c , más la cantidad de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres que no se excluyeron en ninguna de las sub-etapas (i) a (vii) y (xi) a (xv) de la etapa (c), más el número de caracteres en la
 15 segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres que no se excluyeron en ninguna de las sub-etapas (i) a (vii) y (xi) a (xv) de la etapa (c), se suma.

Por lo tanto, el método de la presente invención está destinado a detectar una lista de caracteres que representan una secuencia nucleotídica específica, que se proporciona como un argumento al método, dentro de un archivo de
 20 datos que contiene una mezcla de listas de caracteres cada una representando una secuencia nucleotídica fragmentada en lugares que, desde un punto de vista informático se considera aleatorio. Por lo tanto, la al menos una lista de caracteres en la mezcla tiene longitudes aleatorias y no se conoce de antemano donde puede encontrarse una lista de caracteres que representa una secuencia de nucleótidos específica [en la etapa (b)] en cada al menos una lista de caracteres. Por lo tanto, el método de la presente invención comprende una combinación de
 25 alineación y comparación. Puesto que comprende una mezcla de alineación y comparación, se considera que en la invención, la comparación se realiza solo a partir del primer carácter o secuencia continua de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres y el último carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres (es decir, que incluye y entre los caracteres que coinciden o secuencias continuas más largas de caracteres más cerca de los
 30 extremos de la primera y segunda listas de caracteres), mediante la cual C_c y C_t se determinan sobre una parte de la primera y segunda lista de caracteres que es desde el primer carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda listas de caracteres y al último carácter o secuencia continua más larga de caracteres que son los mismos en la primera y segunda lista de caracteres. En consecuencia, en una realización preferente, cada uno de los caracteres o secuencias continuas más largas de caracteres que se han eliminado en
 35 cualquiera de las sub-etapas (i) a (vii) o (xi) a (xv) de la etapa (c) se colocan en un archivo .adn y en las sub-etapas (vii) y (xvi) puede usar los datos en un archivo .ad para calcular C_c usando una opción de ajuste, de modo que la comparación se realiza solo entre el primero y último carácter o secuencia continua más larga de caracteres en la primera lista que son los mismos que en la segunda lista.

Posterior a la etapa (c), se lleva a cabo una etapa (d), en la que para cada primera lista de caracteres obtenida en la etapa (a), se selecciona el DS del valor más alto, DS_{HV} . Sin embargo, por el hecho de que cada secuencia nucleotídica en las etapas (a) y (b) (definidas en el presente documento de forma arbitraria como la cadena de Watson) tienen una secuencia nucleotídica complementaria (en particular la secuencia complementaria inversa, definida en el presente documento de forma arbitraria como la cadena de Crick), de modo que la primera lista de
 45 caracteres obtenida en la etapa (a) y la segunda lista de caracteres obtenida en la etapa (b) también tienen una primera lista de caracteres complementaria inversa correspondiente y una segunda lista de caracteres complementaria inversa correspondiente, respectivamente, en una realización preferente, no solo las anteriormente mencionadas secuencias de nucleótidos en las etapas (a) y (b) sino que también las secuencias de nucleótidos complementarias pueden someterse a la etapas (a), (b) y (c) del método de la invención. Por lo tanto, en dicha
 50 realización preferente, la etapa (d) comprende:

- amplificar mediante reacción en cadena de polimerasa usando cebadores, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (a) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa
 55 para obtener al menos una primera lista complementaria inversa de caracteres que se leen de izquierda a derecha; y
- amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa previa, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (b) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa para obtener al menos una segunda lista complementaria inversa de caracteres que se
 60 leen de izquierda a derecha; y
- determinar, para cada primera lista complementaria inversa de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista complementaria inversa de caracteres obtenida en la etapa (b), donde un grado de similitud, DS_{rcs} , de una primera lista complementaria inversa de caracteres obtenida en la etapa (a) con una
 65 segunda lista complementaria inversa de caracteres obtenida en la etapa (b) se determina bien mediante:

- (xix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;
- 5 (xx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;
- 10 (xxi)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;
- 15
- 20
- 25 (xxii) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;
- 30 (xxiii)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;
- 35
- 40
- 45 (xxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;
- 50 (xxv) repetición de las etapas (xxii) y (xxiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;
- (xxvi) sumar
- el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv); y
 - el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv) para obtener el número total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera listas complementarias inversas de caracteres, respectivamente;
- 55
- 60
- (xxvii) sumar
- C_c ; y
 - el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista
- 65

complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c); y

- el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda listas complementarias inversas de caracteres; y

(xxviii) calcular el DS según la siguiente fórmula:

$$DS_{RCS} = C_c / C_t$$

o mediante:

(xxix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;

(xxx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xxix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxxi)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;

(xxxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxxiii) repetición de las etapas (xxxii) y (xxii) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;

(xxvi) sumar

- el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv) para obtener el número total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera listas complementarias inversas de caracteres, respectivamente;

(xxvii) sumar

- C_c ; y
- el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda listas complementarias inversas de caracteres; y (xxviii) calcular el DS según la siguiente fórmula:

$$5 \quad DS_{rcs} = C_c / C_t$$

en el que el DS se determina para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (i) a (x), el DS_{rcs} se determina para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xix) a (xxviii) y cuando el DS está determinado para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (xi) a (xvii), el DS_{rcs} se determina para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xxix) a (xxxvi); y seleccionar, para cada primera lista de caracteres obtenida en la etapa (a) y su primera lista complementaria inversa de caracteres, el DS o DS_{rcs} de valor más alto, DS_{HV} .

15 Por lo tanto, un DS o DS_{rcs} de 0,0 significa que ningún carácter en una primera lista de caracteres es el mismo que en la segunda lista de caracteres, mientras que un DS o DS_{rcs} de 1,0 significa que todos los caracteres en una primera lista de caracteres son los mismos que en una segunda lista de caracteres (es decir, una secuencia nucleotídica de una muestra biológica de un sujeto tras el tratamiento de dicha enfermedad es estrictamente igual con una secuencia nucleotídica de una muestra biológica obtenida de un sujeto antes del tratamiento de dicha enfermedad). Por lo tanto, el método de la presente invención proporciona información sobre cómo muchas secuencias de nucleótidos en la muestra biológica de un sujeto contienen la secuencia de argumento (la secuencia nucleotídica de una muestra biológica obtenida de un sujeto con dicha enfermedad), bien en su forma original (Watson) o en su versión complementaria inversa (Crick).

25 Habiendo determinado el DS_{HV} para cada una de las al menos una primera lista de caracteres obtenida en la etapa (a), el número de primeras listas de caracteres obtenidas en la etapa (a) que tienen un DS_{HV} superior a un valor de umbral, T, se suma posteriormente en una etapa (e) para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres. De manera similar, en una etapa (f), L_c y el número de primeras listas de caracteres que no tienen un DS_{HV} superior a T se suman para obtener L_t . L_t se corresponde con el número total de primeras listas de caracteres. Preferentemente dicho valor umbral, T, para el DS y DS_{rcs} se establece en 0,99, más preferentemente de 0,999, más preferentemente de 0,9999, lo más preferentemente 0,99999. Las etapas (e) y (f) pueden llevarse a cabo de forma simultánea o la etapa (e) puede llevarse a cabo antes o después de la etapa (f), preferentemente la etapa (f) se lleva a cabo después de la etapa (e).

35 Finalmente, se lleva a cabo una etapa (g) para calcular el nivel de EMR. El cálculo del nivel de EMR se lleva a cabo según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_t^2$$

40 en la que L_c , D, k y L_t son tal como se ha definido anteriormente y como se define a continuación:

L_c = número total de primeras listas de caracteres que son iguales que una segunda lista de caracteres;

D = cantidad, D, de ADN genómico a partir de una muestra biológica obtenida de un sujeto tras el tratamiento de una enfermedad (a partir de la cual dicha una primera lista de caracteres se obtiene mediante secuenciación);

45 k = peso promedio, k, de ADN genómico por célula diploide de una muestra biológica obtenida a partir de un sujeto después del tratamiento para una enfermedad;

L_t = número total de primeras listas de caracteres.

50 En una realización de la invención, Una EMR de 1 es 100 % indicativo de enfermedad en dicho sujeto y una EMR de 0 es 0% indicativo de enfermedad en dicho sujeto. En otras palabras, una EMR de 1 indica que dicho sujeto padece dicha enfermedad, mientras que una EMR de 0 indica que el sujeto está libre de cualquier enfermedad (al menos en las células o tejido de la muestra biológica). Como tal, la EMR puede usarse para diagnosticar la presencia de dicha enfermedad en dicho sujeto y/o determinar el enfoque terapéutico más adecuado (si se necesita).

