

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 013 487**

51 Int. Cl.:

**G16B 20/00** (2009.01)

**C12Q 1/6827** (2008.01)

**C12Q 1/6886** (2008.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.11.2018** **PCT/US2018/059068**

87 Fecha y número de publicación internacional: **09.05.2019** **WO19090156**

96 Fecha de presentación y número de la solicitud europea: **02.11.2018** **E 18872946 (1)**

97 Fecha y número de publicación de la concesión europea: **22.01.2025** **EP 3704268**

54 Título: **Normalización de la carga de mutaciones tumorales**

30 Prioridad:

**03.11.2017 US 201762581563 P**

45 Fecha de publicación y mención en BOPI de la  
traducción de la patente:  
**14.04.2025**

73 Titular/es:

**GUARDANT HEALTH, INC. (100.00%)**  
**3100 Hanover Street**  
**Palo Alto, CA 94304, US**

72 Inventor/es:

**CHUDOVA, DARYA**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 3 013 487 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

Normalización de la carga de mutaciones tumorales

## 5 ANTECEDENTES

[0001] Un tumor es un crecimiento anormal de células. El ADN fragmentado suele liberarse en los fluidos corporales cuando las células, como las tumorales, mueren. Por lo tanto, parte del ADN libre de células en los fluidos corporales es ADN tumoral. Un tumor puede ser benigno o maligno. Un tumor maligno suele denominarse cáncer.

10

[0002] El cáncer es una de las principales causas de enfermedad en todo el mundo. Cada año, decenas de millones de personas son diagnosticadas con cáncer en todo el mundo y más de la mitad mueren a causa de esta enfermedad. En muchos países, el cáncer ocupa el segundo lugar como causa de muerte más común, después de las enfermedades cardiovasculares. La detección temprana se asocia con mejores resultados en muchos tipos de cáncer.

15

[0003] El cáncer suele ser causado por la acumulación de mutaciones dentro de las células normales de un individuo, al menos algunas de las cuales dan lugar a una división celular regulada de forma inadecuada. Dichas mutaciones suelen incluir variaciones de un solo nucleótido (SNV), fusiones, inserciones y deleciones de genes (indels), transversiones, translocaciones e inversiones. La cantidad de mutaciones dentro de un cáncer es un indicador de la susceptibilidad del cáncer a la inmunoterapia.

20

[0004] Los cánceres se detectan a menudo mediante biopsias de tumores seguidas de análisis de patologías celulares, biomarcadores o ADN extraído de las células. Pero más recientemente se ha propuesto que los cánceres también se pueden detectar a partir de ácidos nucleicos libres de células (por ejemplo, ácido nucleico circulante, ácido nucleico tumoral circulante, exosomas, ácidos nucleicos de células apoptóticas y/o células necróticas) en fluidos corporales, como sangre u orina (ver, por ejemplo, Siravegna et al., Nature Reviews 2017). Estas pruebas tienen la ventaja de que no son invasivas, se pueden realizar sin identificar células cancerosas sospechosas a través de la biopsia y la toma de muestras de ácidos nucleicos de todas las partes de un cáncer. Sin embargo, estas pruebas se complican por el hecho de que la cantidad de ácidos nucleicos liberados a los fluidos corporales es baja y variable, al igual que la recuperación de ácidos nucleicos de dichos fluidos en forma analizable. Estas fuentes de variación pueden oscurecer el valor predictivo de la comparación de la carga de mutación tumoral (TMB) entre muestras.

25

30

[0005] TMB es una medida de las mutaciones que portan las células tumorales en un genoma tumoral. TMB es un biomarcador de tipo que se puede utilizar para evaluar si un sujeto diagnosticado o sospechoso de tener signos de cáncer se beneficiará de una terapia contra el cáncer, como la terapia inmuno-oncológica (I-O).

35

[0006] Zehir, A., Benayed, R., Shah, R. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 23, 703-713 (2017). <https://doi.org/10.1038/nm.4333> revela el uso de una distribución general de TMB para identificar un umbral para considerar tumores altamente mutados adecuados para la identificación de firmas mutacionales.

40

## RESUMEN

[0007] Un aspecto de la divulgación se refiere a un método para proporcionar una medida de la carga de mutación tumoral en una muestra de prueba de ácido nucleico libre de células de un sujeto que tiene un tipo de cáncer o signos de un tipo de cáncer, que comprende: (a) determinar un número de mutaciones presentes en los ácidos nucleicos libres de células de la muestra de prueba, y una fracción de alelos menores basada en una o más mutaciones más altamente representadas en los ácidos nucleicos libres de células de la muestra de prueba; y (b) normalizar el número de mutaciones presentes en la muestra a un número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba.

45

50

[0008] En algunas formas de realización, el número de mutaciones presentes en las muestras de control es un promedio.

55

[0009] En algunas formas de realización, el contenedor tiene un ancho de no más del 20 %, no más del 10 % o no más del 5 %.

[0010] En algunas formas de realización, el método comprende además determinar si el número de mutaciones presentes en la muestra está por encima de un umbral, en donde el umbral se establece para indicar un sujeto que es probable que responda positivamente a una inmunoterapia.

60

[0011] En algunas formas de realización, la normalización comprende dividir el número de mutaciones en la muestra de prueba por un número promedio de mutaciones en las muestras de control.

65

- [0012] En algunas formas de realización, la normalización comprende restar del número determinado de mutaciones en la muestra de prueba de ácido nucleico libre de células un promedio del número de mutaciones en las muestras de control dentro del contenedor.
- 5 [0013] En algunas formas de realización, el método comprende además dividir el número de mutaciones en la muestra de prueba de ácido nucleico libre de células menos el número promedio de mutaciones presentes en las muestras de control por una desviación estándar del número de mutaciones presentes en las muestras de control para calcular una puntuación Z. El promedio puede ser una media.
- 10 [0014] En algunas formas de realización, la normalización comprende determinar el promedio y la dispersión del número de mutaciones en al menos 10, 50, 100 o 500 muestras de control, determinar una puntuación estándar de desviación del promedio en la muestra de prueba y determinar si la puntuación estándar está por encima de un número umbral. El promedio puede ser una media, una mediana o una moda. La dispersión se puede representar como varianza, desviación estándar o rango intercuartil. La puntuación estándar de desviación puede ser una puntuación Z.
- 15 [0015] En algunas formas de realización, la normalización comprende además dividir el número determinado de mutaciones en la muestra de prueba de ácido nucleico libre de células por el número promedio de mutaciones presentes en las muestras de control en el mismo contenedor.
- 20 [0016] En algunas formas de realización, la normalización se implementa en una computadora programada para almacenar valores para el número de mutaciones presentes en una pluralidad de compartimentos de fracciones de alelos menores. Los valores almacenados pueden ser una media y una desviación estándar del número de mutaciones presentes en cada uno de los múltiples compartimentos.
- 25 [0017] En algunas formas de realización, que comprende determinar una puntuación estándar de carga de mutación tumoral en el sujeto y si la puntuación estándar está por encima de un umbral para sujetos de control consistente con la capacidad de respuesta a la inmunoterapia.
- 30 [0018] En algunas formas de realización, (a) comprende determinar secuencias de moléculas de ácido nucleico libres de células en la muestra de prueba y comparar las secuencias resultantes con las secuencias de referencia correspondientes para identificar la cantidad de mutaciones presentes en la muestra y la fracción de alelos menores. Las secuencias de referencia son de hG19 o hG38.
- 35 [0019] En algunas formas de realización, las muestras de control incluyen al menos 25, 50, 100, 200 o 500 muestras de control.
- [0020] En algunas formas de realización, se secuencian al menos 50.000, 100.000 o 150.000 nucleótidos en los segmentos de ácido nucleico.
- 40 [0021] En algunas formas de realización, (a) comprende determinar la presencia o ausencia de un panel de mutaciones predeterminadas que se sabe que ocurren en el cáncer del tipo presente o que se sospecha que está presente en la muestra, opcionalmente en donde las mutaciones son mutaciones somáticas que afectan la secuencia de una proteína codificada.
- 45 [0022] En algunas formas de realización, el paso (a) comprende unir adaptadores a los ácidos nucleicos libres de células, amplificar los ácidos nucleicos libres de células a partir de cebadores que se unen a los adaptadores y secuenciar los ácidos nucleicos amplificados.
- 50 [0023] En algunas formas de realización, la secuenciación es secuenciación de amplificación de puente, pirosecuenciación, secuenciación de semiconductores iónicos, secuenciación de extremos de pares, secuenciación por ligadura o secuenciación de moléculas individuales en tiempo real.
- [0024] En un aspecto, la divulgación se refiere a un método para tratar a un sujeto que comprende: (a) determinar un número de mutaciones presentes en los ácidos nucleicos libres de células de la muestra de prueba, y una fracción de alelos menores basada en una o más mutaciones más altamente representadas en los ácidos nucleicos libres de células de la muestra de prueba; (b) normalizar el número de mutaciones presentes en la muestra al número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba; y (c) administrar inmunoterapia al sujeto si la medida de la carga mutacional del tumor excede un umbral.
- 55 [0025] En algunas formas de realización, el método se realiza en una pluralidad de sujetos para determinar una medida de la carga de mutación tumoral en cada sujeto, en donde una mayor proporción de sujetos con la medida de la carga de mutación del cáncer que excede un umbral reciben inmunoterapia para el cáncer que los sujetos con la medida de la mutación tumoral por debajo del umbral.
- 60
- 65

- [0026] En algunas formas de realización, todos los sujetos en los que la medida está por encima de un primer umbral reciben inmunoterapia y todos los sujetos en los que la medida está por debajo de un segundo umbral no reciben inmunoterapia.
- 5 [0027] En algunas formas de realización, la medida es una puntuación Z.
- [0028] En algunas formas de realización, la inmunoterapia comprende la administración de un anticuerpo inhibidor de puntos de control.
- 10 [0029] En algunas formas de realización, la inmunoterapia comprende la administración de un anticuerpo contra PD-1, PD-2, PD-L1, PD-L2, CTLA-40, OX40, B7.1, B7He, LAG3, CD137, KIR, CCR5, CD27 o CD40.
- [0030] En algunas formas de realización, en donde la inmunoterapia comprende la administración de una citocina proinflamatoria.
- 15 [0031] En algunas formas de realización, la inmunoterapia comprende la administración de células T contra el tipo de cáncer.
- [0032] En algunas formas de realización, el tipo de cáncer es un cáncer sólido.
- 20 [0033] En algunas formas de realización, el tipo de cáncer es cáncer renal, de mesotelioma, de tejidos blandos, del SNC primario, de tiroides, de hígado, de próstata, de páncreas, CUP, neuroendocrino, CPCNP, gastroesofágico, de cabeza y cuello, CPCNP, de mama, melanoma, colangiocarcinoma, ginecológico, colorrectal o urotelial.
- 25 [0034] En algunas formas de realización, el tipo de cáncer es una neoplasia maligna hematopoyética.
- [0035] En algunas formas de realización, el tipo de cáncer es leucemia o linfoma.
- [0036] En un aspecto, la divulgación se refiere a un método para tratar a un sujeto que tiene un cáncer, que comprende administrar un agente de inmunoterapia al sujeto, en donde el sujeto ha sido identificado para inmunoterapia a partir de una medida de la carga de mutación del cáncer del sujeto determinada por: (a) determinar un número de mutaciones presentes en los ácidos nucleicos libres de células de la muestra del sujeto, y una fracción de alelos menores para la mutación más altamente representada en los ácidos nucleicos libres de células de la muestra de prueba; y (b) normalizar el número de mutaciones presentes en la muestra al número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar la medida de la carga de mutación tumoral en la muestra del sujeto; en donde se determina que el sujeto tiene una carga mutacional tumoral por encima de un umbral.
- 30 [0037] La divulgación proporciona además un sistema, que comprende:
- (1) una interfaz de comunicación que recibe, a través de una red de comunicación, lecturas de secuenciación generadas mediante la secuenciación de ácidos nucleicos libres de células en una muestra de prueba; y
- 45 (2) una computadora en comunicación con la interfaz de comunicación, en donde la computadora comprende uno o más procesadores de computadora y un medio legible por computadora que comprende un código ejecutable por máquina que, tras la ejecución por uno o más procesadores de computadora, implementa un método que comprende:
- (a) recibir, a través de la red de comunicación, las lecturas de secuenciación generadas por el secuenciador de ácidos nucleicos;
- 50 (b) determinar un número de mutaciones presentes en las lecturas de secuenciación de la muestra de prueba, y una fracción de alelos menores basada en una o más mutaciones más altamente representadas en las lecturas de secuenciación de la muestra de prueba; y
- 55 (c) normalizar el número de mutaciones presentes en la muestra de prueba con un número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba.
- 60 [0038] En algunas formas de realización, el secuenciador de ácidos nucleicos secuenciar una biblioteca de secuenciación generada a partir de moléculas de ADN libres de células derivadas de un sujeto, en donde la biblioteca de secuenciación comprende las moléculas de ADN libres de células y adaptadores que comprenden códigos de barras. En algunas formas de realización, el secuenciador de ácidos nucleicos realiza la secuenciación por síntesis en la biblioteca de secuenciación para generar las lecturas de secuenciación. En algunas formas de realización, el secuenciador de ácidos nucleicos realiza pirosecuenciación, secuenciación de una sola molécula, secuenciación de nanoporos, secuenciación de semiconductores, secuenciación por ligación o secuenciación por hibridación en la biblioteca de secuenciación para
- 65

generar las lecturas de secuenciación. En algunas formas de realización, el secuenciador de ácidos nucleicos utiliza una matriz de una sola molécula clonal derivada de la biblioteca de secuenciación para generar las lecturas de secuenciación. En algunas formas de realización, el secuenciador de ácidos nucleicos comprende un chip que tiene una matriz de micropocillos para secuenciar la biblioteca de secuenciación para generar las lecturas de secuenciación. En algunas formas de realización, el medio legible por computadora comprende una memoria, un disco duro o un servidor de computadora. En algunas formas de realización, la red de comunicación comprende una red de telecomunicaciones, una internet, una extranet o una intranet. En algunas formas de realización, la red de comunicación incluye uno o más servidores de computadora capaces de computación distribuida. En algunas formas de realización, la computación distribuida es computación en la nube. En algunos métodos, la computadora está ubicada en un servidor de computadora que está ubicado de manera remota desde el secuenciador de ácidos nucleicos. En algunas formas de realización, la biblioteca de secuenciación comprende además códigos de barras de muestra que diferencian una muestra de una o más muestras. En algunas formas de realización, el sistema comprende además una pantalla electrónica en comunicación con la computadora a través de una red, en donde la pantalla electrónica comprende una interfaz de usuario para mostrar resultados al implementar (a)-(c). En algunas formas de realización, la interfaz de usuario es una interfaz gráfica de usuario (IGU) o una interfaz de usuario basada en web. En algunas formas de realización, la pantalla electrónica está en una computadora personal. En algunas formas de realización, la pantalla electrónica está en una computadora habilitada para Internet. En algunas formas de realización, la computadora habilitada para Internet está ubicada en una ubicación remota desde la computadora.

