

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 019 936**

51 Int. Cl.:

C12P 19/34 (2006.01)

C12Q 1/68 (2008.01)

C40B 30/04 (2006.01)

C12Q 1/6811 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.03.2017** **E 21214103 (0)**

97 Fecha y número de publicación de la concesión europea: **05.03.2025** **EP 4071250**

54 Título: **Cribado de ADN combinatorio**

30 Prioridad:

22.03.2016 US 201662311899 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
21.05.2025

73 Titular/es:

MYRIAD WOMEN'S HEALTH, INC. (100.00%)
180 Kimball Way
South San Francisco, CA 94080, US

72 Inventor/es:

MAGUIRE, JARED ROBERT;
CHU, CLEMENT;
HAQUE, IMRAN SAEEDUL;
EVANS, ERIC ANDREW y
WELKER, NOAH

74 Agente/Representante:

BERTRÁN VALLS, Silvia

ES 3 019 936 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

Cribado de ADN combinatorio

5 **Referencia cruzada a solicitudes relacionadas**

La presente solicitud reivindica la prioridad de la solicitud de patente provisional estadounidense con n.º de serie 62/311.899, titulada "Combinatorial DNA Screening", presentada el 22 de marzo de 2016.

10 **Listado de secuencias**

La presente solicitud contiene un listado de secuencias que se ha presentado electrónicamente en formato ASCII. Dicha copia en ASCII, creada el 20 de marzo de 2017, se denomina 04268_012WO1_SL.txt y tiene un tamaño de 3036 bytes.

15 **Campo técnico**

La presente divulgación se refiere a métodos para detectar mutaciones somáticas en ADN libre de células obtenido de muestras de líquidos biológicos. La sensibilidad de estos métodos permite la detección de manera no invasiva de mutaciones asociadas a una enfermedad, por ejemplo, mutaciones oncogénicas, o mutaciones hereditarias en un feto, con proporciones ultrabajas de secuencias portadoras de mutaciones somáticas en comparación con secuencias portadoras de secuencias normales, es decir, no cancerosas o de tipo natural.

25 **Antecedentes**

El descubrimiento del ADN libre de células ha fomentado la detección no invasiva de alteraciones en las secuencias genómicas que se producen en diversos estados patológicos. Sin embargo, en algunos casos, por ejemplo cáncer, la capacidad de determinar la presencia de la enfermedad mediante la detección de mutaciones asociadas a la enfermedad se ha visto obstaculizada por los niveles extremadamente bajos de ADN tumoral libre de células. Sigue siendo deseable contar con métodos que permitan la detección precisa de mutaciones asociadas a la enfermedad. Además, sigue siendo necesario determinar la fracción tumoral en pacientes con cáncer antes y después del tratamiento.

De igual manera, las mutaciones hereditarias que dan como resultado enfermedades en un feto siguen siendo difíciles de detectar debido a la baja concentración de ADN fetal libre en la sangre materna. Las pruebas genéticas prenatales permiten la detección temprana de enfermedades genéticas en un feto. Muchos trastornos genéticos fetales surgen de grandes defectos cromosómicos, tales como aneuploidía. Otros trastornos pueden ser resultado de enfermedades hereditarias o mutaciones *de novo*, tales como variantes de un solo nucleótido, variantes de múltiples nucleótidos, variantes de inserción o delección (variantes "*indel*"), o variantes del número de copias. Durante el embarazo, muchos futuros padres o médicos desean comprender el riesgo de estos trastornos en un feto para prepararse ante cualquier diagnóstico no deseado. Esto es especialmente cierto cuando ambos padres son portadores de la misma enfermedad. Por tanto, sigue siendo necesario detectar con precisión las mutaciones asociadas a enfermedades fetales de manera no invasiva.

45 El documento WO 2016/040901 A1 se refiere a la identificación y el uso de ácidos nucleicos circulantes y divulga adaptadores de polinucleótido y un método de uso de los adaptadores para detectar, diagnosticar, o determinar el pronóstico de cánceres.

50 El documento WO 2014/151117 A1 se refiere a la identificación y el uso de marcadores tumorales de ácidos nucleicos circulantes y divulga métodos para crear un selector de regiones genómicas mutadas y para usar el conjunto de selectores para analizar alteraciones genéticas en una muestra de ácido nucleico libre de células. El conjunto de selectores está diseñado de tal manera que al menos una mutación está presente en la mayoría de los sujetos con el cáncer específico.