55 En una realización, la EMR se produce como un resultado (preferentemente en forma de un archivo, o sobre una pantalla o pieza de papel) después de que la etapa (d) se haya llevado a cabo. Preferentemente, la EMR, junto con al menos una primera lista de caracteres se produce como un resultado (preferentemente en la forma de un archivo, sobre una pantalla o pieza de papel) detallando el carácter o secuencia continua más larga de caracteres que son iguales en la primera y segunda listas de caracteres. Opcionalmente, dicho resultado detalla C_c , C_t , DS, DS_{HV} , D, k, L_c y/o L_t , donde sean relevantes con el DS_{rcs} .

60 En una realización especialmente preferente, la presente invención se refiere a un método para cuantificar la enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad, que comprende:

65

a)

- amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k, por célula diploide de dicha muestra biológica; y
- secuenciar dicho al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres de lectura de izquierda a derecha;

(b)

- amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- secuenciar dicho al menos una secuencia de nucleótidos para obtener al menos una segunda lista de lectura de caracteres de izquierda a derecha;

(c) determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), donde un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

(i) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(v)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de

caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(viii) sumar

- 5
- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

10 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera listas de caracteres, respectivamente;

(ix) sumar

- 15
- C_c ; y
 - el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),
- 20

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y

(x) calcular el DS según la siguiente fórmula:

25
$$DS = C_c / C_t$$

o mediante:

30 (xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

35 (xiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

45 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

50 (xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

55 (xvi) sumar

- 60
- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

65 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

(xvii) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y (xviii) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

(d)

- amplificar mediante reacción en cadena de polimerasa usando cebadores, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (a) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa para obtener al menos una primera lista complementaria inversa de caracteres que se leen de izquierda a derecha; y
- amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa previa, cada al menos una secuencia de nucleótidos que es la secuencia complementaria inversa complementaria a la al menos una secuencia de nucleótidos en la etapa (b) y secuenciar dicha al menos una secuencia de nucleótidos complementaria inversa para obtener al menos una segunda lista complementaria inversa de caracteres que se leen de izquierda a derecha; y
- determinar, para cada primera lista complementaria inversa de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista complementaria inversa de caracteres obtenida en la etapa (b), donde un grado de similitud, DS_{rcs} , de una primera lista complementaria inversa de caracteres obtenida en la etapa (a) con una segunda lista complementaria inversa de caracteres obtenida en la etapa (b) se determina bien mediante:

(xix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;

(xx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxi)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;

(xxii) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria

inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y

- 5 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;

15 (xxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxv) repetición de las etapas (xxii) y (xxiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;

20 (xxvi) sumar

- el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv); y
- 25 - el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv) para obtener el número total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera listas complementarias inversas de caracteres, respectivamente;

30 (xxvii) sumar

- C_c ; y
- el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c); y
- 35 - el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xix) a (xxv) de la etapa (c),

40 para obtener el número total de caracteres, C_t , en la primera y segunda lista complementarias inversas de caracteres; y

(xxviii) calcular el DS según la siguiente fórmula:

45
$$DS_{rcs} = C_c / C_t$$

o mediante:

50 (xxix) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres;

55 (xxx) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xxix) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxxi)

- 60 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas complementarias inversas de caracteres; y
- 65 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del

carácter o secuencia continua más larga de caracteres excluidos en la etapa (xxx) que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas complementarias inversas de caracteres;

(xxxiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xxxiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres;

(xxxiii) repetición de las etapas (xxxii) y (xxii) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista complementaria inversa de caracteres y en la segunda lista complementaria inversa de caracteres sea seleccionado;

(xxxiv) sumar

- el número de caracteres en la primera lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xxix) a (xxxiii); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se excluyeron en cualquiera de las etapas (xxix) a (xxxiii) para obtener la cantidad total de caracteres, C_c , en la primera y segunda listas complementarias inversas de caracteres que son los mismos en la segunda y primera lista complementarias inversas de caracteres, respectivamente;

(xxxv) sumar

- C_c ; y
- el número de caracteres en la primera lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xxix) a (xxxiii) de la etapa (c); y
- el número de caracteres en la segunda lista complementaria inversa de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista complementaria inversa de caracteres y que no se excluyeron en ninguna de las etapas (xxix) a (xxxiii) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista complementarias inversas de caracteres; y

(xxxvi) calcular el DS según la siguiente fórmula:

$$DS_{rcs} = C_c / C_t$$

en el que el DS se determina para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (i) a (x), el DS_{rcs} se determina para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xix) a (xxviii) y cuando el DS se determina para cada primera lista de caracteres obtenida en la etapa (a) usando las sub-etapas (xi) a (xvii), el DS_{rcs} se determina para cada primera lista complementaria inversa de caracteres correspondiente usando las sub-etapas (xxix) a (xxxvi); y seleccionando, para cada primera lista de caracteres obtenida en la etapa (a) y su primera lista complementaria inversa de caracteres, el DS o DS_{rcs} de valor más alto, DS_{HV}

(e) sumar el número de primeras listas de caracteres que tienen un DS_{HV} superior a un valor umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;

(f) sumar

- L_c ; y
- el número de primeras listas de caracteres que no tienen un DS_{HV} superior a T ,

para obtener el número total de primeras listas de caracteres, L_i ; y

(g) calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_i^2;$$

donde en las sub-etapas (iii) y (xiii) de la etapa (c) y las sub-etapas (xxi) y (xxxi) de la etapa (d) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres que se encuentra a la izquierda del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii) de la etapa (c) y sub-etapas (xx) y (xxx) de la etapa (d), respectivamente, y para la secuencia continua de caracteres que se encuentra a la derecha del carácter o la secuencia continua más larga de caracteres excluida en las sub-etapas (ii) y (xii) de la etapa y sub-etapas (xx) y (xxx) de la etapa (d),

respectivamente; y

donde en la sub-etapa (v) de la etapa (c) y la sub-etapa (xxiii) de la etapa (d) de la presente invención, la selección se repite preferentemente de forma simultánea para la secuencia continua de caracteres a la izquierda inmediata de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa, y para la secuencia continua de caracteres a la derecha inmediata de cada carácter o de cada secuencia continua más larga de caracteres excluida en la etapa previa.

La presente divulgación también incluye un kit para cuantificar el nivel de EMR en un sujeto que ha sido tratado para dicha enfermedad. De acuerdo con lo anterior, dicho kit comprende:

- medios para amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k, por célula diploide de dicha muestra biológica; y
- medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres de lectura de izquierda a derecha.

De forma análoga, dicho kit comprende:

- medios para amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de lectura de caracteres de izquierda a derecha.

Los medios para la amplificación de una secuencia nucleotídica mediante reacción en cadena de polimerasa comprende un instrumento de PCR que funciona según cualquiera de las técnicas anteriores (preferentemente un instrumento de PCR de emulsión) y cebadores según la anteriormente mencionada divulgación. De forma análoga, los medios para la secuenciación de al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres que se leen de izquierda a derecha comprende un instrumento de secuenciación de nucleótidos que funciona según cualquiera de las técnicas anteriormente mencionadas (preferentemente un instrumento de secuenciación paralela masiva). Cada uno de los medios para la amplificación y secuenciación de dicha secuencia de nucleótidos en (a) y (b) puede estar comprendido en distintos instrumentos. Como alternativa, estos medios pueden estar comprendidos dentro del mismo instrumento.

Además, el kit de la presente divulgación comprende medios (c) para determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), donde un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

(i) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(v)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(viii) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

(ix) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

o mediante:

(xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

5 (xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(xvi) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y

10 - el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

15 (xvii) sumar

- C_c ; y

- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y

20 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

25 para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y (xviii) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

30 Preferentemente dichos medios (c) determinan el DS, C_c y C_t proporcionando instrucciones para cada una de las etapas (i) a (xi) a (xviii). Además, preferentemente, dichos medios (c) llevan a cabo las etapas (i) a (x) a (xviii) tal como se ha desvelado anteriormente en el presente documento. En una realización, dichos medios son preferentemente al menos un ordenador, al menos un circuito, al menos un circuito integrado, al menos un chip o al menos un microchip.