**[0039]** En algunas formas de realización, los resultados de los sistemas y métodos divulgados en el presente documento se utilizan como entrada para generar un informe en formato papel. Por ejemplo, este informe puede proporcionar una indicación de las variantes llamadas y/o las variantes que se consideran errores de desaminación.

**[0040]** Los diversos pasos de los métodos divulgados en este documento, o los pasos llevados a cabo por los sistemas divulgados en este documento, pueden llevarse a cabo en el mismo momento o en momentos diferentes, en la misma ubicación geográfica o en lugares diferentes, por ejemplo, países, y/o por la misma o diferentes personas.

#### BREVE DESCRIPCIÓN DE LAS FIGURAS

**[0041]** La figura 1 muestra la distribución de la carga de mutación por muestra según el tipo de tumor. Recuento de SNV e indel por muestra en relación con otras muestras dentro de la indicación clínica. Las barras horizontales indican el percentil 95.  
La figura 2 muestra un sistema informático para implementar los métodos descritos.

#### DEFINICIONES

**[0042]** Un sujeto se refiere a un animal, como una especie de mamífero (preferiblemente humano) o una especie de ave (p. ej., pájaro), u otro organismo, como una planta. Más específicamente, un sujeto puede ser un vertebrado, p. ej., un mamífero como un ratón, un primate, un simio o un humano. Los animales incluyen animales de granja, animales de deporte y mascotas. Un sujeto puede ser un individuo sano, un individuo que tiene o se sospecha que tiene una enfermedad o una predisposición a la enfermedad, o un individuo que necesita terapia o se sospecha que necesita terapia.

**[0043]** Por ejemplo, un sujeto es un individuo al que se le ha diagnosticado un cáncer, va a recibir una terapia contra el cáncer y/o ha recibido al menos una terapia contra el cáncer. El sujeto puede estar en remisión de un cáncer. Como otro ejemplo, el sujeto es un individuo al que se le ha diagnosticado una enfermedad autoinmune. Como otro ejemplo, el sujeto puede ser un individuo que está embarazada o que está planeando quedar embarazada, al que se le puede haber diagnosticado o se puede sospechar que tiene una enfermedad, por ejemplo, un cáncer o una enfermedad autoinmune.

**[0044]** Un marcador de cáncer es una variante genética asociada con la presencia o el riesgo de desarrollar un cáncer. Un marcador de cáncer puede proporcionar una indicación de que un sujeto tiene cáncer o un mayor riesgo de desarrollar cáncer que un sujeto de la misma especie, de la misma edad y género, que no tiene el marcador de cáncer. Un marcador de cáncer puede o no ser causa de cáncer.

**[0045]** Los códigos de barras se pueden unir a un extremo o a ambos extremos de los ácidos nucleicos. Los códigos de barras se pueden decodificar para revelar información como la muestra de origen, la forma o el procesamiento de un ácido nucleico. Los códigos de barras se pueden utilizar para permitir la agrupación y el procesamiento paralelo de múltiples muestras que comprenden ácidos nucleicos que tienen diferentes códigos de barras y posteriormente los ácidos nucleicos se deconvolucionan mediante la lectura de códigos de barras. Los códigos de barras también se pueden denominar identificadores moleculares, identificadores de muestra, etiquetas o etiquetas de índice. Los códigos de barras se pueden utilizar para distinguir muestras (identificadores de muestra). Adicionalmente o alternativamente, los códigos de barras se pueden utilizar para distinguir diferentes moléculas en la misma muestra. Esto incluye tanto la codificación de barras única de cada molécula diferente en la muestra, como el uso de una codificación de barras no única de cada molécula. En el caso de la codificación de barras no única, se puede utilizar un número limitado de códigos de barras para codificar cada molécula de manera que se puedan distinguir diferentes moléculas en función de su posición de inicio/fin donde se asignan a un genoma de referencia en combinación con al menos una etiqueta. Por lo general, se utiliza una cantidad suficiente

de códigos de barras diferentes de modo que exista una probabilidad baja (por ejemplo, <10 %, <5 %, <1 % o <0,1 %) de que dos moléculas cualesquiera que tengan el mismo inicio/fin también tengan el mismo código de barras. Algunos códigos de barras incluyen múltiples identificadores moleculares para etiquetar muestras, formas de moléculas dentro de una muestra y moléculas dentro de una forma que tienen los mismos puntos de inicio y fin. Dichos códigos de barras pueden existir en la forma A1i, en la que la letra indica un tipo de muestra, el número arábigo indica una forma de molécula dentro de una muestra y el número romano indica una molécula dentro de una forma.

**[0046]** Los adaptadores son ácidos nucleicos cortos (por ejemplo, de menos de 500, 100 o 50 nucleótidos de longitud) generalmente al menos parcialmente bicatenarios para la unión a uno o ambos extremos de una molécula de ácido nucleico de muestra. Los adaptadores pueden incluir sitios de unión de cebadores para permitir la amplificación de una molécula de ácido nucleico flanqueada por adaptadores en ambos extremos, y/o un sitio de unión de cebador de secuenciación, incluidos sitios de unión de cebadores para la secuenciación de próxima generación (NGS). Los adaptadores también pueden incluir sitios de unión para sondas de captura, como un oligonucleótido unido a un soporte de celda de flujo. Los adaptadores también pueden incluir un código de barras como se describió anteriormente. Los códigos de barras se colocan preferiblemente en relación con los sitios de unión del cebador y del cebador de secuenciación, de modo que se incluye un código de barras en los amplicones y las lecturas de secuenciación de una molécula de ácido nucleico. Los mismos adaptadores o diferentes se pueden unir a los extremos respectivos de una molécula de ácido nucleico. A veces, el mismo adaptador se une a los extremos respectivos, excepto que el código de barras es diferente. Un adaptador preferido es un adaptador en forma de Y en el que un extremo tiene un extremo romo o una cola, como se describe en este documento, para unirse a una molécula de ácido nucleico, que también tiene un extremo romo o una cola con uno o más nucleótidos complementarios. Otro adaptador preferido es un adaptador en forma de campana, igualmente con un extremo romo o una cola para unirse a un ácido nucleico que se va a analizar.

**[0047]** Como se utiliza en este documento, el término "secuenciación" se refiere a cualquiera de una serie de tecnologías utilizadas para determinar la secuencia de una biomolécula, por ejemplo, un ácido nucleico como ADN o ARN. Los métodos de secuenciación ejemplares incluyen, entre otros, secuenciación dirigida, secuenciación en tiempo real de una sola molécula, secuenciación de exones, secuenciación basada en microscopía electrónica, secuenciación en panel, secuenciación mediada por transistores, secuenciación directa, secuenciación aleatoria de escopeta, secuenciación de terminación dideoxi de Sanger, secuenciación de todo el genoma, secuenciación por hibridación, pirosecuenciación, electroforesis capilar, electroforesis en gel, secuenciación dúplex, secuenciación cíclica, secuenciación de extensión de una sola base, secuenciación en fase sólida, secuenciación de alto rendimiento, secuenciación de firma masiva en paralelo, PCR en emulsión, coamplificación a una temperatura de desnaturalización más baja-PCR (COLD-PCR), PCR multiplex, secuenciación por terminador de tinte reversible, secuenciación de extremos emparejados, secuenciación a corto plazo, secuenciación de exonucleasa, secuenciación por ligadura, secuenciación de lectura corta, secuenciación de una sola molécula, secuenciación por síntesis, secuenciación en tiempo real, secuenciación con terminador inverso, secuenciación con nanoporos, secuenciación 454, secuenciación con Solexa Genome Analyzer, secuenciación SOLiD™, secuenciación MS-PET y una combinación de las mismas. En algunas formas de realización, la secuenciación se puede realizar mediante un analizador de genes como, por ejemplo, los analizadores de genes disponibles comercialmente de Illumina o Applied Biosystems.

**[0048]** La frase "secuenciación de próxima generación" o NGS se refiere a tecnologías de secuenciación que tienen un mayor rendimiento en comparación con los enfoques tradicionales basados en electroforesis capilar y de Sanger, por ejemplo, con la capacidad de generar cientos de miles de lecturas de secuencias relativamente pequeñas a la vez. Algunos ejemplos de técnicas de secuenciación de próxima generación incluyen, entre otros, la secuenciación por síntesis, la secuenciación por ligación y la secuenciación por hibridación.

**[0049]** La frase "ejecución de secuenciación" se refiere a cualquier paso o porción de un experimento de secuenciación realizado para determinar alguna información relacionada con al menos una biomolécula (por ejemplo, una molécula de ácido nucleico como ADN o ARN).

**[0050]** El ADN (ácido desoxirribonucleico) es una cadena de nucleótidos que comprende cuatro tipos de nucleótidos: adenina (A), timina (T), citosina (C) y guanina (G). El ARN (ácido ribonucleico) es una cadena de nucleótidos que comprende cuatro tipos de nucleótidos: A, uracilo (U), G y C. Ciertos pares de nucleótidos se unen específicamente entre sí de manera complementaria (llamado apareamiento de bases complementarias). En el ADN, la adenina (A) se empareja con la timina (T) y la citosina (C) se empareja con la guanina (G). En el ARN, la adenina (A) se empareja con el uracilo (U) y la citosina (C) se empareja con la guanina (G). Cuando una primera cadena de ácido nucleico se une a una segunda cadena de ácido nucleico formada por nucleótidos que son complementarios a los de la primera cadena, las dos cadenas se unen para formar una doble cadena. Tal como se utiliza en el presente documento, "datos de secuenciación de ácidos nucleicos", "información de secuenciación de ácidos nucleicos", "secuencia de ácidos nucleicos", "secuencia de nucleótidos", "secuencia genómica", "secuencia genética" o "secuencia de fragmentos" o "lectura de secuenciación de ácidos nucleicos" denota cualquier información o dato que sea indicativo del orden de las bases de nucleótidos (por ejemplo, adenina, guanina, citosina y timina o uracilo) en una molécula (por ejemplo, un genoma completo, transcriptoma completo, exoma, oligonucleótido, polinucleótido o fragmento) de un ácido nucleico como ADN o ARN. Debe entenderse que las presentes enseñanzas contemplan información de secuencia obtenida utilizando todas las variedades disponibles de técnicas, plataformas o tecnologías, incluyendo, pero sin limitarse a: electroforesis capilar, microarreglos, sistemas basados en ligación, sistemas basados en polimerasa, sistemas basados en hibridación, sistemas de identificación de

nucleótidos directos o indirectos, pirosecuenciación, sistemas de detección basados en iones o pH y sistemas basados en firma electrónica.

**[0051]** Un "polinucleótido", "ácido nucleico", "molécula de ácido nucleico" u "oligonucleótido" se refiere a un polímero lineal de nucleósidos (incluidos desoxirribonucleósidos, ribonucleósidos o análogos de los mismos) unidos por enlaces internucleosídicos. Normalmente, un polinucleótido comprende al menos tres nucleósidos. Los oligonucleótidos a menudo varían en tamaño desde unas pocas unidades monoméricas, por ejemplo, 3-4, hasta cientos de unidades monoméricas. Siempre que un polinucleótido esté representado por una secuencia de letras, como "ATGCCTG", se entenderá que los nucleótidos están en orden 5' -> 3' de izquierda a derecha y que "A" denota desoxiadenosina, "C" denota desoxicitidina, "G" denota desoxiguanosina y "T" denota timidina, a menos que se indique lo contrario. Las letras A, C, G y T pueden usarse para referirse a las bases mismas, a los nucleósidos o a los nucleótidos que comprenden las bases, como es estándar en la técnica.

**[0052]** Una secuencia de referencia es una secuencia conocida que se utiliza con fines de comparación con secuencias determinadas experimentalmente. Por ejemplo, una secuencia conocida puede ser un genoma completo, un cromosoma o cualquier segmento del mismo. Una referencia incluye típicamente al menos 20, 50, 100, 200, 250, 300, 350, 400, 450, 500, 1000 o más nucleótidos. Una secuencia de referencia puede alinearse con una única secuencia contigua de un genoma o cromosoma o puede incluir segmentos no contiguos que se alinean con diferentes regiones de un genoma o cromosoma. Los genomas humanos de referencia incluyen, por ejemplo, hG19 y hG38.