55 **Sumario de la invención**

La invención es tal como se expone en las reivindicaciones. Específicamente, en el presente documento se proporciona un método de preparación de una fracción enriquecida en ácido desoxirribonucleico (ADN) de un paciente con cáncer, comprendiendo el método:

60 (a) extraer ADN de una muestra de líquido de plasma sanguíneo o suero sanguíneo del paciente con cáncer para obtener fragmentos de ADN libre de células (ADNlc);

65 (b) producir, a partir de los fragmentos de ADNlc extraídos de (a), una fracción de ADN con fragmentos de ADN tumoral circulante (ADNtc), en el que la fracción de ADN con fragmentos de ADNtc se produce mediante

(i) poner en contacto los fragmentos de ADNlc de (a) con una pluralidad de oligonucleótidos, en el que cada oligonucleótido comprende una secuencia de ácido nucleico que puede hibridarse a un fragmento de ADN que comprende una de un subconjunto de un conjunto de mutaciones somáticas específicas de tumor, en el que el conjunto de mutaciones somáticas específicas de tumor se determina mediante:

(A) secuenciar ADN genómico de una muestra tumoral obtenida del paciente con cáncer y secuenciar ADN genómico de una muestra no tumoral obtenida del paciente con cáncer, obteniéndose de ese modo secuencias de ADN genómico de la muestra tumoral y secuencias de ADN genómico de la muestra no tumoral;

(B) alinear las secuencias de ADN genómico de la muestra tumoral con un genoma humano de referencia que no procede del paciente y alinear las secuencias de ADN genómico de la muestra no tumoral con el genoma de referencia que no procede del paciente, obteniéndose de ese modo un conjunto de secuencias de mutación somática específicas de tumor que están presentes en las secuencias de ADN genómico de la muestra tumoral pero no están presentes en las secuencias de ADN genómico de la muestra no tumoral, y

(ii) enriquecer selectivamente los fragmentos de ADNlc extraídos de (a) con la pluralidad de oligonucleótidos,

en el que la fracción de ADN después de (b) está enriquecida con fragmentos de ADN que comprenden una o más del subconjunto de mutaciones somáticas específicas de tumor.

En algunas realizaciones, el método comprende además analizar los fragmentos de ADN enriquecidos en (b). En algunas realizaciones, analizar los fragmentos de ADN de la fracción producida en (c) comprende la secuenciación de uno o más de los fragmentos de ADN.

En algunas realizaciones, el método comprende una sensibilidad de detección de aproximadamente 20 a aproximadamente 50 fragmentos de ADNtc que comprenden una o más del subconjunto del conjunto de mutaciones somáticas en la muestra de líquido por un fondo total de aproximadamente 500.000 fragmentos de ADNlc.

En algunas realizaciones, la pluralidad de oligonucleótidos puede detectar al menos 10 mutaciones somáticas específicas de tumor diferentes. En algunas realizaciones, la extracción de ADN de la muestra de plasma sanguíneo o suero sanguíneo del paciente con cáncer para obtener los fragmentos de ADN libre de células (ADNlc) comprende separar el ADN libre de células de la muestra de plasma sanguíneo o suero sanguíneo. En algunas realizaciones, el método comprende además determinar una fracción de los fragmentos de ADN libre de células que son fragmentos de ADNtc. En algunas realizaciones, se sabe que las mutaciones somáticas están asociadas a uno cualquiera de una variedad de tumores sólidos.

También se proporciona en el presente documento un método para preparar una fracción enriquecida en ácido desoxirribonucleico (ADN) de un paciente con cáncer, que comprende:

(a) secuenciar ADN genómico de una muestra tumoral obtenida del paciente con cáncer y secuenciar ADN genómico de una muestra no tumoral obtenida del paciente con cáncer, obteniéndose de ese modo secuencias de ADN genómico de la muestra tumoral y secuencias de ADN genómico de la muestra no tumoral;

(b) alinear las secuencias de ADN genómico de la muestra tumoral con un genoma humano de referencia que no procede del paciente y alinear las secuencias de ADN genómico de la muestra no tumoral con el genoma de referencia que no procede del paciente, obteniéndose de ese modo un conjunto de secuencias de mutación somática específicas de tumor que están presentes en las secuencias de ADN genómico de la muestra tumoral pero no están presentes en las secuencias de ADN genómico de la muestra no tumoral;

(c) obtener una muestra de líquido del paciente con cáncer, en el que la muestra de líquido contiene ADN;

(d) poner en contacto el ADN de la muestra de líquido con una pluralidad de sondas de oligonucleótidos, en el que cada sonda de oligonucleótidos individual en la pluralidad de sondas de oligonucleótidos comprende una secuencia de ácido nucleico que puede hibridarse a un fragmento de ADN libre de células (ADNlc) que comprende una de un subconjunto del conjunto de secuencias de mutación somática específicas de tumor; y

(e) enriquecer fragmentos de ADN que se hibridan a una o más de la pluralidad de sondas de oligonucleótidos, preparándose de ese modo una fracción de ADN que está enriquecida para fragmentos de ADN que comprenden el subconjunto del conjunto de mutaciones somáticas específicas de tumor.

En algunas realizaciones, el método comprende además analizar la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor, en el que analizar comprende secuenciar uno o más de los fragmentos de ADN en la fracción.

5 En algunas realizaciones, la detección de una secuencia de uno o más de los fragmentos de ADN indica una recidiva del cáncer del paciente. En algunas realizaciones, la muestra de líquido es plasma sanguíneo o suero sanguíneo libre de células, y el ADN en la muestra de líquido es ADNlc que comprende tanto ADN tumoral circulante (ADNtc) como ADNlc no derivado de un tumor.