35 Además, el kit de la presente divulgación comprende medios (d) para seleccionar, para cada primera lista de caracteres obtenida en la etapa (a), el DS del valor más alto, DS_{HV} , según lo anterior. El kit de la presente divulgación también comprende medios (e) para sumar el número de primeras listas de caracteres que tienen un DS_{HV} superior a un valor umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres. Por otra parte, el kit de la presente divulgación comprende medio (d) para sumar

- L_c ; y

- el número de primeras listas de caracteres que no tienen un DS_{HV} que es superior que T ,

45 para obtener el número total de primeras listas de caracteres, L_t . Además, el kit de la presente invención comprende medio (g) para calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_t^2$$

50 De la misma manera que los medios (c), dichos medios (d) a (g) preferentemente proporcionan instrucciones para cada una de las etapas desveladas en el presente documento. Por lo tanto, preferentemente, el kit de la presente invención comprende adicionalmente instrucciones para calcular el nivel de EMR. El nivel de EMR se calcula según la divulgación anteriormente mencionada de las etapas (c) a (g). Dichas instrucciones se llevan a cabo preferentemente por un operario humano o al menos un ordenador, al menos un circuito, al menos un circuito integrado, al menos un chip o al menos un microchip. En una realización preferente, dichas instrucciones se llevan a cabo por medios (c) siguiendo la entrada de listas de caracteres en dichos medios. En una realización más preferente del método de la invención, las etapas (a) a (g) se llevan a cabo por los mismos medios, en los que dichos medios comprenden al menos un ordenador, al menos un circuito, al menos un circuito integrado, al menos un chip o al menos un microchip. Por lo tanto, en una realización más preferente del kit de la invención, los medios (a), (b) y (c) son parte de los mismos medios, que comprende opcionalmente instrucciones (d), (e), (f) y (g).

65 La presente invención también se refiere al uso del método de la invención o del kit de la invención, según la divulgación anteriormente mencionada, en la cuantificación del nivel de enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha.

Ejemplos

Los siguientes ejemplos ilustran la invención y no deben considerarse como limitantes, sino como ilustrativos de la invención.

5 *Materiales y Métodos*

i) Muestras

10 Se usaron 16 muestras de médula ósea de diagnóstico y 24 muestras médula ósea de seguimiento obtenidas de sujetos con dicha enfermedad (muestras de ensayo) y la serie se expandió a 120 casos incluidos en dos ensayos clínicos consecutivos de pacientes con mieloma españoles.

ii) Extracción de ADN y cuantificación de la muestra

15 Se obtuvieron muestras de ADN de biopsias de médula ósea usando el kit QIAamp DNA mini kit (Qiagen). Se determinó la cantidad y calidad (pureza) de ADN usando NanoDrop 1000 (Thermo Scientific).

iii) Cálculo de concentración de ADN correspondiente a una cantidad deseada de células equivalentes por muestra

20 La concentración de ADN ([ADN], pg/μl) en la muestra biológica obtenida de un sujeto tratado para dicha enfermedad (muestra de seguimiento o de ensayo) se midió usando Qubit® dsDNA HS Assay Kit. Este valor se usó para determinar el número de células equivalentes por microlitro (N) de la muestra de ensayo según la siguiente fórmula:

$$25 \quad N = [\text{ADN}] / k$$

en el que N y [ADN] son tal como se describen anteriormente y k se refiere a un valor de 6,49 pg de ADN por célula diploide.

30 El número de células equivalentes por microlitro (N) de la muestra de ensayo permitió el cálculo del volumen de la muestra (V, μl) que es necesario para usar en la RCP para alcanzar una sensibilidad (S) deseada según la siguiente fórmula:

$$35 \quad V = 1 / (N \times S)$$

La sensibilidad se determinó mediante diluciones en serie de muestras (empezando a partir de 1 μg de ADN que se usó para garantizar la sensibilidad de 10⁻⁵ o más) con una cuantificación conocida de EMR sobre un origen policlonal (usando un grupo de controles sanos). Una sensibilidad de 10⁻⁵ equivale a la que se puede lograr a partir del uso de ADN a partir de al menos 100.000 células equivalentes. El volumen de la muestra (V) determina la cantidad de experimentos de PCR necesarios para obtener suficiente ADN, puesto que cada experimento de PCR realizado en tubos (tubos Eppendorf) con un volumen final de 100 μl permite usar hasta 8 μl de la muestra. Por ejemplo, si la cantidad de células equivalentes por microlitro (N) es 6500 y se desea una sensibilidad (S) de 10⁻⁵, sería necesario usar 1 / (6500 x 10⁻⁵) = 15,4 μl de muestra de ensayo en la PCR y puesto que pueden usarse hasta 8 μl por experimento de RCP, sería necesarios dos experimentos de PCR.

iv) PCR de las muestras

50 Se llevó a cabo la PCR usando Platinum® Taq DNA Polymerase High Fidelity (Life Technologies) y los cebadores que se usaron fueron los descritos en el protocolo de BIOMED-2 para IgH (CDR1/CDR2/CDR3 and DH) y IgK (KVJ-JK, KVJ-KDEL, INTR-KDEL), puesto que estos fragmentos cubren más del 90 % de los casos (Van Dongen Leukemia 2003). Los cebadores usados para amplificar dichos fragmentos del gen IgH se muestran en la **Tabla 1**, mientras que los cebadores usados para amplificar dichos fragmentos del gen IgK se muestran en la **Tabla 2**.

55

Tabla 1: Cebadores para la amplificación de fragmentos de IgH

Cebadores de familia	Tubo	Identificador de secuencia	Nombre del cebador	Secuencia	Sentido
V _H	Tubo A (CDR1)	SEQ ID NO:1	V _H 1-FR1	GGCCTCAGTGAAGGTCTCCTGCAAG	Directo
		SEQ ID NO:2	V _H 2-FR1	GTCTGGTCCTACGCTGGTGAAACCC	
		SEQ ID NO:3	V _H 3-FR1	CTGGGGGGTCCCTGAGACTCTCCTG	
		SEQ ID NO:4	V _H 4-FR1	CTTCGGAGACCCTGTCCCTCACCTG	
		SEQ ID NO:5	V _H 5-FR1	CGGGGAGTCTCTGAAGATCTCCTGT	
		SEQ ID NO:6	V _H 6-FR1	TCGCAGACCCTCTCACTCACCTGTG	
	Tubo B (CDR2)	SEQ ID NO:7	V _H 1-FR2	CTGGGTGCGACAGGCCCTGGACAA	
		SEQ ID NO:8	V _H 2-FR2	TGGATCCGTCAGCCCCCAGGGAAGG	
		SEQ ID NO:9	V _H 3-FR2	GGTCCGCCAGGCTCCAGGGAA	
		SEQ ID NO: 10	V _H 4-FR2	TGGATCCGCCAGCCCCCAGGGAAGG	
		SEQ ID NO: 11	V _H 5-FR2	GGGTGCGCCAGATGCCCGGGAAAGG	
		SEQ ID NO: 12	V _H 6-FR2	TGGATCAGGCAGTCCCCATCGAGAG	
		SEQ ID NO: 13	V _H 7-FR2	TTGGGTGCGACAGGCCCTGGACAA	
	Tubo C (CDR3)	SEQ ID NO: 14	V _H 1-FR3	TGGAGCTGAGCAGCCTGAGATCTGA	
		SEQ ID NO: 15	V _H 2-FR3	CAATGACCAACATGGACCCTGTGGA	
		SEQ ID NO: 16	V _H 3-FR3	TCTGCAAATGAACAGCCTGAGAGCC	
		SEQ ID NO: 17	V _H 4-FR3	GAGCTCTGTGACCGCCGCGGACACG	
	Tubo C (CDR3)	SEQ ID NO: 18	V _H 5-FR3	CAGCACCGCCTACCTGCAGTGGAGC	
		SEQ ID NO: 19	V _H 6-FR3	GTTCTCCCTGCAGCTGAACTCTGTG	
		SEQ ID NO: 20	V _H 7-FR3	CAGCACGGCATATCTGCAGATCAG	

(continuación)

Cebadores de familia	Tubo	Identificador de secuencia	Nombre del cebador	Secuencia	Sentido
D_H	Tubo D	SEQ ID NO: 21	D _{H1}	GGCGGAATGTGTGCAGGC	
		SEQ ID NO: 22	D _{H2}	GCACTGGGCTCAGAGTCCTCT	
		SEQ ID NO: 23	D _{H3}	GTGGCCCTGGGAATATAAAA	
		SEQ ID NO: 24	D _{H4}	AGATCCCCAGGACGCAGCA	
		SEQ ID NO: 25	D _{H5}	CAGGGGGACACTGTGCATGT	
		SEQ ID NO: 26	D _{H6}	TGACCCCAGCAAGGGAAGG	
	Tubo E	SEQ ID NO: 27	D _{H7}	CACAGGCCCCCTACCAGC	
J_H	Tubos A- E	SEQ ID NO: 28	JH57	CTTACCTGAGGAGACGGTGACC	Inverso