**[0053]** Una primera secuencia de ácido nucleico monocatenario se superpone con una segunda secuencia monocatenaria si la primera secuencia de ácido nucleico o su complemento y la segunda secuencia de ácido nucleico o su complemento se alinean con segmentos superpuestos, pero no idénticos de una secuencia de referencia contigua, como la secuencia de un cromosoma humano. Un ácido nucleico total o parcialmente bicatenario se superpone con otro ácido nucleico total o parcialmente bicatenario si cualquiera de sus cadenas se superpone a las del otro ácido nucleico.

**[0054]** "Promedio" se refiere a cualquier medida estadística de tendencia central, incluyendo, sin limitación, media, mediana y moda.

**[0055]** "Extensión" se refiere a cualquier medida estadística de dispersión, incluyendo, sin limitación, varianza, desviación estándar y rango intercuartil.

**[0056]** "Puntuación estándar" se refiere a cualquier medida estadística de distancia respecto de un promedio, incluyendo, sin limitación, una puntuación normalizada o una puntuación Z (número de desviación estándar respecto del promedio).

**[0057]** La carga de mutación tumoral normalizada se refiere a una puntuación estándar de la carga tumoral en comparación con los sujetos de control. Incluye la medición de la carga de mutación tumoral en una muestra de molécula de ácido nucleico de prueba ajustada para tener en cuenta las variaciones aleatorias entre muestras en los factores que afectan la detección de dichas mutaciones, como la liberación de ácidos nucleicos de las células cancerosas a un fluido corporal y la recuperación de ácidos nucleicos del fluido corporal en forma analizable.

**[0058]** Una mutación se refiere a una variación de una secuencia de referencia conocida e incluye mutaciones como, por ejemplo, SNV, variaciones/aberraciones del número de copias, indeles y fusiones de genes. Una mutación puede ser una mutación de línea germinal o somática. Una secuencia de referencia preferida para fines de comparación es una secuencia genómica de tipo salvaje de la especie del sujeto que proporciona una muestra de prueba, típicamente el genoma humano.

**[0059]** Una variante puede denominarse alelo. Una variante suele presentarse con una frecuencia del 50 % (0,5) o del 100 % (1), dependiendo de si el alelo es heterocigoto u homocigoto. Por ejemplo, las variantes de la línea germinal se heredan y suelen tener una frecuencia de 0,5 o 1. Las variantes somáticas, en cambio, son variantes adquiridas y suelen tener una frecuencia de < 0,5.

**[0060]** Los alelos mayores y menores de un locus genético se refieren a los ácidos nucleicos que albergan el locus en el que el locus está ocupado por un nucleótido de una secuencia de referencia y un nucleótido variante diferente de la secuencia de referencia, respectivamente. Las mediciones en un locus pueden tomar la forma de fracciones alélicas (FA), que miden la frecuencia con la que se observa un alelo en una muestra.

**[0061]** Los términos "frecuencia de alelos menores" pueden referirse a la frecuencia con la que aparecen alelos menores (por ejemplo, no el alelo más común) en una población determinada de ácidos nucleicos, como una muestra. Las variantes genéticas con una frecuencia de alelos menores baja pueden tener una frecuencia de presencia relativamente baja en una muestra.

**[0062]** Una "fracción de alelo menor (FAM)" se refiere a la fracción de moléculas de ADN que albergan una alteración alélica (por ejemplo, una mutación) en una posición genómica dada en una muestra dada. Una FAM de una variante somática puede ser menor que 0,5, 0,1, 0,05 o 0,01 de todas las variantes somáticas o alelos presentes en un locus dado.

Por ejemplo, una FAM de una variante somática es  $< 0,05$ . La fracción de alelo menor también se puede utilizar indistintamente con "fracción de alelo mutante".

**[0063]** Los términos "neoplasia" y "tumor" se utilizan indistintamente. Hacen referencia al crecimiento anormal de células en un sujeto. Una neoplasia o tumor puede ser benigno, potencialmente maligno o maligno. Un tumor maligno se denomina cáncer o tumor canceroso.

**[0064]** Los términos "carga de mutación tumoral (TMB)", "carga mutacional tumoral (TMB)" o "carga mutacional del cáncer" se utilizan indistintamente. Se refieren al número total de mutaciones, por ejemplo, mutaciones somáticas, presentes en una porción secuenciada de un genoma tumoral. TMB puede referirse al número de mutaciones codificantes, de sustitución de bases y de indel por megabase de un genoma tumoral que se está examinando. Pueden ser indicativos para detectar, evaluar, calcular o predecir la sensibilidad y/o resistencia a un agente terapéutico o fármaco contra el cáncer, por ejemplo, inhibidores de puntos de control inmunitarios, anticuerpos. Los tumores que tienen niveles más altos de TMB pueden expresar más neoantígenos, un tipo de antígeno específico del cáncer, lo que puede permitir una respuesta inmunitaria más robusta y, por lo tanto, una respuesta más duradera a la inmunoterapia. El sistema inmunitario depende de una cantidad suficiente de neoantígenos para responder adecuadamente; la cantidad de mutaciones somáticas puede actuar como un indicador para determinar la cantidad de neoantígenos en un tumor. La TMB se puede utilizar para deducir la solidez de una respuesta inmunitaria a un tratamiento farmacológico y la eficacia de un tratamiento farmacológico en un sujeto. Las variantes de la línea germinal y somáticas se pueden distinguir bioinformáticamente para identificar variantes somáticas antigénicas, como se describe en PCT/US2018/52087.

**[0065]** Un umbral es un valor predeterminado que se utiliza para caracterizar valores determinados experimentalmente del mismo parámetro para diferentes muestras dependiendo de su relación con el umbral.

**[0066]** Los términos "procesamiento", "cálculo" y "comparación" se pueden utilizar indistintamente. El término puede referirse a determinar una diferencia, por ejemplo, una diferencia en número o secuencia. Por ejemplo, se pueden procesar valores o secuencias de expresión génica, variación del número de copias (CNV), indel y/o variante de nucleótido único (SNV).

**[0067]** "Tipo de cáncer" se refiere al tipo o subtipo definido, por ejemplo, por histopatología. El tipo de cáncer se puede definir por cualquier criterio convencional, como cáncer del mismo tejido (por ejemplo, cánceres de sangre, SNC, cánceres cerebrales, cánceres de pulmón (de células pequeñas y no pequeñas), cánceres de piel, cánceres de nariz, cánceres de garganta, cánceres de hígado, cánceres de huesos, linfomas, cánceres de páncreas, cánceres de intestino, cánceres de recto, cánceres de tiroides, cánceres de vejiga, cánceres de riñón, cánceres de boca, cánceres de estómago, cánceres de mama, cánceres de próstata, cánceres de ovario, cánceres de pulmón, cánceres de intestino, cánceres de tejidos blandos, cánceres de tiroides, cánceres neuroendocrinos, cánceres gastroesofágicos, cánceres de cabeza y cuello, cánceres ginecológicos, cánceres colorrectales, cánceres uroteliales, cánceres de estado sólido, cánceres heterogéneos, cánceres homogéneos), origen primario desconocido y similares, y/o del mismo linaje celular (por ejemplo, carcinoma, sarcoma, linfoma, colangiocarcinoma, leucemia, mesotelioma, melanoma, o glioblastoma) y/o marcadores de cáncer, como Her2, CA15-3, CA19-9, CA-125, CEA, AFP, PSA, HCG, receptor hormonal y NMP-22. Los cánceres también se pueden clasificar por estadio (p. ej., estadio 1, 2, 3 o 4) y si son de origen primario o secundario.

## DESCRIPCIÓN DETALLADA

### I. General

**[0068]** La divulgación se basa en parte en el resultado de que los valores de la carga de mutación tumoral de diferentes muestras pueden hacerse más comparables entre sí o con los estándares de control mediante un régimen de normalización que tenga en cuenta la fracción de alelos menores de mutaciones de alta calificación en una muestra. Dicho análisis puede proporcionar una indicación de dónde se encuentra la carga de mutación tumoral de una muestra de prueba en una distribución de cargas de mutación tumoral en una población de control y, por lo tanto, si es probable que el individuo que proporciona la muestra de prueba sea susceptible a la inmunoterapia para tratar un cáncer.

### II. Determinación de la carga de mutación tumoral y normalización

**[0069]** Los ácidos nucleicos presentes en una muestra pueden procesarse y secuenciarse como se describe más adelante. La secuenciación revela un número total de mutaciones presentes y detectadas en una muestra, es decir, un número total de loci en los que se detecta un alelo menor preferiblemente con una frecuencia suficiente en diferentes moléculas de ácido nucleico en la muestra como para que sea estadísticamente improbable que represente un artefacto de secuenciación (por ejemplo,  $p \leq 0,05$ ). El número total de mutaciones determinadas puede representar mutaciones presentes en cualquier parte del genoma del individuo que proporciona la muestra, o cualquier fracción del mismo, como un cromosoma particular, o un conjunto de segmentos genómicos no contiguos, como un conjunto de dichos segmentos que se sabe que albergan loci en los que ocurren mutaciones asociadas con el cáncer. Las mutaciones determinadas pueden ser exclusivamente mutaciones que causan cambios en la secuencia de una proteína codificada, por ejemplo, SNV, indel, fusión, o pueden incluir cualquier tipo de mutaciones que no causen cambios en la secuencia de una proteína codificada, por ejemplo, variaciones en el número de copias, aberraciones en el número de copias. Si se determina algún



tipo de mutación, se pueden seleccionar las mutaciones que modifican las secuencias de aminoácidos de las proteínas codificadas antes del procesamiento posterior. Las mutaciones pueden incluir mutaciones de la línea germinal, mutaciones somáticas o ambas.

**[0070]** Aunque las mutaciones que codifican cambios de aminoácidos en las proteínas codificadas probablemente estén mejor correlacionadas con la susceptibilidad a la inmunoterapia que otras mutaciones, el muestreo de todas las mutaciones puede correlacionarse mejor con el número de mutaciones que cambian la secuencia en las secuencias de proteínas codificadas, que el recuento directo de dichas mutaciones debido a la pérdida de algunas de esas mutaciones por debajo de un nivel de detección. Por lo tanto, ambos enfoques tienen ventajas y pueden utilizarse.

**[0071]** La secuenciación también proporciona la fracción de alelos menores de cualquiera o todas las mutaciones detectadas (o subconjunto de las mismas seleccionado para el procesamiento posterior). La fracción de alelos menores significa la proporción de todos los ácidos nucleicos secuenciados en una muestra que incluye un locus de mutación que alberga un alelo menor (a diferencia de un alelo de tipo salvaje). Por lo tanto, la fracción de alelos menores se puede representar mediante un número entre 0 y 1. Si puede haber más de un alelo menor en un locus, la fracción de alelos menores se puede definir como la fracción de cualquiera de los alelos menores o la fracción agregada de todos o cualquier subconjunto de los alelos menores.

**[0072]** La fracción de alelos menores del mutante con mayor representación o una fracción de alelos menores promedio de un conjunto de mutantes con alta representación se utiliza en la normalización posterior. Si se utiliza un conjunto de mutantes con alta representación, el conjunto puede representar, por ejemplo, los 2, 3, 5 o 10 mutantes con mayor representación.

**[0073]** El análisis descrito para una muestra de prueba también se puede llevar a cabo en una población de muestras de control para proporcionar un conjunto de datos para comparación. La población de control puede incluir muestras de, por ejemplo, al menos 10, 20, 25, 50, 100, 200, 250, 500, 1.000, 5.000, 10.000, 50.000 o más individuos. Las muestras de control pueden ser muestras de sujetos con el mismo tipo de cáncer que una muestra de prueba. Cada muestra de control se analiza asimismo para la carga total de mutación tumoral y la fracción de alelo menor de la mutación o conjunto de mutaciones con mayor representación. Preferiblemente, la fracción de alelo menor se determina de la misma manera entre las muestras de prueba y las muestras de control (es decir, en función de la mutación con mayor representación o el mismo conjunto de mutaciones con mayor representación). Tal es también el caso para el recuento de mutaciones. Por ejemplo, si las mutaciones que ocurren en cualquier parte del genoma se cuentan en la muestra de prueba, lo mismo es preferiblemente el caso en las muestras de control. Del mismo modo, si para una muestra de prueba solo se cuentan las mutaciones que afectan a una secuencia de proteína codificada, lo mismo ocurre con las muestras de control.

**[0074]** Las muestras de control pueden entonces clasificarse en contenedores según las fracciones de alelos menores determinadas. Los contenedores pueden ser de igual tamaño (por ejemplo, 0,05-0,1, 0,1-0,15, 0,15-0,2, 0,2-0,25) o los contenedores pueden variar en tamaño, por ejemplo, para uniformizar el número de muestras de control que caben en cada contenedor. Los contenedores también pueden definirse como un porcentaje de la variación total en la fracción de alelos menores. Por ejemplo, si la fracción de alelos menores de la mutación o mutaciones más representadas en la población de control varía entre 0,1 y 0,5, los contenedores pueden definirse por un porcentaje de ese rango (por ejemplo, 5 %, 10 % o 20 % por contenedor). A continuación, se determina una carga de mutación tumoral promedio para las muestras de control en cada contenedor. Por ejemplo, si hay tres muestras de control con cargas de mutación de 3, 4 y 5 en un intervalo de 0,1 a 0,15, entonces la carga de mutación de cáncer promedio para ese intervalo es 4. También se puede calcular una desviación estándar para los valores de mutación de cáncer dentro de un intervalo. Una colección de intervalos de este tipo puede completar datos para compararlos con muestras de prueba del mismo tipo de cáncer. El intervalo puede tener un ancho de no más del 20 %, no más del 10 % o no más del 5 %.

**[0075]** La población de control solo necesita analizarse una vez y los datos resultantes pueden servir para comparar con cualquier número de muestras de prueba. Sin embargo, la población de control también puede complementarse con datos de individuos adicionales con el mismo tipo de cáncer.