En algunas realizaciones, el método comprende además:

10 determinar una cantidad total de ADNlc en la muestra de líquido;

determinar una cantidad total de ADNtc en la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor; y

15 comparar la cantidad total de ADNtc en la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor con la cantidad total de ADNlc en la muestra de líquido para determinar una fracción tumoral para el paciente;

en el que la cantidad de ADNtc refleja la carga tumoral del paciente.

20 En algunas realizaciones, las sondas de oligonucleótidos pueden hibridar con fragmentos de ADNlc que, en conjunto, comprenden al menos 10 mutaciones somáticas específicas de tumor diferentes. En algunas realizaciones, se sabe que las mutaciones somáticas están asociadas a uno cualquiera de una variedad de tumores sólidos.

25 **Breve descripción de los dibujos**

La figura 1 ilustra el flujo de trabajo para identificar un panel de mutaciones somáticas en un paciente, por ejemplo, un paciente con cáncer, y determinar la carga tumoral del paciente a partir de secuencias seleccionadas como diana en el ADN libre de células del paciente.

30 La figura 2 ilustra los resultados de un experimento simulado por ordenador para determinar el número de moléculas tumorales en una muestra que tiene entre 3000 y 5000 copias de un genoma. Se hace referencia al ejemplo 1.

35 La figura 3 ilustra la relación entre el número de lecturas tumorales observadas y la fracción tumoral de la muestra de ácido nucleico. Se hace referencia al ejemplo 1.

La figura 4 muestra un gráfico que ilustra el nivel significativo de detección de mutaciones somáticas en una muestra que comprende secuencias "de cáncer" cada vez más diluidas. Se hace referencia al ejemplo 2.

40 La figura 5 muestra en (A) el gráfico de la figura 5 y el número de lecturas normales (B) y lecturas de cáncer (C) obtenidas para las condiciones descritas en el gráfico anterior. Se hace referencia al ejemplo 2.

La figura 6 muestra una tabla con los resultados obtenidos para la detección de sitios "de cáncer" en un fondo de 1.000.000 de secuencias en total, es decir, secuencias "de cáncer" y normales. Se hace referencia al ejemplo 2.

45 Las figuras 7A-7C son gráficos que resumen los resultados del ejemplo 3.

La figura 8 ilustra una visión general de un laboratorio de diagnóstico automático y un sistema de gestión de información de laboratorio.

50 La figura 9 ilustra una vista detallada de un sistema de ejecución de laboratorio para facilitar la automatización del laboratorio.

La figura 10 ilustra una vista detallada de un sistema de gestión de información de laboratorio para facilitar la automatización del laboratorio.

55 La figura 11 ilustra un sistema informático general en el que pueden implementarse uno o más sistemas.

La figura 12 ilustra un diagrama de flujo de trabajo a modo de ejemplo para el procesamiento de muestras.

60 La figura 13 ilustra un proceso a modo de ejemplo para controlar una unidad robótica para transportar uno o más tubos de muestra.

La figura 14 ilustra un sistema robótico para gestionar procesos automáticos de laboratorio.

65 La figura 15 ilustra una vista en ángulo de un brazo robótico.

Descripción detallada

A continuación se describirá la invención con detalle sólo a modo de referencia usando las siguientes definiciones y ejemplos.

5 A menos que se definan de otro modo en el presente documento, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que entiende habitualmente un experto habitual en la técnica a la que pertenece esta invención. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2.^a ED., John Wiley and Sons, Nueva York (1994), y Hale y Marham, THE HARPER COLLINS DICTIONARY OF
10 BIOLOGY, Harper Perennial, NY (1991) proporcionan a los expertos un diccionario general de muchos de los términos usados en esta invención. Aunque pueden usarse materiales y métodos similares o equivalentes a los descritos en el presente documento en la práctica o las pruebas de la presente invención, se describen los materiales y métodos preferidos. Se recomienda particularmente a los profesionales consultar Sambrook *et al.*, 1989, y Ausubel FM *et al.*, 1993, para definiciones y términos de la técnica. Debe entenderse que esta invención no
15 se limita a la metodología, los protocolos y los reactivos particulares descritos, ya que estos pueden variar.

Tal como se usa en el presente documento, las formas en singular “un(o)”, “una” y “el/la” incluyen la referencia en plural a menos que el contexto indique claramente lo contrario.

20 La referencia a “aproximadamente” un valor o parámetro en el presente documento incluye (y describe) variaciones relacionadas con dicho valor o parámetro en sí mismo. Por ejemplo, la descripción que hace referencia a “aproximadamente X” incluye la descripción de “X”. Además, el uso de “aproximadamente” antes de cualquier serie de números incluye “aproximadamente” cada uno de los números mencionados en esa serie. Por ejemplo, la descripción que hace referencia a “aproximadamente X, Y, o Z” pretende describir “aproximadamente X,
25 aproximadamente Y o aproximadamente Z”.