Tabla 2: Cebadores para la amplificación de fragmentos de IgK

Cebadores de familia	Tubo	Identificador de secuencia	Nombre del cebador	Secuencia	Sentido
V_k	Tubo F, G	SEQ ID NO:29	V _{k1f/6}	TCAAGGTTTCAGCGGCAGTGGATCTG	Directo
		SEQ ID NO:30	V _{k2f}	GGCCTCCATCTCCTGCAGGTCTAGTC	
		SEQ ID NO:31	V _{k3f}	CCCAGGCTCCTCATCTATGATGCATCC	
		SEQ ID NO:32	V _{k4}	CAACTGCAAGTCCAGCCAGAGTGTFFF	
		SEQ ID NO:33	V _{k5}	CCTGCAAAGCCAGCCAAGACATTGAT	
		SEQ ID NO:34	V _{k6}	GACCGATTTACCCCTCACAATTAATCC	
J_k	Tubo F	SEQ ID NO:35	J _{k1-4}	CTTACGTTTGATCTCCACCTTGGTCCC	Inverso
		SEQ ID NO:36	J _{k5}	CTTACGTTTAATCTCCAGTCGTGTCCC	
KDEL	Tubo G, H	SEQ ID NO:37	KDEL	CCTCAGAGGTCAGAGCAGGTTGTCCTA	
J_k C_k Intron	Tubo H	SEQ ID NO:38	INTR	CGTGGCACCGCGAGCTGTAGAC	Directo

5

La amplificación de la muestra de ensayo se realizó usando la cantidad de experimentos de PCR (es decir, tubos de PCR) que se calcularon necesarios basándose en el volumen de la muestra de ensayo (V, µl), el número de células equivalentes por microlitro (N) de dicha muestra, la sensibilidad (S) que era deseada alcanzar y el volumen del tubo de PCR final. En consecuencia, las cantidades de cada componente (por tubo de PCR) para cada mezcla de reacción de PCR para la muestra de ensayo fue como sigue:

10

a) PCR CDR1/CDR2/CDR3

- 15 84 µl de mezcla maestra de Platinum HIFI
 4 µl de mezcla de cebadores CDR1/CDR2/CDR3 (Tubos A, B, C)
 4 µl de cebador JH57
 8 µl de ADNg

b) PCR KVJ

- 5 80 µl de mezcla maestra de Platinum HIFI
 4 µl de mezcla KV de cebadores (Tubo F)
 4 µl de cebador KJ5
 4 µl de cebador KJ1-4
 8 µl de ADNg

c) PCR DH

- 10 84 µl de mezcla maestra de Platinum HIFI
 4 µl de cebadores DH 1-6 (Tubo D) o DH7 (Tubo E)
 4 µl de JH57
 8 µl de ADNg

d) PCR KDEL

- 20 80 µl de mezcla maestra de Platinum HIFI
 4 µl de mezcla KV de cebadores (Tubo G)
 4 µl INTR (Tubo H)
 4 µl de KDEL
 8 µl de ADNg

25 La muestra de diagnóstico se amplifica con las mismas reacciones, pero usando 1 µl de ADN (1 µl de ADN = aproximadamente 20ng de ADNg, es decir [ADN] = 20 µg/ml) ya que no es necesario alcanzar una sensibilidad dada en la muestra de diagnóstico.

v) Preparación de bibliotecas de amplicones sin fragmentación

30 Los productos amplificados de ambas muestras (de diagnóstico y seguimiento) se usaron para preparar bibliotecas de amplicones respectivas sin fragmentación usando Ion Plus Fragment Library kit y Agencourt Ampure XP (Thermo-Fisher). Se encontró que era posible usar la mitad de los volúmenes de los reactivos usando el Ion Plus Fragment Library kit. La concentración de la biblioteca final se determinó usando PCRq en el GeneRead Library Quant kit (Qiagen). Las bibliotecas se generaron usando códigos de barras específicos para cada muestra.

vi) Secuenciación paralela masiva

40 El clon o clones principales se identificaron en la muestra de diagnóstico mediante secuenciación paralela masiva del producto de la amplificación de CDR1, CDR2, CD3, KVJ, DH y KDEL. Las muestras clonales con fragmentos superiores a 250 pb se secuenciaron sobre la plataforma de PGM (plataforma de Ion Torrent Personal Genome Machine™) usando OneTouch™ Ion v2 Kit Template 400 DL, 400 Ion PGM™ Sequencing Kit v2 y Ion Chip 318™ Kit según las instrucciones del fabricante (Thermo-Fisher). Los fragmentos inferiores a 250 pb (principalmente de muestras con los reordenamientos KVJ y KDEL) se secuenciaron usando la plataforma PROTON: Ion Proton™ I emulsion OT2 Template Kit y sequencing Ion Proton™ I Sequencing Kit (Thermo-Fisher). Todos los reactivos se compraron de Lifetech usando sus protocolos con ligeras modificaciones: la tecnología de la plataforma PGM secuencia fragmentos de hasta 250 pb, pero es posible secuencias fragmentos de hasta 400 pb con otro kit comercial de Lifetech usando química distinta.

vii) Análisis bioinformático

50 Después de la secuenciación, se obtuvieron archivos FASTQ de las dos muestras a partir del buscador Torrent según el código de barras correspondiente. Cada archivo FASTQ comprende una lista de caracteres que se leen de izquierda a derecha que representan la secuencia nucleotídica del ADN comprendido en dicha muestra y adicionalmente comprenden la puntuación de calidad correspondiente a cada carácter de dicha lista de caracteres.

55 La cuantificación de cada secuencia o secuencias clonotípicas en la muestra de diagnóstico se determinó usando métodos matemáticos y computacionales (herramientas IT), a saber, usando el FrequencyRank.sh Bourne shell script (frequency_rank.sh) para clasificar secuencias en orden de frecuencia descendiente. Una vez se determinaron las secuencias que son las mismas en la muestra de diagnósticos como en la muestra de seguimiento, se generó un archivo adn que comprendía cada una de dichas secuencias clonales como una lista de caracteres que se leían de izquierda a derecha y que tenía una cantidad total de caracteres

65 El número de secuencias clonales (primeras listas de caracteres) identificadas en la muestra de diagnóstico que se consideraron las mismas que la secuencia de argumento (segundas listas de caracteres se contaron usando el programa SeqSearchFastq.java con la opción de ajuste y una relación de coincidencia (grado de similitud) de 0,99, para proporcionar un valor, L_c . L_t se determinó a partir del número total de primeras listas de caracteres.

Como el método de la invención implica una mezcla de alineación y comparación, la comparación se realizó solo entre la primera y la última posición de coincidencia y la opción de ajuste indica que el proceso actúe de este modo, limitando la comparación desde la primera y las últimas posiciones de coincidencia en lugar de la primera y última posiciones (independientemente de la coincidencia) en la secuencia de la muestra. El resultado, L_c , del programa SeqSearchFastq.java se usó posteriormente, junto con los valores para L_t , k y D , para calcular la EMR.

Ejemplo 1. Cuantificación de EMR en mieloma múltiple usando secuenciación paralela masiva de genes de inmunoglobulinas

Lo que sigue presenta un método para la cuantificación de secuencias clonotípicas tumorales dentro de reordenamientos de origen policlonal de genes de inmunoglobulinas (Ig) mediante secuenciación paralela masiva (MPS). La detección de reordenamiento clonal en neoplasmas de linfocitos B y T permite controlar la evolución de estas patologías. Para cuantificar estos reordenamientos en linfocitos B, se usaron los cebadores que se describen en las **Tablas 1 y 2** para CDR3, VDJ, IgH, IgK, KVJ, KDEL y IgL, puesto que estos fragmentos cubren más del 90 % de los casos (Van Dongen, Leukemia 2003). La selección de estos reordenamientos particulares se debe al diseño de los cebadores que solo amplifican secuencias cortas (inferiores a 200 pb); permitiendo secuencias estos fragmentos en la plataforma PROTON, capaz de 10 Gb.

Los pacientes negativos en el diagnóstico de VDJ, IgH, GDR3, KVJ, KDEL pueden secuenciarse con el resto de cebadores BIOMED como IgH, VDJ, CDRI y IκL DJ. Como el tamaño de estos fragmentos es de entre 300 y 400 pares base (pb), es necesario usar la plataforma PGM con el kit reactivo para 400 pb. La capacidad del PGM es casi 1 Gbase, mediante la cual la capacidad se refiere a la cantidad de bases que pueden leerse en una puesta en marcha de PGM.

Resultados:

Diluciones en serie de muestras experimentales clonotípicas indicaron una sensibilidad de 10^{-5} para 150.000 células. La correlación de todas las muestras con los datos de la citometría de flujo logró un $R = 0,59$ (Pearson = 0,765, $p < 0,0001$) y $R = 0,51$ (Pearson = 0,716, $p < 0,0001$) para las muestras de seguimiento. El promedio de las lecturas de los cebadores BIOMED fue similar a las frecuencias de los distintos fragmentos descritos en mieloma múltiple. La reproducibilidad de la técnica fue más de 90 %. De los 24 seguimientos analizados, dos fueron positivos por secuenciación paralela masiva y negativos por citometría de flujo y uno fue negativo por secuenciación paralela masiva y positivo por citometría de flujo. En la plataforma de PGM, se logró el análisis en paralelo de 12 seguimientos en una semana con una cobertura de 500,000x y a un precio aproximado de 100 Euros/muestra.