**[0076]** El mismo tipo de análisis también se puede realizar en poblaciones de control adicionales con otros tipos de cáncer para compararlas con muestras de prueba con estas otras formas de cáncer.

**[0077]** El número de mutaciones tumorales medidas en una muestra de prueba se puede comparar entonces con el número promedio de mutaciones de un grupo de muestras de control definido por un rango de frecuencias de alelos menores que incluye el de la muestra de prueba. Por ejemplo, si la muestra de prueba tiene una frecuencia de alelo menor del alelo menor más representado de 0,125, entonces se puede elegir un grupo que incluya frecuencias de alelos menores de 0,1 a 0,15 para la comparación. Una comparación numérica simple (por ejemplo, restando la carga de mutación tumoral promedio de las muestras de control de la de la muestra de prueba, o dividiendo la carga de mutación tumoral de la muestra de prueba por el valor promedio de las muestras de control) indica si la carga de mutación tumoral de una muestra está en, por encima o por debajo del promedio. Por ejemplo, si la carga de mutación tumoral de una muestra de prueba es 5 y el promedio del grupo de representación coincidente con el alelo menor es 3, entonces la carga de mutación tumoral de la muestra de prueba se puede representar como 2 mutaciones más que el promedio o  $5/3=167$  % del promedio.

**[0078]** Sin embargo, se puede realizar una comparación más cuantitativa calculando una puntuación Z. La puntuación Z se calcula restando de la carga de mutación tumoral de la muestra de prueba la carga de mutación tumoral promedio del grupo coincidente y dividiendo el resultado por la desviación estándar de la variación en la mutación tumoral dentro del grupo. La puntuación Z puede ser positiva (carga de mutación más alta que el promedio), negativa (carga de mutación de cáncer más baja que el promedio) o cero (carga de mutación promedio). La magnitud de la puntuación Z (positiva o negativa) es una indicación del grado en el que una muestra de prueba está por encima o por debajo del promedio en carga de mutación tumoral.

**[0079]** Una carga de mutación tumoral normalizada de una muestra de prueba de un sujeto (como se representa, por ejemplo, mediante una puntuación Z) de un sujeto proporciona una indicación de la susceptibilidad del sujeto a la inmunoterapia. En general, cuanto mayor sea la carga de mutación tumoral normalizada (como se puede representar mediante una puntuación Z positiva más alta), más susceptible es el sujeto a la inmunoterapia. Sin estar limitado a ninguna teoría, cuantas más mutaciones haya, más neoepítomos que forman dianas no propias para la inmunoterapia. Por el contrario, una mutación normalizada más baja representada, por ejemplo, mediante una puntuación Z negativa, menos susceptible es un sujeto a la inmunoterapia.

**[0080]** Se pueden establecer uno o más umbrales de carga de mutación tumoral normalizada para determinar o al menos proporcionar una indicación (que se puede utilizar en combinación con otros factores) para determinar si un sujeto recibe o continúa recibiendo, o deja de recibir, una inmunoterapia. Por ejemplo, se puede establecer un umbral de modo que los sujetos que se encuentran en el umbral o por encima de él reciban o continúen recibiendo inmunoterapia, y los sujetos que se encuentran por debajo del umbral no reciban o dejen de recibir inmunoterapia. Alternativamente, se pueden establecer dos umbrales con sujetos que se encuentran en el umbral superior o por encima de él recibiendo o continuando a recibir inmunoterapia y sujetos que se encuentran en el umbral inferior o por debajo de él no recibiendo o dejando de recibir inmunoterapia. Los sujetos entre los umbrales se pueden evaluar mediante factores adicionales en cuanto a si deben recibir o continuar recibiendo inmunoterapia.

**[0081]** Los umbrales se pueden determinar empíricamente observando las respuestas a la inmunoterapia en sujetos caracterizados por una carga de mutación tumoral normalizada para determinar los umbrales que mejor se correlacionan con una respuesta beneficiosa o la falta de ella a la inmunoterapia. Los umbrales se pueden establecer alternativamente en puntos predefinidos en una escala, como, por ejemplo, sujetos con una puntuación Z positiva que reciben o continúan recibiendo terapia, y/o sujetos con una puntuación Z negativa que no reciben o interrumpen la recepción de inmunoterapia. Como otro ejemplo, los sujetos con una puntuación Z superior a 1, 2 o 3 pueden recibir o continuar recibiendo terapia. Como otro ejemplo, los sujetos con una puntuación Z superior a menos de 1 pueden interrumpir la recepción de la terapia. Como otro ejemplo, los sujetos con una puntuación Z positiva, lo que coloca al sujeto con una puntuación Z que representa, por ejemplo, al menos el 75 %, 50 %, 25 %, 15 %, 10 % o 5 % más altos de las puntuaciones Z de los sujetos con ese tipo de cáncer, pueden recibir o continuar recibiendo inmunoterapia con otros sujetos que no reciben inmunoterapia.

**[0082]** Como se mencionó, la carga de mutación tumoral normalizada se puede utilizar con o sin otros factores para determinar si se administra o se continúa administrando inmunoterapia. Dichos otros factores pueden incluir la condición del sujeto, la respuesta del sujeto a otras terapias previamente probadas y la disponibilidad de otras terapias aún no probadas en el sujeto, entre otros factores. Por lo tanto, no necesariamente todos los sujetos por encima de un umbral reciben o continúan recibiendo inmunoterapia o por debajo de un umbral no la reciben, pero en general una mayor proporción de sujetos con una carga de mutación tumoral normalizada por encima de un umbral reciben o continúan recibiendo inmunoterapia que en el caso de sujetos con una carga de mutación tumoral normalizada por debajo de un umbral.

### III. Inmunoterapia

**[0083]** La inmunoterapia se refiere al tratamiento con uno o más agentes que actúan para estimular el sistema inmunológico de modo de matar o al menos inhibir el crecimiento de células de un cáncer, y preferiblemente para reducir un mayor crecimiento del cáncer, reducir el tamaño del cáncer y/o eliminar el cáncer. Algunos de estos agentes se unen a un objetivo presente en las células del cáncer; algunos se unen a un objetivo presente en las células inmunes y no en el cáncer; algunos se unen a un objetivo presente tanto en las células del cáncer como en las células inmunes. Dichos agentes incluyen, pero no se limitan a, inhibidores de puntos de control y/o anticuerpos. Los inhibidores de puntos de control son inhibidores de vías del sistema inmunológico que mantienen la autotolerancia y modulan la duración y amplitud de las respuestas inmunes fisiológicas en los tejidos periféricos para minimizar el daño tisular colateral (véase, por ejemplo, Pardoll, Nature Reviews Cancer 12, 252-264 (2012)). Los agentes ejemplares incluyen anticuerpos contra cualquiera de los siguientes: PD-1, PD-2, PD-L1, PD-L2, CTLA-40, OX40, B7.1, B7He, LAG3, CD137, KIR, CCR5, CD27 o CD40. Otros agentes ejemplares incluyen citocinas proinflamatorias, como IL-1, IL-6 y TNF- $\alpha$ . Otros agentes ejemplares son células T activadas contra un tumor, como, por ejemplo, mediante la expresión de un antígeno quimérico que se dirige a un antígeno tumoral de la célula T. En algunas formas de realización, la inmunoterapia estimula el sistema inmunitario para atacar a los antígenos tumorales que se distinguen de sus homólogos de tipo salvaje por la presencia de una o más mutaciones.

### IV. Otras aplicaciones

[85] Las cargas de mutación tumoral normalizadas determinadas por los métodos actuales se pueden utilizar para diagnosticar la presencia de afecciones, en particular cáncer, en un sujeto, para caracterizar afecciones (por ejemplo, estadificación del cáncer o determinación de la heterogeneidad de un cáncer), controlar la respuesta al tratamiento de una afección, determinar el riesgo de pronóstico de desarrollar una afección o el curso posterior de una afección. Las cargas de mutación tumoral normalizadas también se pueden utilizar para caracterizar una forma específica de cáncer. Los cánceres suelen ser heterogéneos tanto en composición como en estadificación. Los datos del perfil genético pueden permitir la caracterización de subtipos específicos de cáncer que pueden ser importantes en el diagnóstico o tratamiento de ese subtipo específico. Esta información también puede proporcionar a un sujeto o médico pistas sobre el pronóstico de un tipo específico de cáncer y permitir que un sujeto o médico adapte las opciones de tratamiento de acuerdo con el progreso de la enfermedad. Algunos cánceres progresan, volviéndose más agresivos y genéticamente inestables. Otros cánceres pueden permanecer benignos, inactivos o latentes. La carga de mutación tumoral normalizada puede ser útil para determinar la progresión de la enfermedad.

[86] Las cargas de mutación tumoral normalizadas también se pueden utilizar para seleccionar tratamientos más allá de la inmunoterapia y para determinar la eficacia de una opción de tratamiento en particular. Las opciones de tratamiento exitosas pueden aumentar inicialmente la carga de mutación tumoral normalizada si el tratamiento es exitoso, ya que más cánceres pueden morir y desprender ácidos nucleicos, seguido de una disminución a medida que el cáncer se encoge o muere. El tratamiento exitoso también puede disminuir la carga de mutación tumoral y/o la fracción de alelos menores sin un aumento inicial. Además, si se observa que un cáncer está en remisión después del tratamiento, la carga de mutación tumoral normalizada se puede utilizar para monitorear la enfermedad residual o la recurrencia de la enfermedad, como lo indica un recuento de mutaciones normalizadas en un fluido corporal.

## V. Implementación informática

[87] Los métodos presentes pueden implementarse por computadora, de modo que cualquiera o todos los pasos descritos en la especificación o en las reivindicaciones adjuntas, excepto los pasos de química húmeda se puedan realizar en una computadora programada adecuada. La computadora puede ser una computadora central, una computadora personal, una tableta, un teléfono inteligente, una nube, un almacenamiento de datos en línea, un almacenamiento de datos remoto o similar. La computadora puede operarse en una o más ubicaciones.

[88] Un programa informático para analizar una población de ácidos nucleicos puede incluir códigos para realizar cualquiera de los pasos distintos de los pasos de química húmeda descritos en la especificación o en las reivindicaciones adjuntas; por ejemplo, código para recibir datos de secuenciación sin procesar, código para determinar secuencias de ácidos nucleicos a partir de dichos datos, códigos para determinar un número de mutaciones presentes en las secuencias determinadas, código para categorizar mutaciones como que afectan a la secuencia de una proteína codificada o de otro modo, código para determinar la fracción de alelos menores de cualquiera de las mutaciones y código para comparar el número de mutaciones presentes en una muestra con el número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba y código para generar una carga de mutación normalizada opcionalmente con un tratamiento de inmunoterapia asociado.

[89] Los presentes métodos se pueden implementar en un sistema (por ejemplo, un sistema de procesamiento de datos) para analizar una población de ácidos nucleicos. El sistema también puede incluir un procesador, un bus de sistema, una memoria principal y opcionalmente una memoria auxiliar acopladas entre sí para realizar uno o más de los pasos descritos en la especificación o en las reivindicaciones adjuntas, como los siguientes: recibir datos de secuenciación en bruto, determinar secuencias de ácidos nucleicos a partir de dichos datos, identificar mutaciones dentro de las secuencias determinadas, categorizar las mutaciones como que afectan a una secuencia de proteína codificada o de otro modo, determinar la fracción de alelos menores para cualquier mutación determinada y comparar el número de mutaciones presentes en la muestra con el número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos menores dentro de un contenedor de fracciones de alelos menores que incluye la fracción de alelos menores de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba, y generar una carga de mutación normalizada opcionalmente con un tratamiento de inmunoterapia. La memoria del sistema también puede almacenar datos de control de varias poblaciones con diferentes tipos de cáncer. Para cualquier población de este tipo, los datos pueden incluir números de mutaciones presentes en sujetos, frecuencias de alelos menores de algunas o todas estas mutaciones, grupos de frecuencias de alelos menores caracterizados por desviaciones estándar de frecuencias de mutación promedio para números de mutaciones de muestras que caen dentro de un grupo. El sistema también puede incluir una pantalla o impresora para generar resultados, como la carga de mutación de cáncer de una muestra expresada, por ejemplo, como una puntuación Z, y/o un tratamiento futuro recomendado, como administrar inmunoterapia o continuar con la inmunoterapia. El sistema también puede incluir un teclado y/o un puntero para proporcionar información al usuario, como definir tipos de cáncer, un conjunto de mutaciones en las que se realizará el análisis o establecer umbrales, entre otros accesorios. El sistema también puede incluir un aparato de secuenciación acoplado a la memoria para proporcionar datos de secuenciación sin procesar.

[90] Diversos pasos de los métodos presentes pueden utilizar información y/o programas y generar resultados que se almacenan en medios legibles por computadora (por ejemplo, disco duro, memoria auxiliar, memoria externa, servidor;

base de datos, dispositivo de memoria portátil (por ejemplo, CD-R, DVD, disco ZIP, tarjetas de memoria flash), y similares. Por ejemplo, la información utilizada y los resultados generados por los métodos que se pueden almacenar en medios legibles por computadora incluyen datos de control de varias poblaciones con diferentes tipos de cáncer como se describió anteriormente, secuencias de referencias, datos de secuenciación sin procesar, ácidos nucleicos secuenciados, mutaciones, fracciones de alelos menores, medidas de carga de mutación normalizada, como puntuaciones Z, umbrales y regímenes de tratamiento de inmunoterapia asociados con cargas de mutación normalizadas por encima de un umbral en varios tipos de cáncer.