Se entiende que los aspectos y variaciones de la invención descrita en el presente documento incluyen aspectos y variaciones que “consisten” y/o “consisten esencialmente en”.

30 Un “conjunto” de lecturas se refiere a todas las lecturas de secuenciación con una hebra de ácido nucleico parental común, que pueden haber tenido o no errores introducidos durante la secuenciación o amplificación de la hebra de ácido nucleico parental.

35 Los intervalos numéricos incluyen los números que definen los intervalos. El término “aproximadamente” se usa en el presente documento para indicar más o menos el diez por ciento (10 %) de un valor. Por ejemplo, “aproximadamente 100” se refiere a cualquier número entre 90 y 110.

40 A menos que se indique de otro modo, los ácidos nucleicos se escriben de izquierda a derecha en orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en orientación amino a carboxilo, respectivamente.

45 Los encabezados que se proporcionan en el presente documento no limitan los diversos aspectos o realizaciones de la invención, que pueden obtenerse mediante referencia a la memoria descriptiva en su conjunto. Por consiguiente, los términos definidos justo a continuación se definen con mayor detalle mediante referencia a la memoria descriptiva en su conjunto.

Definiciones

50 El término “panel distintivo” se refiere a un conjunto de secuencias que comprende mutaciones somáticas que son específicas para un paciente, o marcadores que distinguen entre dos o más individuos. Un panel distintivo puede distinguir una muestra de otra.

55 El término “carga tumoral” en el presente documento se refiere a la cantidad total de material tumoral presente en un paciente, que puede reflejarse mediante la fracción tumoral determinada según el método proporcionado en el presente documento.

60 El término “fracción tumoral” en el presente documento se refiere a la proporción de ADN tumoral libre de células circulante (ADN_{tc}) en relación con la cantidad total de ADN libre de células (ADN_{lc}). Se cree que la fracción tumoral indica el tamaño del tumor.

65 El término “ADN genómico” o “ADN” en el presente documento se refiere al ADN de un genoma celular. El ADN genómico puede ser celular, es decir, contenido dentro de una célula, o puede ser libre de células.

El término “muestra” en el presente documento se refiere a cualquier sustancia que contenga o se suponga que contiene ácido nucleico. La muestra puede ser una muestra biológica obtenida de un sujeto. Los ácidos nucleicos pueden ser ARN, ADN, por ejemplo, ADN genómico, ADN mitocondrial, ADN viral, ADN sintético, o ADN_{tc} sometido

a transcripción inversa a partir de ARN. Los ácidos nucleicos en una muestra de ácido nucleico sirven generalmente como moldes para la extensión de un cebador hibridado. En algunas realizaciones, la muestra biológica es una muestra de líquido biológico. La muestra de líquido puede ser sangre completa, plasma, suero, ascitis, líquido cefalorraquídeo, sudor, orina, lágrimas, saliva, muestra bucal, enjuague de cavidades, o enjuague de órganos. La muestra de líquido puede ser una muestra de líquido esencialmente libre de células (por ejemplo, plasma, suero, sudor, orina, lágrimas, etc.). En otras realizaciones, la muestra biológica es una muestra biológica sólida, por ejemplo, heces o biopsia de tejido, por ejemplo, una biopsia de tumor. Una muestra también puede comprender componentes de cultivo celular *in vitro* (incluyendo pero sin limitarse a, el medio condicionado resultante del crecimiento de células en medio de cultivo celular, células recombinantes y componentes celulares). En algunas realizaciones, la muestra es una muestra biológica que consiste en una mezcla de ácidos nucleicos de múltiples fuentes, es decir, hay más de un contribuyente a una muestra biológica, por ejemplo, dos o más individuos.

En el presente documento, el término “secuencia diana” se refiere a un polinucleótido diana seleccionado, por ejemplo, una secuencia presente en una molécula de ADNlc, cuya presencia, cantidad, y/o secuencia de nucleótidos, o cambios en estas, se desea determinar. Las secuencias diana se exploran para detectar la presencia o ausencia de una mutación somática. El polinucleótido diana puede ser una región de un gen asociada a una enfermedad. En algunas realizaciones, la región es un exón. La enfermedad puede ser cáncer.

Los términos “aparear”, “hibridar” o “unir” pueden referirse a dos secuencias, segmentos o hebras de polinucleótido, y pueden usarse indistintamente y tienen el significado habitual en la técnica. Dos secuencias complementarias (por ejemplo, ADN y/o ARN) pueden aparearse o hibridarse formando enlaces de hidrógeno con bases complementarias para producir un polinucleótido bicatenario o una región bicatenaria de un polinucleótido.