En la **Figura 3**, se muestra la sensibilidad lograda en diluciones en serie de reordenamientos clonales de mieloma múltiple sobre un origen policlonal obtenido a partir de pacientes sanos o a partir de líneas celulares que no son de linfocitos B, . La sensibilidad en la caracterización de los reordenamientos génicos de inmunoglobulinas se determina por la cantidad de ADN de entrada o el número de células equivalente. La sensibilidad de esta técnica depende de la cantidad de variantes encontrada en una lectura clonal y muestra que el método de la invención muestra una sensibilidad extremadamente alta en la detección de EMR.

La figura 4 muestra la correlación de reordenamiento clonal principal en muestras de paciente que han sido tratados para mieloma múltiple (muestras de seguimiento) entre datos de EMR medidos con citometría de flujo (eje y) y secuenciación paralela masiva según la presente invención (eje x). La cuantificación de EMR en mieloma múltiple realizada mediante citometría de flujo se refiere a la cantidad celular total, con una ecuación equivalente a la de esta invención. La correlación entre las dos técnicas es alta, con $R^2 = 0,51$.

Conclusiones:

Secuenciación profunda de los reordenamientos de los genes Ig por tecnología Ion Torrent es una técnica eficaz para definir y cuantificar los clones patológicos en mieloma múltiple. Esta técnica es una alternativa metódica y económicamente viable a la citometría de flujo y a otros métodos de control de la EMR.

Por lo tanto, la cuantificación con alta sensibilidad de las secuencias específicas que pertenecen a clones patológicos que definen la afección permite el control de la evolución y respuesta celular a tratamientos específicos, la definición de un nuevo foco de enfermedad y el control de enfermedad mínima residual en pacientes con alteraciones genéticas definidas.

Ejemplo 2. Cuantificación de SNV, MNV e índices

El método descrito a continuación es aplicable a la detección de cualquier tipo de mutación, dadas algunas limitaciones, del siguiente modo. El error promedio basado en plataformas de secuenciación masiva es 0,5 % o, en otras palabras, una lectura errónea por cada 200 para cada posición en el genoma. La probabilidad de que ocurra un error en la lectura de la variante buscada es del 0,5 % / 4 bases o aproximadamente del 0,1 %. Esta limitación teórica se ha verificado experimentalmente para mutaciones puntuales (SNV: DN3A y IDH2) en casos de LMA

(leucemia mieloide aguda, **Figura 5**), donde este error es del 0,1 % para cada posición. En esas mutaciones que incluyen más de dos posiciones para leer, tales como una mutación múltiple (MNV) o un indel, el error será $(0,1 \times n)\%$, donde n es el número de variantes clonales presentes en la lectura.

5 *Ejemplo 3. Cuantificación de inserciones largas y translocaciones*

En este caso el límite de sensibilidad no se alcanza debido al hecho de que no existe un origen frente al cual comparar las lecturas, ya que los cebadores amplifican solo aquellos fragmentos de ADN que tienen una translocación o inversión. Para aliviar este problema, se usó un control de ADN, siendo este uno de los genes implicados en la translocación, en su forma de tipo silvestre. Por lo tanto, la relación de secuencias clonotípicas / secuencias totales tiene en cuenta la cantidad de lecturas en ambos.

LISTADO DE SECUENCIAS

15 <110> Fundación de Investigación Hospital 12 de Octubre
 <120> Método para cuantificar el nivel de enfermedad mínima residual de un sujeto
 <130> EP-06895
 20 <160> 38
 <170> BiSSAP 1.3
 25 <210> 1
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Cebador VH1-FR1
 <400> 1
 ggcctcagtg aaggtctcct gcaag 25
 35 <210> 2
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> Cebador VH2-FR1
 <400> 2
 45 gtctggtcct acgctggtga aaccc 25
 <210> 3
 <211> 25
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 <223> Cebador VH3-FR1
 55 <400> 3
 ctggggggtc cctgagactc tcctg 25
 <210> 4
 60 <211> 25
 <212> ADN
 <213> Secuencia artificial
 <220>
 65 <223> Cebador VH4-FR1

<400> 4
 cttcggagac cctgtccctc acctg 25

5 <210> 5
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador VH5-FR1

<400> 5
 cggggagtct ctgaagatct cctgt 25

15 <210> 6
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador VH6-FR1

<400> 6
 tcgcagaccc tctcactcac ctgtg 25

25 <210> 7
 <211> 25
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador VH1-FR2

<400> 7
 ctgggtgcga caggcccctg gacaa 25

35 <210> 8
 <211> 25
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador VH2-FR2

<400> 8
 tggatccgtc agccccagg gaagg 25

45 <210> 9
 <211> 21
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> Cebador VH3-FR2

<400> 9
 ggtccgccag gctccaggga a 21

55 <210> 10
 <211> 25
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> Cebador VH4-FR2

65

<400> 10
 tggatccgcc agccccagg gaagg 25

5 <210> 11
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador VH5-FR2

<400> 11
 ggggtgcgcca gatgccggg aaagg 25

15 <210> 12
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador VH6-FR2

<400> 12
 tggatcaggc agtccccatc gagag 25

25 <210> 13
 <211> 25
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador VH7-FR2

<400> 13
 ttgggtgcga caggcccctg gacaa 25

35 <210> 14
 <211> 25
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador VH1-FR3

45 <400> 14
 tggagctgag cagcctgaga tctga 25

50 <210> 15
 <211> 25
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador VH2-FR3

<400> 15
 caatgaccaa catggaccct gtgga 25

60 <210> 16
 <211> 25
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Cebador VH3-FR3

<400> 16
 tctgcaaatg aacagcctga gagcc 25

5 <210> 17
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador VH4-FR3

<400> 17
 gagctctgtg accgccgagg acacg 25

15 <210> 18
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador VH5-FR3

<400> 18
 cagcaccgcc tacctgcagt ggagc 25

25 <210> 19
 <211> 25
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador VH6-FR3

<400> 19
 gttctccctg cagctgaact ctgtg 25

35 <210> 20
 <211> 24
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador VH7-FR3

45 <400> 20
 cagcacggca tatctgcaga tcag 24

50 <210> 21
 <211> 18
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador DH1

<400> 21
 ggccgaatgt gtgcaggc 18

60 <210> 22
 <211> 21
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Cebador DH2

<400> 22
 gcactgggct cagagtctc t 21

5 <210> 23
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador DH3

<400> 23
 gtggccctgg gaatataaaa 20

15 <210> 24
 <211> 19
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador DH4

<400> 24
 agatccccag gacgcagca 19

25 <210> 25
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador DH5

<400> 25
 cagggggaca ctgtgcatgt 20

35 <210> 26
 <211> 19
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador DH6

45 <400> 26
 tgaccccagc aaggaagg 19

50 <210> 27
 <211> 18
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> Cebador DH7

<400> 27
 cacaggcccc ctaccagc 18

60 <210> 28
 <211> 22
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Cebador JH57

<400> 28
 cttacctgag gagacggtga cc 22

5 <210> 29
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador VK1f/6

<400> 29
 tcaaggttca gcggcagtg atctg 25

15 <210> 30
 <211> 26
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador VK2f

<400> 30
 ggctccatc tctgcaggt ctagtc 26

25 <210> 31
 <211> 27
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador VK3f

<400> 31
 cccaggctcc tcatctatga tgcattc 27

35 <210> 32
 <211> 27
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador VK4

45 <400> 32
 caactgcaag tccagccaga gtgtttt 27

<210> 33
 <211> 26
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> Cebador VK5

55 <400> 33
 cctgcaaagc cagccaagac attgat 26

60 <210> 34
 <211> 27
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Cebador VK6

<400> 34
 gaccgatttc accctcacia ttaatcc 27

5 <210> 35
 <211> 27
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Cebador JK1-4

<400> 35
 cttacgtttg atctccacct tggccc 27

15 <210> 36
 <211> 27
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador JK5

<400> 36
 cttacgttta atctccagtc ggtccc 27

25 <210> 37
 <211> 27
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador Kde

<400> 37
 cctcagaggt cagagcaggt tgccta 27

35 <210> 38
 <211> 22
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Cebador INTR