**[91]** La presente divulgación también incluye un kit que comprende instrucciones para proporcionar una medida de la carga de mutación tumoral en una muestra. El kit puede incluir un medio legible por máquina que contiene uno o más programas que, cuando se ejecutan, implementan los pasos de los presentes métodos. El kit puede no incluir un medio físico legible por máquina, sino más bien acceso a la nube o a un almacenamiento de datos en línea que proporciona una plataforma a través de la cual un usuario puede realizar un análisis de la carga de mutación tumoral en la muestra.

**[92]** La divulgación puede implementarse en hardware y/o software. Por ejemplo, diferentes aspectos de la divulgación pueden implementarse en lógica del lado del cliente o en lógica del lado del servidor. La divulgación o sus componentes pueden incorporarse en un componente de programa de medios fijos que contiene instrucciones lógicas y/o datos que, cuando se cargan en un dispositivo informático configurado adecuadamente, hacen que ese dispositivo funcione de acuerdo con la divulgación. Un medio fijo que contiene instrucciones lógicas puede entregarse a un espectador en un medio fijo para cargarlo físicamente en la computadora de un espectador o un medio fijo que contiene instrucciones lógicas puede residir en un servidor remoto al que un espectador accede a través de un medio de comunicación para descargar un componente de programa.

**[93]** La presente divulgación proporciona sistemas de control informático que están programados para implementar métodos de la divulgación. La FIG. 2 muestra un sistema informático 901 que está programado o configurado de otro modo para implementar métodos de la presente divulgación. El sistema informático 901 incluye una unidad central de procesamiento (CPU, también "procesador" y "procesador informático" en el presente documento) 905, que puede ser un procesador de un solo núcleo o de múltiples núcleos, o una pluralidad de procesadores para procesamiento en paralelo. El sistema informático 901 también incluye una memoria o ubicación de memoria 910 (por ejemplo, memoria de acceso aleatorio, memoria de solo lectura, memoria flash), una unidad de almacenamiento electrónico 915 (por ejemplo, disco duro), una interfaz de comunicación 920 (por ejemplo, adaptador de red) para comunicarse con uno o más sistemas diferentes, y dispositivos periféricos 925, como caché, otra memoria, almacenamiento de datos y/o adaptadores de pantalla electrónica. La memoria 910, la unidad de almacenamiento 915, la interfaz 920 y los dispositivos periféricos 925 están en comunicación con la CPU 905.

**[94]** A través de un bus de comunicación (líneas continuas), como una placa base. La unidad de almacenamiento 915 puede ser una unidad de almacenamiento de datos (o repositorio de datos) para almacenar datos. El sistema informático 901 puede estar acoplado operativamente a una red informática ("red") 930 con la ayuda de la interfaz de comunicación 920. La red 930 puede ser Internet, una red de Internet y/o extranet, o una intranet y/o extranet que esté en comunicación con Internet. La red 930 en algunos casos es una red de telecomunicaciones y/o datos. La red 930 puede incluir una red de área local. La red 930 puede incluir uno o más servidores informáticos, que pueden permitir la computación distribuida, como la computación en la nube. La red 930, en algunos casos con la ayuda del sistema informático 901, puede implementar una red de igual a igual, que puede permitir que los dispositivos acoplados al sistema informático 901 se comporten como un cliente o un servidor.

**[95]** La CPU 905 puede ejecutar una secuencia de instrucciones legibles por máquina, que pueden estar incorporadas en un programa o software. Las instrucciones pueden estar almacenadas en una ubicación de memoria, como la memoria 910. Las instrucciones pueden dirigirse a la CPU 905, que posteriormente puede programar o configurar de otro modo la CPU 905 para implementar métodos de la presente divulgación. Los ejemplos de operaciones realizadas por la CPU 905 pueden incluir búsqueda, decodificación, ejecución y escritura diferida.

**[96]** La CPU 905 puede ser parte de un circuito, como un circuito integrado. Uno o más componentes adicionales del sistema 901 pueden estar incluidos en el circuito. En algunos casos, el circuito es un circuito integrado de aplicación específica (ASIC).

**[97]** La unidad de almacenamiento 915 puede almacenar archivos, como controladores, bibliotecas y programas guardados. La unidad de almacenamiento 915 puede almacenar datos de usuario, por ejemplo, preferencias de usuario y programas de usuario. El sistema informático 901 en algunos casos puede incluir una o más unidades de almacenamiento de datos adicionales que son externas al sistema informático 901, como, por ejemplo, ubicadas en un servidor remoto que está en comunicación con el sistema informático 901 a través de una intranet o Internet.

**[98]** El sistema informático 901 puede comunicarse con uno o más sistemas informáticos remotos a través de la red 930. Por ejemplo, el sistema informático 901 puede comunicarse con un sistema informático remoto de un usuario. Entre los ejemplos de sistemas informáticos remotos se incluyen ordenadores personales (por ejemplo, PC portátiles), tabletas o pizarras electrónicas (por ejemplo, iPad de Apple®, Galaxy Tab de Samsung®), teléfonos, teléfonos inteligentes (por

ejemplo, iPhone de Apple®, dispositivo habilitado con Android, Blackberry®) o asistentes digitales personales. El usuario puede acceder al sistema informático 901 a través de la red 930.

**[99]** Los métodos descritos en este documento se pueden implementar por medio de un código ejecutable por máquina (por ejemplo, un procesador de ordenador) almacenado en una ubicación de almacenamiento electrónico del sistema informático 901, como, por ejemplo, en la memoria 910 o en la unidad de almacenamiento electrónico 915. El código ejecutable por máquina o legible por máquina se puede proporcionar en forma de software. Durante el uso, el código se puede ejecutar por el procesador 905. En algunos casos, el código se puede recuperar de la unidad de almacenamiento 915 y almacenar en la memoria 910 para que el procesador 905 pueda acceder a él fácilmente. En algunas situaciones, se puede excluir la unidad de almacenamiento electrónico 915 y las instrucciones ejecutables por máquina se almacenan en la memoria 910.

**[100]** El código puede ser precompilado y configurado para su uso con una máquina que tenga un procesador adaptado para ejecutar el código, o puede ser compilado durante el tiempo de ejecución. El código puede ser suministrado en un lenguaje de programación que puede ser seleccionado para permitir que el código se ejecute de manera precompilada o tal como fue compilado.

**[101]** Los aspectos de los sistemas y métodos proporcionados en este documento, como el sistema informático 901, pueden incorporarse en la programación. Diversos aspectos de la tecnología pueden considerarse como "productos" o "artículos de fabricación" típicamente en forma de código ejecutable por máquina (o procesador) y/o datos asociados que se transportan o incorporan en un tipo de medio legible por máquina. El código ejecutable por máquina puede almacenarse en una unidad de almacenamiento electrónico, como una memoria (por ejemplo, memoria de solo lectura, memoria de acceso aleatorio, memoria flash ) o un disco duro.

**[102]** Los medios de tipo "almacenamiento" pueden incluir cualquiera o toda la memoria tangible de los ordenadores, procesadores o similares, o módulos asociados a los mismos, como diversas memorias de semiconductores, unidades de cinta, unidades de disco y similares, que pueden proporcionar almacenamiento no transitorio en cualquier momento para la programación del software. Todo o partes del software pueden comunicarse en ocasiones a través de Internet o de varias otras redes de telecomunicaciones. Dichas comunicaciones, por ejemplo, pueden permitir la carga del software desde un ordenador o procesador a otro, por ejemplo, desde un servidor de gestión o un ordenador anfitrión a la plataforma informática de un servidor de aplicaciones. Por tanto, otro tipo de medios que pueden soportar los elementos del software incluyen ondas ópticas, eléctricas y electromagnéticas, como las que se utilizan en interfaces físicas entre dispositivos locales, a través de redes terrestres cableadas y ópticas y en varios enlaces aéreos. Los elementos físicos que transportan dichas ondas, como enlaces cableados o inalámbricos, enlaces ópticos o similares, también pueden considerarse como medios que transportan el software. Tal como se utiliza en el presente documento, a menos que se restrinja a medios no transitorios, tangibles,

**[103]** Los medios de "almacenamiento", términos como "medio legible por computadora" o "medio legible por máquina" se refieren a cualquier medio que participe en el suministro de instrucciones a un procesador para su ejecución.

**[104]** Por lo tanto, un medio legible por máquina, como un código ejecutable por computadora, puede adoptar muchas formas, incluyendo, entre otras, un medio de almacenamiento tangible, un medio de onda portadora o un medio de transmisión física. Los medios de almacenamiento no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, como cualquiera de los dispositivos de almacenamiento en cualquier computadora o similar, como los que se pueden utilizar para implementar las bases de datos, etc., que se muestran en los dibujos. Los medios de almacenamiento volátiles incluyen memoria dinámica, como la memoria principal de dicha plataforma informática. Los medios de transmisión tangibles incluyen cables coaxiales, alambre de cobre y fibra óptica, incluidos los cables que forman un bus dentro de un sistema informático. Los medios de transmisión de onda portadora pueden adoptar la forma de señales eléctricas o electromagnéticas, u ondas acústicas o de luz como las generadas durante las comunicaciones de datos por radiofrecuencia (RF) e infrarrojos (IR). Por lo tanto, las formas comunes de medios legibles por computadora incluyen, por ejemplo: un disquete, un disco flexible, un disco duro, una cinta magnética, cualquier otro medio magnético, un CD-ROM, un DVD o un DVD-ROM, cualquier otro medio óptico, tarjetas perforadas, cinta de papel, cualquier otro medio de almacenamiento físico con patrones de agujeros, una RAM, una ROM, una PROM y una EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, una onda portadora que transporte datos o instrucciones, cables o enlaces que transporten dicha onda portadora, o cualquier otro medio desde el cual una computadora pueda leer código de programación y/o datos. Muchas de estas formas de medios legibles por computadora pueden estar involucradas en llevar una o más secuencias de una o más instrucciones a un procesador para su ejecución.

**[105]** El sistema informático 901 puede incluir o estar en comunicación con una pantalla electrónica 935 que comprende una interfaz de usuario (IU) 940 para proporcionar, por ejemplo, un informe. Los ejemplos de IU incluyen, sin limitación, una interfaz gráfica de usuario (IGU) y una interfaz de usuario basada en web.

**[106]** Los métodos y sistemas de la presente divulgación pueden implementarse mediante uno o más algoritmos. Un algoritmo puede implementarse mediante software al ser ejecutado por la unidad central de procesamiento 905.

## VI. Características generales de los métodos

## 1. Muestras

**[107]** Una muestra puede ser cualquier muestra biológica aislada de un sujeto. Las muestras pueden incluir tejidos corporales, tales como tumores sólidos conocidos o sospechosos, sangre completa, plaquetas, suero, plasma, heces, glóbulos rojos, glóbulos blancos o leucocitos, células endoteliales, biopsias de tejido, líquido cefalorraquídeo, líquido sinovial, líquido linfático, líquido ascítico, líquido intersticial o extracelular, el líquido en espacios entre células, incluyendo líquido crevicular gingival, médula ósea, derrames pleurales, líquido cefalorraquídeo, saliva, mucosidad, esputo, semen, sudor, orina. Las muestras son preferiblemente fluidos corporales, particularmente sangre y fracciones de la misma tales como plasma y suero, y orina. Dichas muestras incluyen ácidos nucleicos desprendidos de tumores. Los ácidos nucleicos pueden incluir ADN y ARN y pueden estar en formas bicatenarias y/o monocatenarias. Una muestra puede estar en la forma en que se aisló originalmente de un sujeto o puede haber sido sometida a un procesamiento posterior para eliminar o agregar componentes, como células, enriquecer un componente en relación con otro o convertir una forma de ácido nucleico en otra, como ARN en ADN o ácidos nucleicos monocatenarios en bicatenarios. Así, por ejemplo, un fluido corporal para análisis es plasma o suero que contiene ácidos nucleicos libres de células, por ejemplo, ADN libre de células (ADNlc).

**[108]** El volumen del fluido corporal puede depender de la profundidad de lectura deseada para las regiones secuenciadas. Los volúmenes ejemplares son 0,4-40 ml, 5-20 ml, 10-20 ml. Por ejemplo, el volumen puede ser 0,5 ml, 1 ml, 5 ml, 10 ml, 20 ml, 30 ml o 40 ml. Un volumen de plasma muestreado puede ser de 5 a 20 ml.

**[109]** La muestra puede comprender diversas cantidades de ácido nucleico que contienen equivalentes de genoma. Por ejemplo, una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10.000 ( $10^4$ ) equivalentes de genoma humano haploide y, en el caso de ADNlc, aproximadamente 200 mil millones ( $2 \times 10^4$ ) de moléculas de polinucleótidos individuales. De manera similar, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30.000 equivalentes de genoma humano haploide y, en el caso de ADNlc, aproximadamente 600 mil millones de moléculas individuales.

**[110]** Una muestra puede comprender ácidos nucleicos de diferentes fuentes, por ejemplo, de células y de células libres. Una muestra puede comprender ácidos nucleicos que portan mutaciones. Por ejemplo, una muestra puede comprender ADN que porta mutaciones de la línea germinal y/o mutaciones somáticas. Una muestra puede comprender ADN que porta mutaciones asociadas al cáncer (por ejemplo, mutaciones somáticas asociadas al cáncer).

**[111]** Las cantidades ejemplares de ácidos nucleicos libres de células en una muestra antes de la amplificación varían de aproximadamente 1 fg a aproximadamente 1  $\mu$ g, por ejemplo, de 1 pg a 200 ng, de 1 ng a 100 ng, de 10 ng a 1000 ng. Por ejemplo, la cantidad puede ser de hasta aproximadamente 600 ng, hasta aproximadamente 500 ng, hasta aproximadamente 400 ng, hasta aproximadamente 300 ng, hasta aproximadamente 200 ng, hasta aproximadamente 100 ng, hasta aproximadamente 50 ng o hasta aproximadamente 20 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de al menos 1 fg, al menos 10 fg, al menos 100 fg, al menos 1 pg, al menos 10 pg, al menos 100 pg, al menos 1 ng, al menos 10 ng, al menos 100 ng, al menos 150 ng o al menos 200 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de hasta 1 femtogramo (fg), 10 fg, 100 fg, 1 picogramo (pg), 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 150 ng o 200 ng de moléculas de ácido nucleico libres de células. El método puede comprender la obtención de 1 femtogramo (fg) a 200 ng.