El término “marcador” o “marcador de segregación” se refiere a un resto que se usa para discriminar entre dos o más muestras, por ejemplo, dos o más individuos o tejidos. Un marcador puede ser un ácido nucleico (por ejemplo, un gen), una molécula pequeña, un péptido, un ácido graso, un metabolito, una proteína, un lípido, etc. Un marcador puede ser una mutación. Un marcador puede ser un ácido nucleico sintético. Un marcador o un conjunto de marcadores puede definir el motivo distintivo genético de una entidad, por ejemplo, un individuo, en relación con un segundo ácido nucleico, por ejemplo, una secuencia de ácido nucleico de referencia.

El término “mutación” en el presente documento se refiere a un cambio introducido en una secuencia de referencia, incluyendo, pero sin limitarse a, sustituciones, inserciones, deleciones (incluyendo truncamientos) con relación a la secuencia de referencia. Las mutaciones pueden implicar grandes secciones de ADN (por ejemplo, variación en el número de copias). Las mutaciones pueden implicar cromosomas completos (por ejemplo, aneuploidía). Las mutaciones pueden implicar pequeñas secciones de ADN. Los ejemplos de mutaciones que implican pequeñas secciones de ADN incluyen, por ejemplo, mutaciones puntuales o polimorfismos de un solo nucleótido (SNP), polimorfismos de múltiples nucleótidos, inserciones (por ejemplo, inserción de uno o más nucleótidos en un locus, pero menos que todo el locus), cambios de múltiples nucleótidos, deleciones (por ejemplo, delección de uno o más nucleótidos en un locus), e inversiones (por ejemplo, inversión de una secuencia de uno o más nucleótidos). Las consecuencias de una mutación incluyen, pero no se limitan a, la creación de un nuevo carácter, propiedad, función, fenotipo o rasgo que no se encuentra en la proteína codificada por la secuencia de referencia. En algunas realizaciones, la secuencia de referencia es una secuencia parental. En algunas realizaciones, la secuencia de referencia es un genoma humano de referencia, por ejemplo, h19. En algunas realizaciones, la secuencia de referencia deriva de una secuencia no cancerosa (o no tumoral). En algunas realizaciones, la mutación es hereditaria. En algunas realizaciones, la mutación es espontánea o *de novo*.

Los términos “tratar”, “tratamiento” y “que trata” se refieren a la reducción o mejora de la progresión, gravedad y/o duración de un trastorno proliferativo, por ejemplo, cáncer, o la mejora de un trastorno proliferativo resultante de la administración de una o más terapias.

Los términos “cáncer” y “tumor” se usan indistintamente en el presente documento. Estos términos se refieren a la presencia de células que tienen características típicas de las células cancerígenas, tales como descontrol de la proliferación, inmortalidad, potencial metastásico, tasa de proliferación y crecimiento rápida, y ciertos rasgos morfológicos característicos. Las células cancerosas suelen presentarse en forma de tumor, pero tales células pueden existir solas en un animal, o pueden ser células cancerosas no tumorígenas, tales como una célula de leucemia. Estos términos incluyen tumor sólido, tumor de tejidos blandos, o lesión metastásica. Tal como se usa en el presente documento, el término “cáncer” incluye cánceres tanto premalignos como malignos.

Tal como se usa en el presente documento, el término “código de barras” (también denominado identificador de una sola molécula [SMI]) se refiere a una secuencia de ácido nucleico conocida que permite identificar algún rasgo de un polinucleótido al que está asociado el código de barras. En algunas realizaciones, el rasgo del polinucleótido que va a identificarse es la muestra de la que deriva el polinucleótido. En algunas realizaciones, los códigos de barras tienen una longitud de aproximadamente o al menos aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, o más nucleótidos. En algunas realizaciones, los códigos de barras tienen una longitud menor de 10, 9, 8, 7, 6, 5, o 4 nucleótidos. En algunas realizaciones, los códigos de barras asociados a algunos polinucleótidos tienen una longitud diferente a la de los códigos de barras asociados a otros polinucleótidos. En general, los códigos de barras tienen