45 <400> 38
 cgtggcaccg cgagctgtag ac 22

REIVINDICACIONES

1. Un método para cuantificar el nivel de enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad, que comprende:

- 5
- (a)
- 10 - amplificar mediante reacción en cadena de polimerasa usando cebadores, al menos una secuencia de nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, donde el ADN genómico tiene un peso promedio, k, por célula diploide de dicha muestra biológica; y
- secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de caracteres de lectura de izquierda a derecha;
- 15 (b)
- amplificar mediante reacción en cadena de polimerasa usando los mismos cebadores que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- 20 - secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de caracteres de lectura de izquierda a derecha;
- (c) determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), donde un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:
- 25
- (i) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;
- 30 (ii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- 35 (iii)
- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 40 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 45 (iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- 50 (v)
- 55 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 60 - selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 65

- (vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- 5 (vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;
- (viii) sumar
- 10 - el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)
- 15 para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;
- (ix) sumar
- 20 - C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y
- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que
- 25 no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),
- para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
- (x) calcular el DS según la siguiente fórmula:
- 30
$$DS = C_c / C_t$$
- o mediante:
- 35 (xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;
- (xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los
- 40 mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- (xiii)
- 45 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y
- 50 - selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;
- 55 (xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;
- (xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de
- 60 caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;
- (xvi) sumar
- 65 - el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y
- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas

(xi) a (xv)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;
5 (xvii) sumar

- C_c ; y
- el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
10 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

15 para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
(xviii) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

20 (d) seleccionar, para cada primera lista de caracteres obtenida en la etapa (a), el DS del valor más alto, DS_{HV} ;
(e) sumar el número de primeras listas de caracteres que tienen un DS_{HV} superior a un valor umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;
25 (f) sumar

- L_c ; y
- el número de primeras listas de caracteres que no tienen un DS_{HV} superior a T , para obtener el número total de primeras listas de caracteres, L_t ; y

30 (g) calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_t^2.$$

35 2. El método de acuerdo con la reivindicación 1, en el que

- cada carácter en cada una de la primera y segunda listas de caracteres respectivamente obtenidas en las etapas (a) y (b), comprende una letra asociada con un número o símbolo, donde dicho número o símbolo representa la calidad (Q), donde la calidad (Q) es un número entero que representa la probabilidad de que la letra que representa un nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica sea incorrecta y donde dicha letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica que tiene la mayor calidad (Q); y
40 - un carácter en la primera lista de caracteres se determina en la etapa (c) como el mismo que el carácter en la segunda lista de caracteres, cuando la letra que tiene la mayor calidad es la misma en la primera y segunda listas de caracteres.

45 3. El método de acuerdo con la reivindicación 2, en el que un carácter en la primera lista de caracteres se determina en la etapa (c) como el mismo que el carácter en la segunda lista de caracteres, cuando la letra que tiene la calidad más alta (Q) es la misma en la primera y segunda lista de caracteres, y la calidad de la letra de la primera lista de caracteres se encuentra dentro del 1 por ciento de la calidad de la letra en la segunda lista de caracteres.

50 4. El método de acuerdo con cualquiera de las anteriores reivindicaciones 1 a 3, en el que la secuenciación es secuenciación paralela masiva.

55 5. El método de acuerdo con cualquiera de las anteriores reivindicaciones 1 a 4, en el que una EMR de 1 es 100 % indicativo de enfermedad en dicho sujeto y una EMR de 0 es 0% indicativo de enfermedad en dicho sujeto.

6. El método de acuerdo con cualquiera de las anteriores reivindicaciones 1 a 5, en el que la enfermedad se selecciona entre cáncer o leucemia.

60 7. El método de acuerdo con la reivindicación 6, en el que la enfermedad se selecciona entre leucemia linfoblástica aguda, leucemia mieloide aguda, leucemia linfocítica crónica, leucemia mielógena crónica, linfoma folicular, linfoma de células del manto, mieloma múltiple, cáncer de mama o neuroblastoma.

65 8. Un kit para la cuantificación del nivel de enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad que comprende:

(a)

- 5 - medios para amplificar en un instrumento de PCR, mediante reacción en cadena de polimerasa, al menos una secuencia de nucleótidos comprendida en una cantidad, D, de ADN genómico de una muestra biológica obtenida a partir de dicho sujeto después del tratamiento para dicha enfermedad, en el que el ADN genómico tiene un peso promedio, k, por célula diploide de dicha muestra biológica, en el que dichos medios comprenden cebadores directos específicos de locus y cebadores inversos específicos de locus, para identificar una variante específica de una secuencia nucleotídica presente en la muestra biológica, en la que dicha una variante o la ausencia de dicha una variante es indicativo de enfermedad; y
- 10 - medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una primera lista de lectura de caracteres de izquierda a derecha, en el que la secuenciación comprende una plataforma de secuenciación paralela masiva;

(b)

- 15 - medios para amplificar, mediante reacción en cadena de polimerasa, en un instrumento de PCR, usando los mismos cebadores específicos de locus que en la etapa (a), al menos una secuencia de nucleótidos en una muestra biológica obtenida a partir de dicho sujeto antes del tratamiento de dicha enfermedad; y
- 20 - medios para secuenciar dicha al menos una secuencia de nucleótidos para obtener al menos una segunda lista de caracteres de lectura de izquierda a derecha, en el que la secuenciación comprende una plataforma de secuenciación paralela masiva;

(c) medios para determinar, para cada primera lista de caracteres obtenida en la etapa (a), el grado de similitud con cada segunda lista de caracteres obtenida en la etapa (b), en los que dichos medios comprenden preferentemente al menos un ordenador, al menos un circuito, al menos un circuito integrado, al menos un chip o al menos un microchip y en el que un grado de similitud, DS, de una primera lista de caracteres obtenida en la etapa (a) con una segunda lista de caracteres obtenida en la etapa (b) se determina bien mediante:

(i) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(ii) la exclusión del carácter o secuencia continua más larga de caracteres seleccionado en la etapa (i) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(iii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (ii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(iv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (iii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(v)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la izquierda de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra en la secuencia continua de caracteres inmediatamente a la derecha de cada carácter o cada secuencia continua más larga de caracteres excluidos en la etapa previa que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia

continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(vi) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (v) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(vii) repetición de las etapas (v) y (vi) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(viii) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii); y

- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (i) a (vii)

para obtener el número total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

(ix) sumar

- C_c ; y

- le número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c); y

- el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (i) a (vii) de la etapa (c),

para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y

(x) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

o mediante:

(xi) selección del carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres;

(xii) exclusión del carácter o secuencia continua más larga de caracteres seleccionados en la etapa (xi) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xiii)

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la izquierda del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la derecha en las listas de caracteres; y

- selección del carácter o secuencia continua más larga de caracteres que se encuentra a la derecha del carácter o secuencia continua más larga de caracteres excluidos en la etapa (xii) que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres, donde cuando son seleccionados dos o más caracteres o dos o más secuencias continuas más largas de la misma longitud, solo se selecciona el carácter o secuencia continua más larga de caracteres que se encuentra más a la izquierda en las listas de caracteres;

(xiv) exclusión de cada carácter y/o cada secuencia continua más larga de caracteres seleccionados en la etapa (xiii) de etapas posteriores de selección de un carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres;

(xv) repetición de las etapas (xii) y (xiv) hasta que ningún carácter o secuencia continua más larga de caracteres que son los mismos en la primera lista de caracteres y en la segunda lista de caracteres sea seleccionado;

(xvi) sumar

- el número de caracteres en la primera lista de caracteres que se excluyeron en cualquiera de las etapas (xi) a (xv); y

- el número de caracteres en la segunda lista de caracteres que se excluyeron en cualquiera de las etapas (xix) a (xxv) para obtener la cantidad total de caracteres, C_c , en la primera y segunda listas de caracteres que son los mismos en la segunda y primera lista de caracteres, respectivamente;

5 (xvii) sumar

- C_c ; y
 - el número de caracteres en la primera lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la primera lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c); y
 - el número de caracteres en la segunda lista de caracteres que se encuentra entre los caracteres y/o secuencias continuas más largas de caracteres que se excluyeron de la segunda lista de caracteres y que no se excluyeron en ninguna de las etapas (xi) a (xv) de la etapa (c),

15 para obtener el número total de caracteres, C_t , en la primera y segunda lista de caracteres; y
 (xviii) calcular el DS según la siguiente fórmula:

$$DS = C_c / C_t$$

20 (d) medios o instrucciones para seleccionar, para cada primera lista de caracteres obtenida en la etapa (a), el DS del valor más alto, DS_{HV} ;

(e) medios o instrucciones para sumar el número de primeras listas de caracteres que tienen un DS_{HV} que es superior a un valor umbral, T , para obtener el número total de primeras listas de caracteres, L_c , que son las mismas que una segunda lista de caracteres;

25 (f) medios o instrucciones para sumar

- L_c ; y
 - la cantidad de primeras listas de caracteres que no tienen un DS_{HV} que es superior a T ,

30 para obtener el número total de primeras listas de caracteres, L_t ; y

(g) medios e instrucciones para calcular el nivel de enfermedad mínima residual (EMR) según la siguiente fórmula:

$$EMR = L_c \times (D / k) / L_t^2$$

35 y en la que dichos medios de las etapas (a) a (g) son capaces de calcular la EMR o proporcionan instrucciones para calcular la EMR, sin la necesidad de acceder a bases de datos externas; y en la que dichos medios comprende: al menos un ordenador o al menos un circuito o al menos un circuito integrado o al menos un chip o al menos un microchip; y en el que dichas instrucciones pueden llevarse a cabo: por un operario humano o al menos un ordenador, o al menos un circuito o al menos un circuito integrado o al menos un chip o al menos un microchip.