**[112]** Una muestra de ácido nucleico libre de células se refiere a una muestra que contiene ácidos nucleicos libres de células. Los ácidos nucleicos libres de células son ácidos nucleicos que no están contenidos dentro de una célula ni unidos de otro modo a ella o, en otras palabras, ácidos nucleicos que permanecen en una muestra de células intactas extraídas. Los ácidos nucleicos libres de células pueden referirse a todos los ácidos nucleicos no encapsulados procedentes de un fluido corporal (por ejemplo, sangre, orina, LCR, etc.) de un sujeto. Los ácidos nucleicos libres de células incluyen ADN (ADNlc), ARN (ARNlc) e híbridos de los mismos, incluyendo ADN genómico, ADN mitocondrial, ADN circulante, ARNip, miARN, ARN circulante (ARNc), ARNt, ARNr, ARN nucleolar pequeño (ARNnp), ARN que interactúa con Piwi (ARNpi), ARN largo no codificante (ARNnc largo) o fragmentos de cualquiera de estos. Los ácidos nucleicos libres de células pueden ser bicatenarios, monocatenarios o un híbrido de los mismos. Un ácido nucleico libre de células puede liberarse en el líquido corporal a través de procesos de secreción o muerte celular, por ejemplo, necrosis celular y apoptosis. Algunos ácidos nucleicos libres de células se liberan en el líquido corporal a partir de células cancerosas, por ejemplo, ADN tumoral circulante (ADNtc). Otros se liberan a partir de células sanas. El ADNtc puede ser ADN fragmentado derivado de tumores no encapsulado. El ADN fetal libre de células (ADNflc) es ADN fetal que circula libremente en el torrente sanguíneo materno.

**[113]** Un ácido nucleico libre de células o proteínas asociadas a él pueden tener una o más modificaciones epigenéticas, por ejemplo, un ácido nucleico libre de células puede estar acetilado, 5-metilado, ubiquitinado, fosforilado, sumoilado, ribosilado y/o citrulinado.

**[114]** Los ácidos nucleicos libres de células tienen una distribución de tamaño ejemplar de aproximadamente 100-500 nucleótidos, con moléculas de 110 a aproximadamente 230 nucleótidos que representan aproximadamente el 90 % de las moléculas, con una moda de aproximadamente 168 nucleótidos en humanos y un segundo pico menor en un rango entre 240 a 440 nucleótidos. Los ácidos nucleicos libres de células pueden tener aproximadamente de 160 a

aproximadamente 180 nucleótidos, o aproximadamente de 320 a aproximadamente 360 nucleótidos, o aproximadamente de 440 a aproximadamente 480 nucleótidos.

**[115]** Los ácidos nucleicos libres de células se pueden aislar de los fluidos corporales mediante un paso de partición en el que los ácidos nucleicos libres de células, tal como se encuentran en solución, se separan de las células intactas y otros componentes no solubles del fluido corporal. La partición puede incluir técnicas como la centrifugación o la filtración. Alternativamente, las células en fluidos corporales se pueden lisar y los ácidos nucleicos libres de células y celulares se pueden procesar juntos. Generalmente, después de la adición de tampones y pasos de lavado, los ácidos nucleicos libres de células se pueden precipitar con un alcohol. Se pueden utilizar pasos de limpieza adicionales, como columnas a base de sílice, para eliminar contaminantes o sales. Se pueden añadir ácidos nucleicos portadores a granel no específicos, por ejemplo, a lo largo de la reacción para optimizar ciertos aspectos del procedimiento, como el rendimiento.

**[116]** Después de dicho procesamiento, las muestras pueden incluir diversas formas de ácidos nucleicos, incluyendo ADN bicatenario, ADN monocatenario y/o ARN monocatenario. Opcionalmente, el ADN monocatenario y/o el ARN monocatenario se pueden convertir en formas bicatenarias para que se incluyan en los pasos de procesamiento y análisis posteriores.

## 2. Amplificación

**[117]** Los ácidos nucleicos de muestra flanqueados por adaptadores se pueden amplificar mediante PCR y otros métodos de amplificación que normalmente se preparan a partir de cebadores que se unen a los sitios de unión de cebadores en adaptadores que flanquean una molécula de ADN que se va a amplificar. Los métodos de amplificación pueden implicar ciclos de extensión, desnaturalización y anillamiento resultantes del termociclado o pueden ser isotérmicos como en la amplificación mediada por transcripción. Otros métodos de amplificación incluyen la reacción en cadena de la ligasa, la amplificación por desplazamiento de cadena, la amplificación basada en secuencias de ácidos nucleicos y la replicación basada en secuencias autosostenidas.

**[118]** Se pueden aplicar una o más amplificaciones para introducir códigos de barras en una molécula de ácido nucleico utilizando métodos de amplificación de ácido nucleico convencionales. La amplificación se puede llevar a cabo en una o más mezclas de reacción. Los códigos de barras se pueden introducir simultáneamente o en cualquier orden secuencial. Los códigos de barras se pueden introducir antes y/o después de la captura de secuencia. En algunos casos, solo se introducen códigos de barras para etiquetar moléculas de ácido nucleico individuales antes de la captura de la sonda, mientras que los códigos de barras para etiquetar muestras se introducen después de la captura de secuencia. En algunos casos, tanto los códigos de barras para etiquetar ácidos nucleicos individuales como los códigos de barras para etiquetar muestras se introducen antes de la captura de la sonda. En algunos casos, los códigos de barras para etiquetar muestras se introducen después de la captura de secuencia. Normalmente, la captura de secuencia implica introducir una molécula de ácido nucleico monocatenario complementaria a una secuencia diana, por ejemplo, una secuencia codificante de una región genómica y la mutación de dicha región está asociada con un tipo de cáncer. Por lo general, las amplificaciones generan una pluralidad de amplicones de ácidos nucleicos con códigos de barras únicos o no únicos, con códigos de barras que etiquetan ácidos nucleicos individuales y/o muestras con un tamaño que varía de 200 nt a 700 nt, 250 nt a 350 nt o 320 nt a 550 nt. En algunas formas de realización, los amplicones tienen un tamaño de aproximadamente 300 nt. En algunas formas de realización, los amplicones tienen un tamaño de aproximadamente 500 nt.

## 3. Código de barras

**[119]** Los códigos de barras se pueden incorporar o unir de otro modo a los adaptadores mediante síntesis química, ligadura, PCR de extensión por superposición, entre otros métodos. En general, la asignación de códigos de barras únicos o no únicos en reacciones sigue los métodos y sistemas descritos en las solicitudes de patente de EE. UU. 20010053519, 20110160078 y las patentes de EE. UU. n.º 6.582.908 y n.º 7.537.898 y n.º 9.598.731.

**[120]** Los códigos de barras se pueden vincular a ácidos nucleicos de muestra de forma aleatoria o no aleatoria. En algunos casos, se introducen en una proporción esperada de códigos de barras (por ejemplo, una combinación de códigos de barras únicos o no únicos) a micropocillos. Por ejemplo, los códigos de barras se pueden cargar de modo que se carguen más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10 000, 50 000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 de códigos de barras por muestra de genoma. En algunos casos, los códigos de barras se pueden cargar de manera que se carguen menos de aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10 000, 50 000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 de códigos de barras por muestra de genoma. En algunos casos, el número promedio de códigos de barras cargados por muestra de genoma es menor o mayor que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10000, 50 000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 códigos de barras por muestra de genoma. Los códigos de barras pueden ser únicos o no únicos.

**[121]** Un formato preferido utiliza entre 20 y 50 códigos de barras diferentes, ligados a ambos extremos de una molécula diana, creando 20-50 x 20-50 etiquetas, por ejemplo, entre 400 y 2500 códigos de barras. Dicha cantidad de códigos de barras es suficiente para que diferentes moléculas que tienen los mismos puntos de inicio y de fin tengan una alta probabilidad (por ejemplo, al menos 94 %, 99,5 %, 99,99 %, 99,999 %) de recibir diferentes combinaciones de etiquetas.

**[122]** En algunos casos, los códigos de barras pueden ser oligonucleótidos de secuencia predeterminados o aleatorios o semialeatorios. En otros casos, se puede utilizar una pluralidad de códigos de barras de modo que los códigos de barras no sean necesariamente únicos entre sí en la pluralidad. En este ejemplo, los códigos de barras se pueden unir (por ejemplo, mediante ligadura o amplificación por PCR) a moléculas individuales de modo que la combinación del código de barras y la secuencia a la que se puede unir crea una secuencia única que se puede rastrear individualmente. Como se describe en el presente documento, la detección de códigos de barras no únicos en combinación con datos de secuencia de partes de inicio (inicio) y final (fin) de lecturas de secuencia puede permitir la asignación de una identidad única a una molécula particular. La longitud o el número de pares de bases de una lectura de secuencia individual también se puede utilizar para asignar una identidad única a dicha molécula. Como se describe en el presente documento, los fragmentos de una sola hebra de ácido nucleico a los que se les ha asignado una identidad única pueden permitir de ese modo la identificación posterior de fragmentos de la hebra original y/o una hebra complementaria.

#### 4. Secuenciación

**[123]** Los ácidos nucleicos de muestra, opcionalmente flanqueados por adaptadores, con o sin amplificación previa, pueden ser objeto de secuenciación. Los métodos de secuenciación incluyen, por ejemplo, secuenciación de Sanger, secuenciación de alto rendimiento, pirosecuenciación, secuenciación por síntesis, secuenciación de moléculas individuales, secuenciación de nanoporos, secuenciación de semiconductores, secuenciación por ligación, secuenciación por hibridación, RNA-Seq (Illumina), expresión génica digital (Helicos), secuenciación de próxima generación (NGS), secuenciación de moléculas individuales por síntesis (SMSS) (Helicos), secuenciación masiva en paralelo, matriz de moléculas individuales clonales (Solexa), secuenciación shotgun, Ion Torrent, Oxford Nanopore, Roche Genia, secuenciación Maxim-Gilbert, secuenciación por primer walking, secuenciación utilizando plataformas PacBio, SOLiD, Ion Torrent o Nanopore. Las reacciones de secuenciación se pueden realizar en una variedad de unidades de procesamiento de muestras, que pueden ser de múltiples carriles, múltiples canales, múltiples pocillos u otros medios para procesar múltiples conjuntos de muestras de manera prácticamente simultánea. La unidad de procesamiento de muestras también puede incluir múltiples cámaras de muestras para permitir el procesamiento de múltiples ejecuciones simultáneamente.

**[124]** Las reacciones de secuenciación se pueden realizar en uno o más tipos de fragmentos que se sabe que contienen marcadores de cáncer u otra enfermedad. Las reacciones de secuenciación también se pueden realizar en cualquier fragmento de ácido nucleico presente en la muestra. Las reacciones de secuenciación pueden proporcionar una cobertura de secuencia del genoma de al menos 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 %. En otros casos, la cobertura de secuencia del genoma puede ser inferior a 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 %.

**[125]** Se pueden realizar reacciones de secuenciación simultáneas utilizando secuenciación multiplex. En algunos casos, los polinucleótidos libres de células se pueden secuenciar con al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000, 100 000 reacciones de secuenciación. En otros casos, los polinucleótidos libres de células se pueden secuenciar con menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000, 100 000 reacciones de secuenciación. Las reacciones de secuenciación se pueden realizar de forma secuencial o simultánea. El análisis posterior de los datos puede realizarse en todas o en parte de las reacciones de secuenciación. En algunos casos, el análisis de los datos puede realizarse en al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000, 100 000 reacciones de secuenciación. En otros casos, el análisis de los datos puede realizarse en menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000, 100 000 reacciones de secuenciación. Una profundidad de lectura ejemplar es de 1000 a 50 000 lecturas por locus (base).

**[126]** En algunos métodos, se prepara una población de ácidos nucleicos para la secuenciación mediante la terminación enzimática de ácidos nucleicos bicatenarios con salientes monocatenarios en uno o ambos extremos. La población se puede tratar con una proteína con una actividad de polimerasa 5'-3' y una actividad de exonucleasa 3'-5' en presencia de los nucleótidos (por ejemplo, A, C, G y T o U). Las proteínas ejemplares son el fragmento grande de Klenow y la polimerasa T4. En los salientes 5', la proteína extiende el extremo 3' hundido en la cadena opuesta hasta que está al mismo nivel que el extremo 5' produciendo un extremo romo. En los salientes 3', la proteína se digiere desde el extremo 3' hasta y, a veces, más allá del extremo 5' de la cadena opuesta. Si la digestión continúa más allá del extremo 5' de la cadena opuesta, el hueco se puede rellenar mediante la actividad de la polimerasa como en el caso de un saliente 5'. La terminación roma de los ácidos nucleicos bicatenarios facilita la unión de adaptadores y la amplificación posterior.

**[127]** Las poblaciones de ácidos nucleicos pueden ser sometidas a un procesamiento adicional, como la conversión de ácidos nucleicos monocatenarios a bicatenarios y/o la conversión de ARN a ADN. Estas formas de ácidos nucleicos también pueden unirse a adaptadores y amplificarse.