- una longitud suficiente e incluyen secuencias lo suficientemente diferentes como para permitir la identificación de muestras basándose en códigos de barras a los que están asociados. En algunas realizaciones, un código de barras, y la fuente de muestra a la que está asociado, pueden identificarse con precisión después de la mutación, inserción, o deleción de uno o más nucleótidos en la secuencia de código de barras, tal como la mutación, inserción, o deleción de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más nucleótidos. En algunas realizaciones, cada código de barras en una pluralidad de códigos de barras difiere de todos los demás códigos de barras en la pluralidad en al menos tres posiciones de nucleótido, tales como al menos 3, 4, 5, 6, 7, 8, 9, 10, o más posiciones de nucleótido. Una pluralidad de códigos de barras puede estar representada en una agrupación de muestras, incluyendo cada muestra polinucleótidos que comprenden uno o más códigos de barras que difieren de los códigos de barras contenidos en los polinucleótidos derivados de las demás muestras de la agrupación. Las muestras de polinucleótidos que incluyen uno o más códigos de barras pueden agruparse basándose en las secuencias de código de barras a las que están unidos, de tal manera que las cuatro bases de nucleótidos A, G, C, y T estén representadas de manera aproximadamente uniforme en una o más posiciones a lo largo de cada código de barras en la agrupación (tal como en 1, 2, 3, 4, 5, 6, 7, 8, o más posiciones, o en todas las posiciones del código de barras).
- El término “par de bases” o “pb” tal como se usa en el presente documento se refiere a una asociación (es decir, emparejamiento por enlaces de hidrógeno) de adenina (A) con timina (T), o de citosina (C) con guanina (G) en una molécula de ADN bicatenario. En algunas realizaciones, un par de bases puede incluir A emparejado con uracilo (U), por ejemplo, en un dúplex de ADN/ARN.
- Una “variante genética causal” es una variante genética para la cual existen evidencias estadísticas, biológicas y/o funcionales de asociación con una enfermedad o un rasgo.
- El término “variante del número de copias” o “CNV” se refiere a cualquier duplicación o deleción de un segmento genómico. Una “variante con pérdida del número de copias” o “CNLV” se refiere a una deleción de un segmento genómico de más de aproximadamente 100 pares de bases.
- El término “variante *indel*” se refiere a una variante de inserción o deleción.
- El término “microdeleción” se refiere a una deleción de aproximadamente 2 millones de pares de bases a aproximadamente 7 millones de pares de bases.
- El término “error aleatorio o sistemático” significa un artefacto de secuencia introducido artificialmente.
- El término “polimorfismo de nucleótido pequeño” o “SNP” se refiere a una variante de un solo nucleótido (SNV), una variante de múltiples nucleótidos (MNV) o una variante *indel* de aproximadamente 100 pares de bases o menos.
- En general, un “complemento” de una secuencia de ácido nucleico dada es una secuencia que es totalmente complementaria y puede hibridarse con la secuencia dada. En general, una primera secuencia que puede hibridarse con una segunda secuencia o un conjunto de segundas secuencias puede hibridarse específica o selectivamente con la segunda secuencia o un conjunto de segundas secuencias, de tal manera que se prefiere la hibridación con la segunda secuencia o un conjunto de segundas secuencias (por ejemplo, termodinámicamente más estables en un conjunto dado de condiciones, tales como las condiciones rigurosas usadas habitualmente en la técnica) en comparación con la hibridación con secuencias no diana durante una reacción de hibridación. Normalmente, las secuencias que pueden hibridarse comparten un grado de complementariedad de secuencia a lo largo de la totalidad o una parte de sus longitudes respectivas, tales como una complementariedad del 25 %-100 %, que incluye una complementariedad de secuencia de al menos aproximadamente el 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, y 100 %.
- El término “complementario” en el presente documento se refiere al concepto amplio de complementariedad de secuencias en regiones dúplex de una sola hebra de polinucleótido o entre dos hebras de polinucleótido entre pares de nucleótidos mediante emparejamiento de bases. Se sabe que un nucleótido de adenina puede formar enlaces de hidrógeno específicos (“emparejamiento de bases”) con un nucleótido, que es timina o uracilo. De igual manera, se sabe que un nucleótido de citosina puede dar emparejamiento de bases con un nucleótido de guanina. Sin embargo, en ciertas circunstancias, también pueden formarse enlaces de hidrógeno entre otros pares de bases, por ejemplo, entre adenina y citosina, etc. “Esencialmente complementario” en el presente documento se refiere a la complementariedad de secuencias en regiones dúplex de una sola hebra de polinucleótido o entre dos hebras de polinucleótido, por ejemplo, en las que la complementariedad es menor del 100 % pero mayor del 90 %, y conserva la estabilidad de la región dúplex.
- El término “derivado de” abarca los términos “originado a partir de”, “obtenido de”, “que puede obtenerse de”, “aislado de” y “creado a partir de”, e indica generalmente que un material específico encuentra su origen en otro material específico o tiene rasgos que pueden describirse con referencia al otro material específico.
- Un “gen” se refiere a un segmento de ADN implicado en la producción de un polipéptido e incluye regiones anteriores y posteriores a las regiones codificantes, así como secuencias intermedias (intrones) entre segmentos

codificantes individuales (exones).

La "hibridación" y el "apareamiento" se refieren a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza mediante enlaces de hidrógeno entre las bases de los residuos de nucleótido. Los enlaces de hidrógeno pueden producirse mediante emparejamiento de bases de Watson-Crick, unión de Hoogsteen o cualquier otra manera específica de secuencia. El complejo puede incluir dos hebras de ácido nucleico que forman una estructura de dúplex, tres, o más hebras que forman un complejo multicatenario, una sola hebra autohíbrida, o cualquier combinación de las mismas. Una reacción de hibridación puede constituir una etapa en un proceso más extenso, tal como la iniciación de la reacción en cadena de la polimerasa (PCR), reacción de ligamiento, reacción de secuenciación o reacción de escisión, por ejemplo, la escisión enzimática de un polinucleótido por una ribozima. Una primera secuencia de ácido nucleico que puede estabilizarse mediante enlaces de hidrógeno con las bases de los residuos de nucleótido de una segunda secuencia se dice que "puede hibridar" con la segunda secuencia. En tal caso, también puede decirse que la segunda secuencia puede hibridar con la primera secuencia. El término "hibridado" se refiere a un polinucleótido en un complejo estabilizado mediante enlaces de hidrógeno entre las bases de los residuos de nucleótido.