9. El kit de acuerdo con la reivindicación 8, en el que

45 - cada carácter en cada una de las primeras y segunda listas de caracteres respectivamente obtenidas en las etapas (a) y (b), comprende una letra asociada con un número o símbolo, en la que dicho número o símbolo representa la calidad (Q), en la que la calidad (Q) es un número entero que representa la probabilidad de que la letra que representa un nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica sea incorrecta y en la que dicha letra representa el nucleótido que se identifica en la posición correspondiente en la secuencia nucleotídica que tiene la mayor calidad (Q); y
 50 - un carácter en la primera lista de caracteres se determina en (c) como el mismo que el carácter en la segunda lista de caracteres, cuando una letra que tiene la mayor calidad es la misma en la primera y segunda listas de caracteres.

55 10. El kit de acuerdo con la reivindicación 9, en el que un carácter en la primera lista de caracteres se determina como el mismo que el carácter en la segunda lista de caracteres, cuando la letra que tiene la calidad más alta (Q) es la misma en la primera y segunda lista de caracteres, y la calidad de la letra de la primera y segunda lista de caracteres se encuentra dentro del 1 por ciento de la calidad de la letra en la segunda lista de caracteres.

60 11. El kit de acuerdo con cualquiera de las anteriores reivindicaciones 8 a 10, en el que una EMR de 1 es 100 % indicativo de enfermedad en dicho sujeto y una EMR de 0 es 0% indicativo de enfermedad en dicho sujeto.

12. El kit de acuerdo con cualquiera de las anteriores reivindicaciones 8 a 11, en el que la enfermedad se selecciona de cáncer o leucemia.

65 13. El kit de acuerdo con la reivindicación 12, en el que la enfermedad se selecciona de leucemia linfoblástica aguda,

leucemia mieloide aguda, leucemia linfocítica crónica, leucemia mielógena crónica, linfoma folicular, linfoma de células del manto, mieloma múltiple, cáncer de mama o neuroblastoma.

5 14. El kit de acuerdo con la reivindicación 8 en el que los medios para la secuenciación usados en las etapas (a) y/o (b) se seleccionan a partir de: secuenciación de nucleótidos multiplex y/o de alto rendimiento; secuenciación masiva paralela de firmas (MPSS) y secuenciación paralela masiva usando PCR de emulsión.

10 15. Uso del método de acuerdo con cualquiera de las reivindicaciones 1 a 7 o el kit de acuerdo con cualquiera de las reivindicaciones 8 a 14 en la cuantificación del nivel de enfermedad mínima residual (EMR) en un sujeto que ha sido tratado para dicha enfermedad.