**[128]** Con o sin amplificación previa, los ácidos nucleicos sujetos a terminación roma como se describió anteriormente, y opcionalmente otros ácidos nucleicos en una muestra, pueden secuenciarse para producir ácidos nucleicos secuenciados. Un ácido nucleico secuenciado puede referirse a la secuencia de un ácido nucleico o a un ácido nucleico cuya secuencia ha sido determinada. La secuenciación puede realizarse de manera que proporcione datos de secuencia de moléculas de ácido nucleico individuales en una muestra, ya sea directamente o indirectamente a partir de una secuencia de consenso de productos de amplificación de una molécula de ácido nucleico individual en la muestra.



**[129]** En algunos métodos, los ácidos nucleicos de doble cadena con salientes de cadena sencilla en una muestra después de la terminación roma se unen en ambos extremos a adaptadores que incluyen códigos de barras, y la secuenciación determina las secuencias de ácidos nucleicos, así como los códigos de barras en línea introducidos por los adaptadores. Las moléculas de ADN de extremos romos se pueden ligar por extremos romos con un extremo romo de un adaptador al menos parcialmente de doble cadena (por ejemplo, un adaptador en forma de Y o de campana). Alternativamente, los extremos romos de los ácidos nucleicos de muestra y los adaptadores se pueden unir con nucleótidos complementarios para facilitar la ligadura.

**[130]** La muestra puede ponerse en contacto con una cantidad suficiente de adaptadores de modo que exista una baja probabilidad (por ejemplo,  $<1$  o  $0,1\%$ ) de que dos instancias cualesquiera del mismo ácido nucleico reciban la misma combinación de códigos de barras de adaptadores de los adaptadores unidos en ambos extremos. El uso de adaptadores de esta manera permite la identificación de familias de secuencias de ácidos nucleicos con los mismos puntos de inicio y de fin en un ácido nucleico de referencia y unidos a la misma combinación de códigos de barras. Dicha familia representa secuencias de productos de amplificación de un ácido nucleico en la muestra antes de la amplificación. Las secuencias de los miembros de la familia pueden compilarse para derivar nucleótidos de consenso o una secuencia de consenso completa para una molécula de ácido nucleico en la muestra original, modificada por terminación roma y unión de adaptadores. En otras palabras, el nucleótido que ocupa una posición específica de un ácido nucleico en la muestra se determina como el consenso de nucleótidos que ocupan esa posición correspondiente en las secuencias de miembros de la familia. Las familias pueden incluir secuencias de una o ambas cadenas de un ácido nucleico bicatenario. Si los miembros de una familia incluyen secuencias de ambas cadenas de un ácido nucleico bicatenario, las secuencias de una cadena se convierten en su complemento con el fin de compilar todas las secuencias para derivar nucleótidos o secuencias de consenso. Algunas familias incluyen solo una secuencia de un solo miembro. En este caso, esta secuencia se puede tomar como la secuencia de un ácido nucleico en la muestra antes de la amplificación. Alternativamente, las familias con solo una secuencia de un solo miembro se pueden eliminar del análisis posterior.

**[131]** Las variaciones de nucleótidos en los ácidos nucleicos secuenciados se pueden determinar comparando los ácidos nucleicos secuenciados con una secuencia de referencia. La secuencia de referencia es a menudo una secuencia conocida, por ejemplo, una secuencia de genoma completo o parcial conocida de un objeto, una secuencia de genoma completo de un objeto humano. La secuencia de referencia puede ser hG19. Los ácidos nucleicos secuenciados pueden representar secuencias determinadas directamente para un ácido nucleico en una muestra, o un consenso de secuencias de productos de amplificación de dicho ácido nucleico, como se describió anteriormente. Se puede realizar una comparación en una o más posiciones designadas en una secuencia de referencia. Se puede identificar un subconjunto de ácidos nucleicos secuenciados que incluya una posición que corresponda con una posición designada de la secuencia de referencia cuando las secuencias respectivas estén alineadas al máximo. Dentro de dicho subconjunto se puede determinar cuáles, si los hay, los ácidos nucleicos secuenciados incluyen una variación de nucleótido en la posición designada, y opcionalmente cuáles, si los hay, incluyen un nucleótido de referencia (es decir, el mismo que en la secuencia de referencia). Si el número de ácidos nucleicos secuenciados en el subconjunto que incluye una variante de nucleótido supera un umbral, entonces se puede llamar a un nucleótido variante en la posición designada. El umbral puede ser un número simple, tal como al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 ácidos nucleicos secuenciados dentro del subconjunto que incluye la variante de nucleótido o puede ser una proporción, tal como al menos 0,5, 1, 2, 3, 4, 5, 10, 15 o 20 de ácidos nucleicos secuenciados dentro del subconjunto que incluye la variante de nucleótido, entre otras posibilidades. La comparación se puede repetir para cualquier posición designada de interés en la secuencia de referencia. A veces se puede realizar una comparación para posiciones designadas que ocupan al menos 20, 100, 200 o 300 posiciones contiguas en una secuencia de referencia, por ejemplo, 20-500 o 50-300 posiciones contiguas.

**[132]** Si se asocian distintas versiones de una secuencia a un número de acceso en distintos momentos, se hace referencia a la versión asociada al número de acceso en la fecha de presentación efectiva de esta solicitud. La fecha de presentación efectiva significa la fecha anterior entre la fecha de presentación real o la fecha de presentación de una solicitud de prioridad que haga referencia al número de acceso, si corresponde. Del mismo modo, si se publican distintas versiones de una publicación, sitio web o similar en distintos momentos, se hace referencia a la versión publicada más recientemente en la fecha de presentación efectiva de la solicitud, a menos que se indique lo contrario.

## Ejemplos

**[133]** Este ejemplo determina la distribución de las puntuaciones Z en individuos con distintos tipos de cáncer.

## Extracción de sangre, envío y aislamiento de plasma

**[134]** Toda la extracción, procesamiento y secuenciación de ADNlc se realizó en un laboratorio certificado por la CLIA y acreditado por el CAP. Brevemente, para las muestras clínicas, se aisló plasma de 10 ml de sangre completa recolectada en tubos de recolección de sangre libre de células mediante centrifugación doble, de la cual se extrajo ADNlc, se etiquetó con códigos de barras no aleatorios y se utilizaron de 5 a 30 ng para preparar bibliotecas de secuenciación, que luego se enriquecieron mediante captura híbrida, se agruparon y se secuenciaron mediante síntesis de extremos emparejados (NextSeq 500 y/o HiSeq 2500, Illumina, Inc.). Se generaron 34 muestras analíticas artificiales utilizando ADNlc preparado de manera similar de donantes sanos y ADNlc aislado como se mencionó anteriormente del sobrenadante de cultivo de

líneas celulares modelo y seleccionado por tamaño en serie utilizando perlas Agencourt Ampure XP (Beckman Coulter, Inc.) hasta que no quedó ADN genómico detectable.

#### Análisis bioinformático y detección de variantes

[135] Todos los análisis de detección de variantes se realizaron utilizando el flujo de trabajo bioinformático clínico bloqueado Guardant360 y se informaron sin modificaciones mediante análisis post-hoc. Todos los umbrales de decisión se determinaron utilizando cohortes de entrenamiento independientes, se bloquearon y se aplicaron prospectivamente a todas las muestras de validación y clínicas. Como se describió anteriormente [PMID 26474073], los archivos de llamadas de base generados por el software RTA de Illumina (v2.12) se desmultiplexaron utilizando bcl2fastq (v2.19) y se procesaron con una secuencia personalizada para la detección de códigos de barras de moléculas, el recorte del adaptador de secuenciación y el recorte de la calidad de la base (descartando las bases por debajo de Q20 en los extremos de las lecturas). Las lecturas procesadas se alinearon luego con hg19 utilizando BWA-MEM [Li et al. 2013 arXiv:1303.3997v2] y se utilizaron para construir representaciones de consenso de doble cadena de moléculas de ADNlc únicas originales utilizando códigos de barras inferidos y posiciones de inicio/fin de lectura. Las SNV se detectaron comparando las características de las moléculas de consenso y lectura con los perfiles de ruido de error de referencia específicos de la plataforma de secuenciación y de la posición determinados de forma independiente para cada posición en el panel mediante la secuenciación de un conjunto de entrenamiento de 62 donantes sanos tanto en NextSeq 500 como en HiSeq 2500. Los perfiles de error de SNV posicionales observados se utilizaron para definir los puntos de corte de llamada para la detección de SNV con respecto al número y las características de las moléculas variantes, que diferían según la posición pero que eran más comúnmente moléculas únicas, lo que en una muestra promedio (cobertura de 5000 moléculas únicas) corresponde a un límite de detección de -0,04 % de fracción alélica. Para detectar indels, se construyó un modelo de ruido de fondo generativo para tener en cuenta los artefactos de PCR que surgen con frecuencia en contextos homopoliméricos o repetitivos, lo que permite errores de PCR específicos de la cadena y tardíos. A continuación, la detección se determinó mediante la puntuación de la razón de verosimilitud para el soporte de la molécula variante ponderada por características observadas frente a la distribución del ruido de fondo. Los umbrales de notificación fueron específicos del evento, tal como se determinó por el rendimiento en las muestras de entrenamiento, pero lo más común fue 2 moléculas únicas para indels clínicamente procesables, lo que en una muestra promedio corresponde a un límite de detección de -0,02 % de fracción alélica. Los eventos de fusión se detectaron fusionando lecturas superpuestas de extremos emparejados para formar una representación de la molécula de ADNlc secuenciada, que luego se alineó y se mapeó a moléculas de ADNlc únicas iniciales según el código de barras y la información de alineación, incluido el recorte suave. Las lecturas recortadas suavemente se analizaron utilizando la direccionalidad y la proximidad del punto de ruptura para identificar grupos de moléculas que representan eventos de fusión candidatos, que luego se usaron para construir referencias fusionadas contra las cuales se realinearon las lecturas recortadas suavemente por el alineador en la primera pasada. Los umbrales de informe específicos se determinaron mediante análisis retrospectivos y de conjuntos de entrenamiento, pero generalmente fueron 1 molécula única posterior a la realineación que cumplía con los requisitos de calidad, lo que en una muestra promedio corresponde a un límite de detección de -0,04 % de fracción alélica. Para detectar CNA, la cobertura de moléculas únicas a nivel de sonda se normalizó para el rendimiento general de moléculas únicas, la eficiencia de la sonda, el contenido de GC y la saturación de la señal y se resumió de manera robusta a nivel del gen. Las determinaciones de CNA se basaron en umbrales de decisión establecidos en el conjunto de entrenamiento tanto para la desviación absoluta del número de copias con respecto a la línea base diploide por muestra como para la desviación de la variación de la línea base de la señal normalizada a nivel de sonda en el contexto de la variación de fondo dentro de la línea base diploide de cada muestra. La carga tumoral normalizada por muestra se determinó mediante la normalización a la carga mutacional esperada para el tipo de tumor y la fracción de ADNtc y se informó como una puntuación Z.

[136] La Figura 1 representa gráficamente las puntuaciones Z para la carga de mutación del cáncer en muestras de diferentes individuos que tienen uno de los tipos de cáncer que se muestran en los ejes X. La distribución de las puntuaciones Z varía para los diferentes tipos de cáncer, pero generalmente es asimétrica, con la moda generalmente por debajo de cero, pero con unos pocos individuos que muestran puntuaciones Z altamente positivas.

## REIVINDICACIONES

1. Un método para proporcionar una medida de la carga de mutación tumoral en una muestra de prueba de ácidos nucleicos libres de células de un sujeto que tiene un tipo de cáncer o signos de un tipo de cáncer, que comprende:

- (a) determinar, mediante una computadora, un número de mutaciones presentes en la muestra de prueba de ácidos nucleicos libres de células, y una fracción de alelos menores basada en una o más mutaciones más altamente representadas en la muestra de prueba de ácidos nucleicos libres de células; y
- (b) normalizar, mediante un ordenador, el número de mutaciones presentes en la muestra con respecto a un número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos minoritarios dentro de un compartimento de fracciones de alelos minoritarios que incluye la fracción de alelos minoritarios de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba.

2. El método de la reivindicación 1:

- (a) en el que el número de mutaciones presentes en las muestras de control es un promedio;
- (b) en el que el compartimento tiene un ancho de no más del 20 %, no más del 10 % o no más del 5 %; y/o
- (c) que comprende además determinar si el número de mutaciones presentes en la muestra está por encima de un umbral, en el que el umbral se establece para indicar un sujeto que es probable que responda positivamente a una inmunoterapia.

3. El método de la reivindicación 1 o la reivindicación 2, en el que

- (a) la normalización comprende dividir el número de mutaciones en la muestra de prueba por un número promedio de mutaciones en las muestras de control;
- (b) la normalización comprende restar del número determinado de mutaciones en la muestra de prueba de ácidos nucleicos libres de células un promedio del número de mutaciones en las muestras de control dentro del contenedor, opcionalmente comprendiendo además dividir el número de mutaciones en la muestra de prueba de ácidos nucleicos libres de células menos el número promedio de mutaciones presentes en las muestras de control por una desviación estándar del número de mutaciones presentes en las muestras de control para calcular una puntuación Z, opcionalmente en donde:

- (i) el promedio es una media; o
- (ii) las muestras de control incluyen al menos 25, 50, 100, 200 o 500 muestras de control;

(c) la normalización comprende determinar el promedio y la dispersión del número de mutaciones en al menos 10, 50, 100 o 500 muestras de control, determinar una puntuación estándar de desviación, a partir del promedio en la muestra de prueba, y determinar si la puntuación estándar está por encima de un número umbral, opcionalmente en donde:

- (i) el promedio es una media, mediana o moda;
- (ii) la dispersión se representa como varianza, desviación estándar o rango intercuartil; y/o
- (iii) la puntuación estándar de desviación es una puntuación Z; o

(d) la normalización comprende además dividir el número determinado de mutaciones en la muestra de prueba de ácidos nucleicos libres de células por el número promedio de mutaciones presentes en las muestras de control en el mismo contenedor.