El término "biblioteca" en el presente documento se refiere a una colección o pluralidad de moléculas molde, es decir dúplex de ADN diana, que comparten secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'. El uso del término "biblioteca" para referirse a una colección o pluralidad de moléculas molde no debe tomarse como que implica que los moldes que componen la biblioteca deriven de una fuente particular ni que la "biblioteca" tenga una composición particular. A modo de ejemplo, el uso del término "biblioteca" no debe tomarse como que implica que los moldes individuales dentro de la biblioteca deban tener una secuencia de nucleótidos diferente ni que los moldes deban estar relacionados en cuanto a secuencia y/o fuente.

El término "secuenciación de próxima generación (NGS)" en el presente documento se refiere a métodos de secuenciación que permiten la secuenciación masiva en paralelo de moléculas de ácido nucleico amplificadas de manera clonal e individuales, durante la cual se secuencian simultáneamente una pluralidad, por ejemplo millones, de fragmentos de ácido nucleico procedentes de una única muestra o de múltiples muestras diferentes. Ejemplos no limitativos de NGS incluyen la secuenciación por síntesis, la secuenciación por ligamiento, la secuenciación en tiempo real, y la secuenciación por nanoporos.

El término "nucleótido" en el presente documento se refiere a una unidad monomérica de ADN o ARN que consiste en un resto de azúcar (pentosa), un fosfato, y una base heterocíclica nitrogenada. La base se une al resto de azúcar a través del carbono glicosídico (carbono 1' de la pentosa), y esa combinación de base y azúcar constituye un nucleósido. Cuando el nucleósido contiene un grupo fosfato unido a la posición 3' o 5' de la pentosa, se denomina nucleótido. Una secuencia de nucleótidos poliméricos operativamente unidos se denomina habitualmente en el presente documento "secuencia de bases" o "secuencia de nucleótidos", o "hebra" de ácido nucleico o polinucleótido, y se representa en el presente documento mediante una fórmula cuya orientación de izquierda a derecha es la dirección convencional del extremo 5' terminal al extremo terminal 3', haciendo referencia al grupo fosfato en 5' terminal y al grupo hidroxilo en 3' terminal en los extremos "5'" y "3'" de la secuencia polimérica, respectivamente.

El término "análogo de nucleótido" en el presente documento se refiere a análogos de nucleósidos trifosfato, por ejemplo, (S)-glicerol-nucleósidos trifosfato (gNTP) de las bases nitrogenadas comunes: adenina, citosina, guanina, uracilo, y timidina (Horhota *et al.*, *Organic Letters*, 8:5345-5347 [2006]) También se incluyen los nucleósidos tetrafosfato, los nucleósidos pentafofosfato y los nucleósidos hexafofosfato.

El término "operativamente unido" se refiere a una yuxtaposición o disposición de elementos especificados que les permite actuar en conjunto para producir un efecto. Por ejemplo, un promotor está operativamente unido a una secuencia codificante si controla la transcripción de la secuencia codificante.

El término "polimerasa" en el presente documento se refiere a una enzima que cataliza la polimerización de nucleótidos (es decir, la actividad polimerasa). El término polimerasa abarca las ADN polimerasas, las ARN polimerasas, y las transcriptasas inversas. Una "ADN polimerasa" cataliza la polimerización de desoxirribonucleótidos. Una "ARN polimerasa" cataliza la polimerización de ribonucleótidos. Una "transcriptasa inversa" cataliza la polimerización de desoxirribonucleótidos complementarios a un molde de ARN.

Los términos "polinucleótido", "secuencia de nucleótidos", "ácido nucleico", y "oligonucleótido" se usan indistintamente. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida, y pueden ser monocatenarios o multicatenarios (por ejemplo, monocatenarios, bicatenarios, de triple hélice, etc.), que contienen desoxirribonucleótidos, ribonucleótidos y/o análogos o formas modificadas de desoxirribonucleótidos o ribonucleótidos, incluyendo bases o nucleótidos modificados o sus análogos. Debido a que el código genético es redundante, puede usarse más de un codón para codificar un aminoácido en particular, y la presente divulgación abarca polinucleótidos que codifican para una secuencia de aminoácidos en particular. Puede usarse cualquier tipo