Figura 1

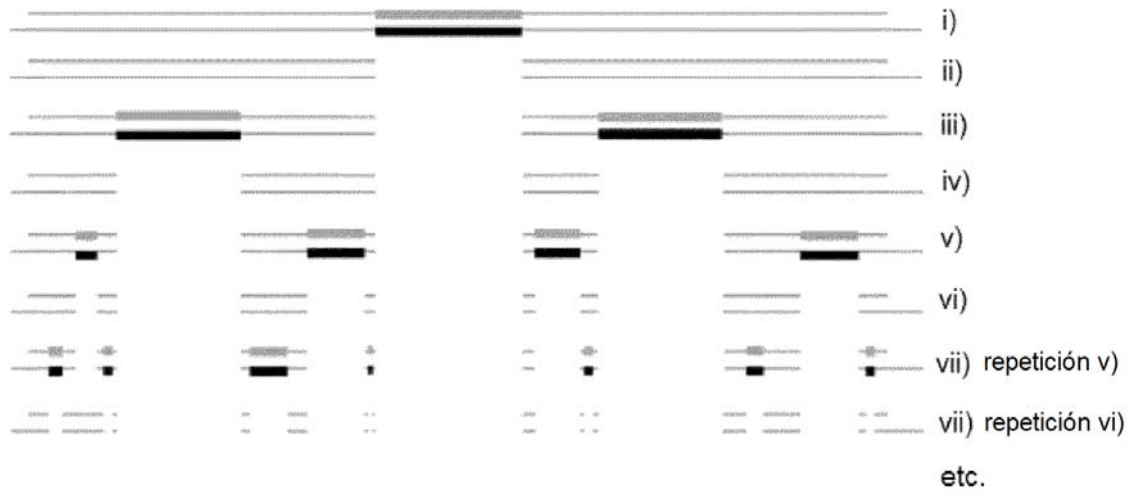


Figura 2

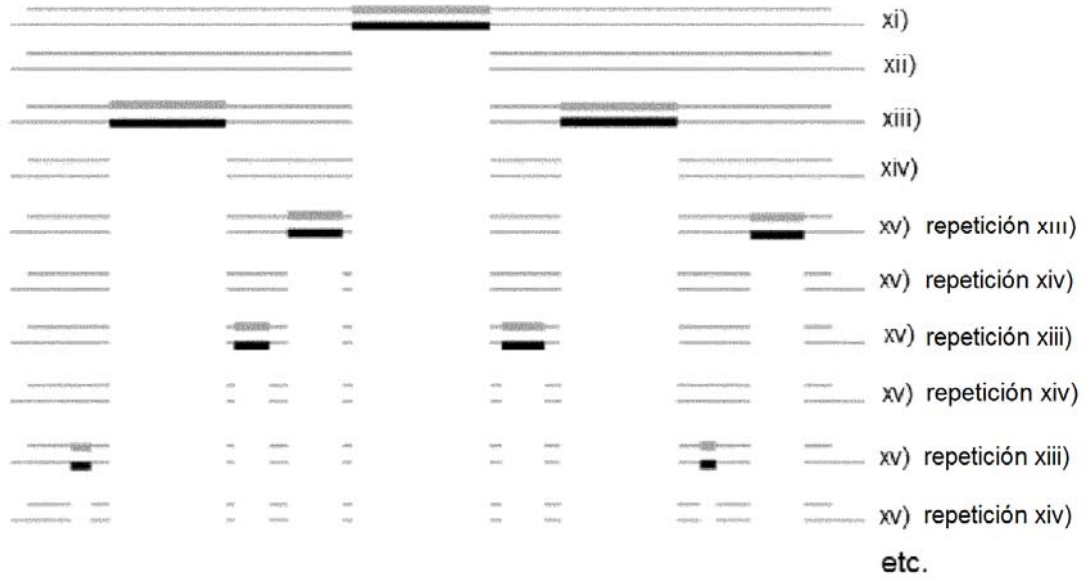


Figura 3

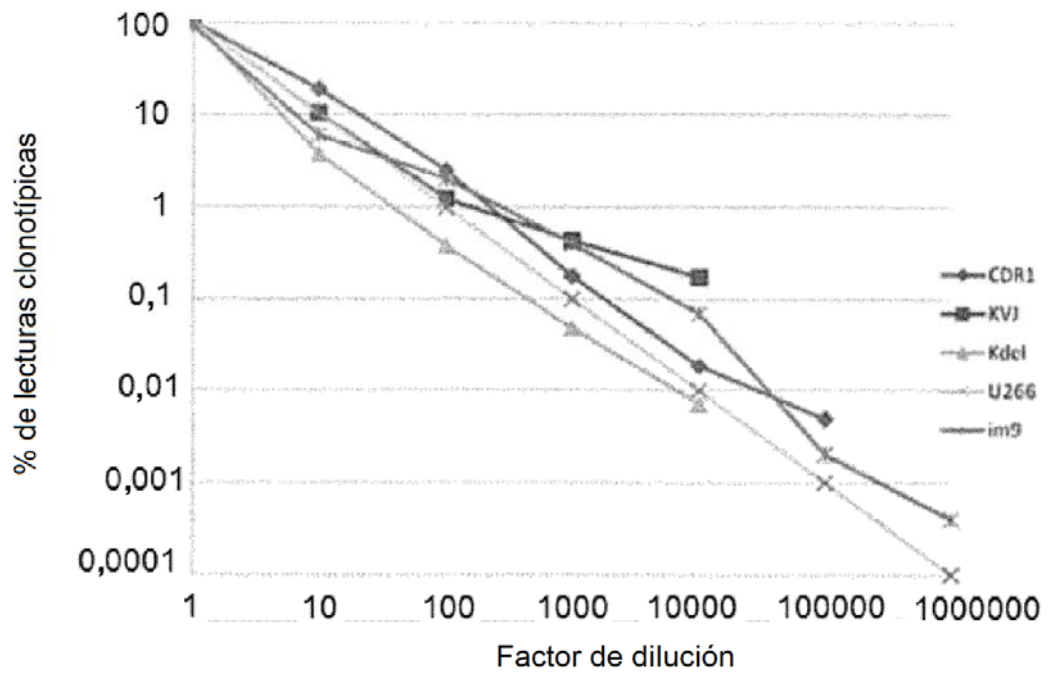


Figura 4

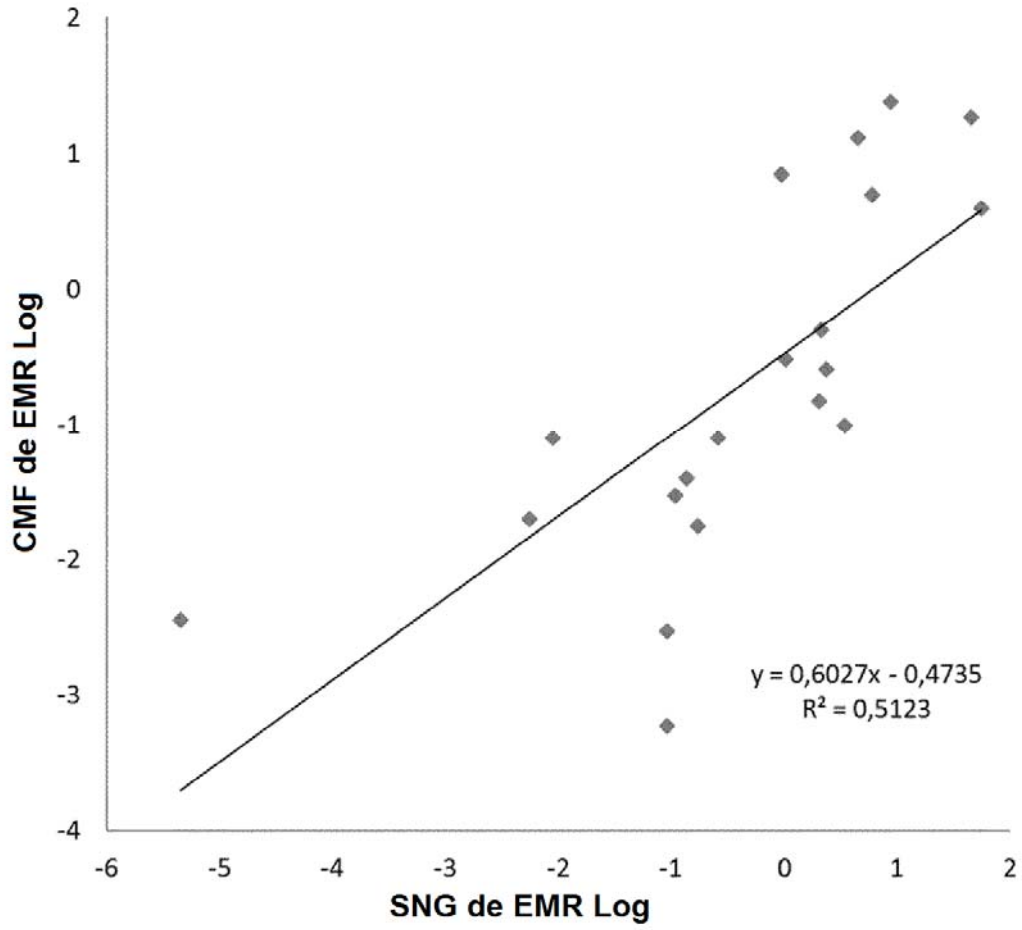
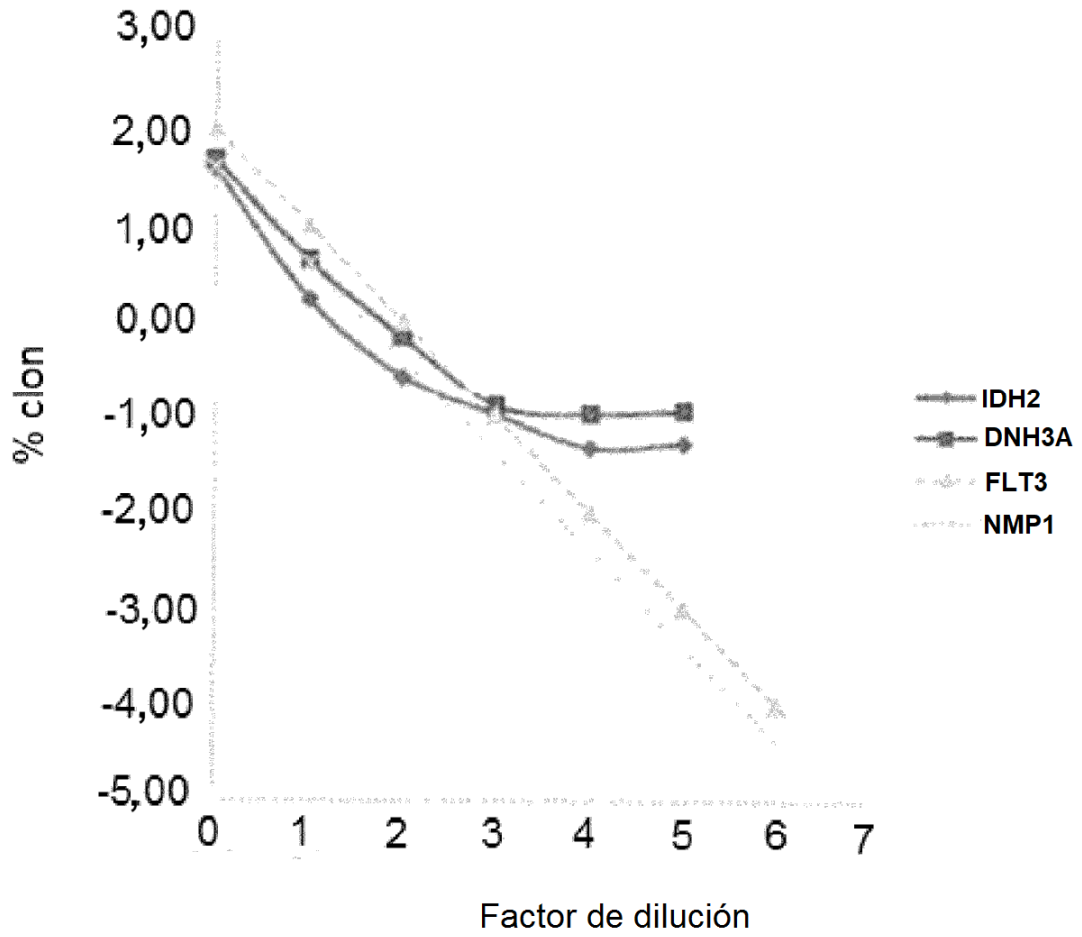


Figura 5



REFERENCIAS CITADAS EN LA DESCRIPCIÓN

La lista de referencias citadas por el solicitante es, únicamente, para conveniencia del lector. No forma parte del documento de patente europea. Si bien se ha tenido gran cuidado al compilar las referencias, no pueden excluirse errores u omisiones y la OEP declina toda responsabilidad a este respecto.

Documentos de patente citados en la descripción

- WO 2004033728 A2 [0004]
- US 7785783 B2 [0004]
- US 8628927 B2 [0005]
- WO 2013188471 A2 [0005]

10 **Literatura no patente citada en la descripción**

- **MARTINEZ-LOPEZ J. et al.** Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. *Blood*, 2014, vol. 123 (20), 3073-3079 [0004]
- **VIJ R. et al.** Deep sequencing reveals myeloma cells in peripheral blood in majority of multiple myeloma patients. *J. Clin. Lymphoma Myeloma Leuk.*, 2013, vol. 14 (2), 131-139 [0004]
- **LADETTO M. et al.** Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. *Leukemia*, 2014, vol. 28 (6), 1299-1307 [0005]
- **FAHAM et al.** *Blood*, 2012, vol. 120 (26), 5173-5180 [0005]
- **HAUWEL M. ; MATTHES T.** Minimal residual disease monitoring: the new standard for treatment evaluation of haematological malignancies?. *Swiss Med Wkly*, 2014, vol. 144, w13907 [0012]