4. El método de cualquiera de las reivindicaciones anteriores, en el que la normalización se implementa en un ordenador programado para almacenar valores para el número de mutaciones presentes en una pluralidad de contenedores de fracciones de alelos menores, opcionalmente en el que los valores almacenados son una media y una desviación estándar del número de mutaciones presentes en cada uno de la pluralidad de contenedores, opcionalmente en el que al menos 50.000, 100.000 o 150.000 nucleótidos están secuenciados en segmentos de los ácidos nucleicos libres de células.

5. Procedimiento según cualquiera de las reivindicaciones anteriores, que comprende determinar una puntuación estándar de carga de mutación tumoral en el sujeto y si la puntuación estándar está por encima de un umbral para sujetos de control coherente con la capacidad de respuesta a la inmunoterapia.

6. Procedimiento según la reivindicación 1, en el que (a) comprende

- (i) determinar secuencias de moléculas de ácido nucleico libres de células en la muestra de prueba y comparar las secuencias resultantes con las secuencias de referencia correspondientes para identificar el número de mutaciones presentes en la muestra y la fracción de alelos minoritarios, opcionalmente en el que las secuencias de referencia son de hG19 o hG38;

(ii) determinar la presencia o ausencia de un panel de mutaciones predeterminadas que se sabe que se producen en el cáncer del tipo presente o que se sospecha que está presente en la muestra, opcionalmente en el que: las mutaciones son mutaciones somáticas que afectan a la secuencia de una proteína codificada; o  
 (iii) unir adaptadores a los ácidos nucleicos libres de células, amplificar los ácidos nucleicos libres de células a partir de cebadores que se unen a los adaptadores y secuenciar los ácidos nucleicos amplificados, por ejemplo, en el que la secuenciación es una secuenciación de amplificación de puente, pirosecuenciación, secuenciación de semiconductores de iones, secuenciación de extremo emparejado, secuenciación por ligación o secuenciación en tiempo real de molécula única.

7. El método de cualquier reivindicación precedente, en el que el tipo de cáncer es:

- (a) un cáncer sólido;
- (b) renal, mesotelioma, tejido blando, SNC primario, tiroides, hígado, próstata, páncreas, CUP, neuroendocrino, NSCLC, gastroesofágico, cabeza y cuello, SCLC, mama, melanoma, colangiocarcinoma, cáncer ginecológico, colorrectal o urotelial;
- (c) una leucemia o linfoma; o
- (d) una neoplasia maligna hematopoyética.

8. Un sistema, que comprende:

una interfaz de comunicación que recibe, a través de una red de comunicación, lecturas de secuenciación generadas por la secuenciación de ácidos nucleicos libres de células en una muestra de prueba; y  
 un ordenador en comunicación con la interfaz de comunicación, en donde el ordenador comprende uno o más procesadores de ordenador y un medio legible por ordenador, que comprende un código ejecutable por máquina que, tras la ejecución por uno o más procesadores de ordenador, implementa un método que comprende:

- (a) recibir, a través de la red de comunicación, las lecturas de secuenciación generadas por el secuenciador de ácidos nucleicos;
- (b) determinar un número de mutaciones presentes en las lecturas de secuenciación de la muestra de prueba, y una fracción de alelos minoritarios basada en una o más mutaciones más altamente representadas en las lecturas de secuenciación de la muestra de prueba; y,
- (c) normalizar el número de mutaciones presentes en la muestra de prueba a un número de mutaciones presentes en muestras de control de otros sujetos con el mismo tipo de cáncer y una fracción de alelos minoritarios dentro de un contenedor de fracciones de alelos minoritarios que incluye la fracción de alelos minoritarios de la muestra de prueba para determinar una medida de la carga de mutación del cáncer en la muestra de prueba.

9. El sistema de la reivindicación 8, en el que el secuenciador de ácidos nucleicos secuenciar una biblioteca de secuenciación generada a partir de moléculas de ADN libre de células derivadas de un sujeto, en el que la biblioteca de secuenciación comprende las moléculas de ADN libre de células y adaptadores que comprenden códigos de barras, opcionalmente en el que la biblioteca de secuenciación comprende además códigos de barras de muestra que diferencian una muestra de una o más muestras.

10. El sistema de la reivindicación 8 o la reivindicación 9, en el que:

- (a) el medio legible por ordenador comprende una memoria, un disco duro o un servidor informático;
- (b) la red de comunicación comprende una red de telecomunicaciones, una internet, una extranet o una intranet;
- (b) la red de comunicación incluye uno o más servidores informáticos capaces de realizar computación distribuida, opcionalmente en el que la computación distribuida es computación en la nube; y/o
- (c) el ordenador está situado en un servidor informático que está situado de forma remota respecto del secuenciador de ácidos nucleicos.

11. El sistema de cualquiera de las reivindicaciones 8 a 10 en el que el secuenciador de ácidos nucleicos:

- (a) realiza la secuenciación por síntesis en la biblioteca de secuenciación para generar las lecturas de secuenciación;
- (b) realiza pirosecuenciación, secuenciación de una sola molécula, secuenciación de nanoporos, secuenciación de semiconductores, secuenciación por ligación o secuenciación por hibridación en la biblioteca de secuenciación para generar las lecturas de secuenciación;
- (c) utiliza una matriz de una sola molécula clonal derivada de la biblioteca de secuenciación para generar las lecturas de secuenciación; y/o
- (d) comprende un chip que tiene una matriz de micropocillos para secuenciar la biblioteca de secuenciación para generar las lecturas de secuenciación.

12. El sistema de cualquiera de las reivindicaciones 8-11, que comprende además una pantalla electrónica en comunicación con el ordenador a través de una red, en donde la pantalla electrónica comprende una interfaz de usuario para mostrar resultados al implementar (a)-(c).

5 13. El sistema de la reivindicación 12, en el que la interfaz de usuario es una interfaz de usuario gráfica (IUG) o interfaz de usuario basada en web.

10 14. El sistema de la reivindicación 12 o la reivindicación 13, en el que la pantalla electrónica está en una computadora personal y/o una computadora con acceso a Internet, opcionalmente en el que la computadora con acceso a Internet está ubicada en una ubicación remota de la computadora.

15

20

25

30

35

40

45

50

55

60

65

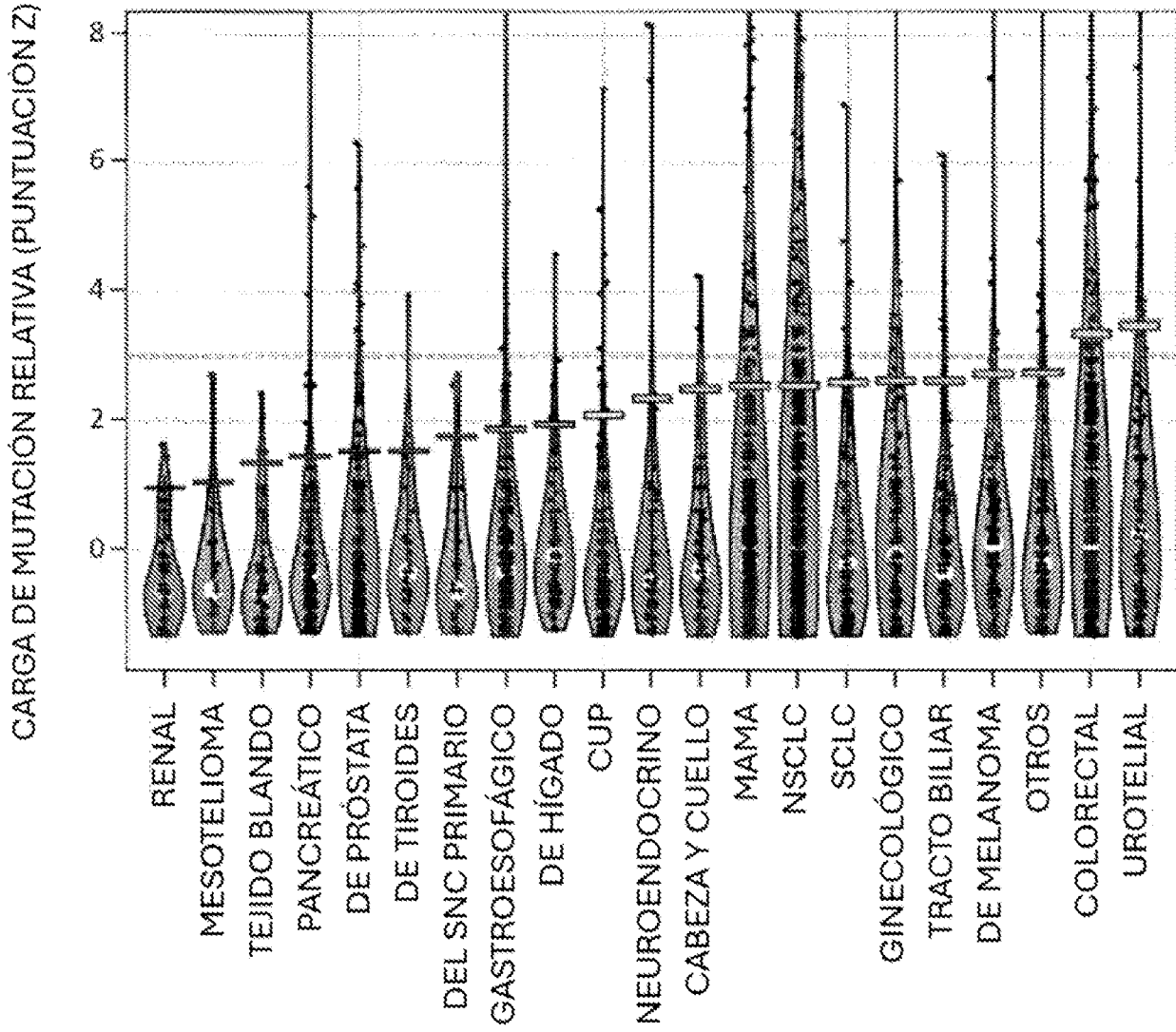


FIG. 1

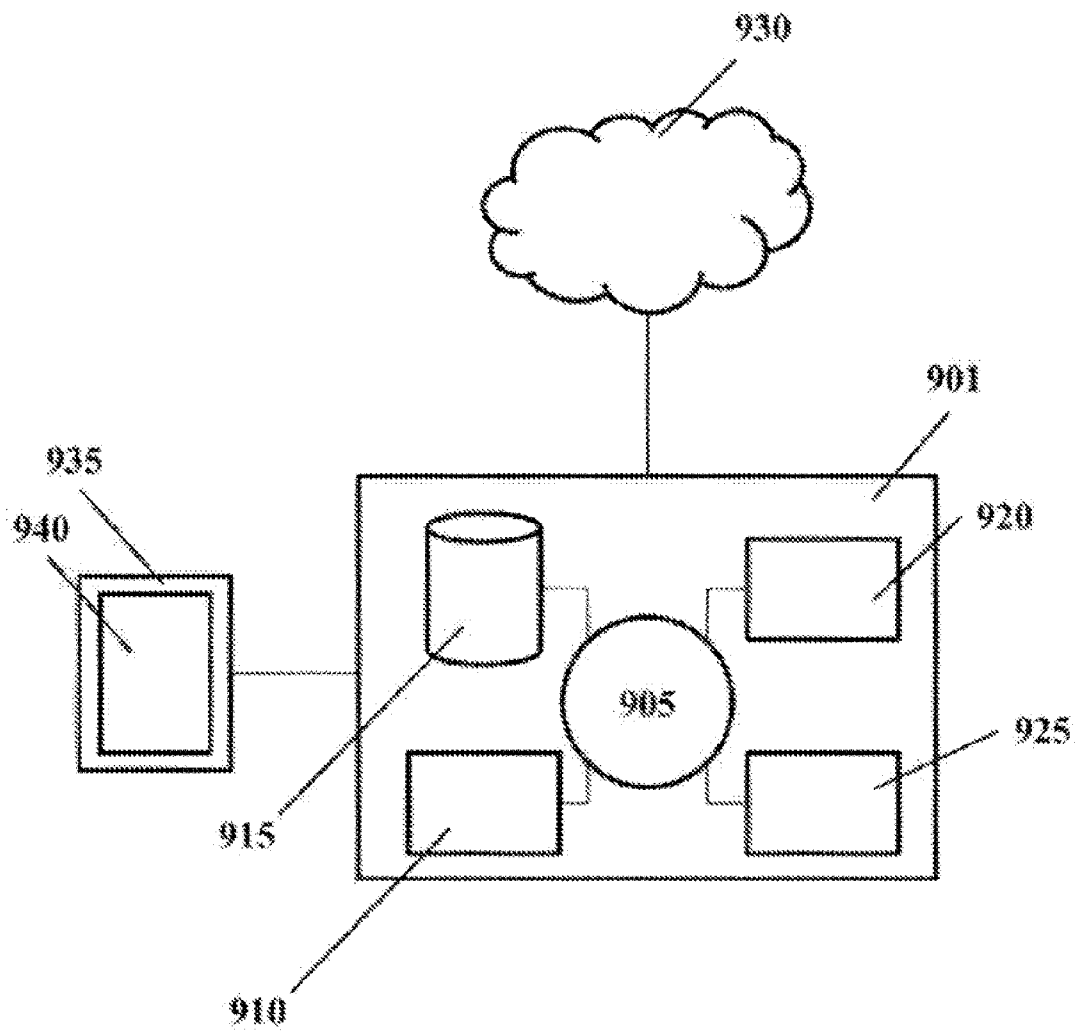


Fig. 2