de nucleótido modificado o análogo de nucleótido, siempre que el polinucleótido conserve la funcionalidad deseada en las condiciones de uso, incluyendo modificaciones que aumentan la resistencia a las nucleasas (por ejemplo, desoxi, 2'-O-Me, fosforotioatos, etc.). También pueden incorporarse marcadores con propósitos de detección o captura, por ejemplo, marcadores o anclajes radiactivos o no radiactivos, por ejemplo, biotina. El término polinucleótido también incluye los ácidos nucleicos peptídicos (ANP). Los polinucleótidos pueden producirse de manera natural o no natural. Los polinucleótidos pueden contener ARN, ADN o ambos, y/o formas modificadas y/o análogos de los mismos. Una secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Uno o más enlaces fosfodiéster pueden sustituirse por grupos de unión alternativos. Estos grupos de unión alternativos incluyen, pero no se limitan a, realizaciones en las que el fosfato se sustituye por P(O)S ("tioato"), P(S)S ("ditioato"), (O)NR₂ ("amidato"), P(O)R, P(O)OR', CO o CH₂ ("formacetal"), en los que cada R o R' es independientemente H o alquilo (1-20 C) sustituido o no sustituido que contiene opcionalmente una unión éter (—O—), arilo, alquenilo, cicloalquilo, cicloalquenilo o araldilo. No todas las uniones en un polinucleótido necesitan porciones circulares. Los siguientes son ejemplos no limitativos de polinucleótidos: regiones codificantes o no codificantes de un gen o fragmento de gen, ADN intergénico, loci (locus) definidos a partir del análisis de ligamiento, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ARN de interferencia corto (ARNic), ARN en horquilla corto (ARNhc), microARN (miARN), ARN nucleolar pequeño, ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico, adaptadores, y cebadores. Un polinucleótido puede incluir nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótido. Si están presentes, las modificaciones en la estructura del nucleótido pueden conferirse antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Un polinucleótido puede modificarse adicionalmente después de la polimerización, tal como mediante la conjugación con un componente de marcaje, una etiqueta, un resto reactivo, o una pareja de unión. Las secuencias de polinucleótido, cuando se proporcionan, se anotan en dirección 5' a 3', a menos que se indique lo contrario.

Tal como se usa en el presente documento, "polipéptido" se refiere a una composición que se compone de aminoácidos y reconocida como proteína por los expertos en la técnica. En el presente documento, se usa el código convencional de una letra o de tres letras para los residuos de aminoácido. Los términos "polipéptido" y "proteína" se usan indistintamente en el presente documento para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede incluir aminoácidos modificados, y puede estar interrumpido por componentes distintos de aminoácidos. Los términos también abarcan un polímero de aminoácidos modificado de manera natural o mediante intervención; por ejemplo, formación de enlaces disulfuro, glicosilación, lipidación, acetilación, fosforilación, o cualquier otra manipulación o modificación, tal como conjugación con un componente de marcaje. También se incluyen en la definición, por ejemplo, polipéptidos que contienen uno o más análogos de un aminoácido (incluyendo, por ejemplo, aminoácidos no naturales, etc.), así como otras modificaciones conocidas en la técnica.

El término "cebador" en el presente documento se refiere a un oligonucleótido, ya sea que se produce de manera natural o sintética, que puede actuar como punto de iniciación de la síntesis de ácidos nucleicos cuando se pone en condiciones en las que se induce la síntesis de un producto de extensión de cebador que es complementario a una hebra de ácido nucleico, por ejemplo, en presencia de cuatro nucleótidos trifosfato diferentes y una enzima polimerasa, por ejemplo, una enzima termoestable, en un tampón apropiado ("tampón" incluye pH, fuerza iónica, cofactores, etc.) y a una temperatura adecuada. El cebador es preferiblemente monocatenario para una máxima eficiencia en la amplificación, pero alternativamente puede ser bicatenario. Si es bicatenario, el cebador se trata en primer lugar para separar sus hebras antes de usarse para preparar productos de extensión. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo suficientemente largo como para cebar la síntesis de productos de extensión en presencia de la polimerasa, por ejemplo, la enzima polimerasa termoestable. Las longitudes exactas de un cebador dependerán de muchos factores, incluyendo la temperatura, la fuente del cebador y el uso del método. Por ejemplo, dependiendo de la complejidad de la secuencia diana, el cebador de oligonucleótidos contiene normalmente 15-25 nucleótidos, aunque puede contener más o menos nucleótidos. Las moléculas de cebador cortas requieren generalmente temperaturas más frías para formar complejos híbridos suficientemente estables con el molde.

Un "promotor" se refiere a una secuencia reguladora implicada en la unión de la ARN polimerasa para iniciar la transcripción de un gen. Un promotor puede ser un promotor inducible o un promotor constitutivo. Un "promotor inducible" es un promotor que está activo en condiciones ambientales o reguladoras del desarrollo.

El término "biblioteca de secuenciación" en el presente documento se refiere al ADN que se procesa para secuenciación, por ejemplo, usando métodos masivos en paralelo, por ejemplo, NGS. El ADN puede amplificarse opcionalmente para obtener una población de múltiples copias de ADN procesado, que pueden secuenciarse mediante NGS.

El término "proyección monocatenaria" o "proyección" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bicatenario (bc) que se extiende más allá del extremo terminal de la hebra complementaria de la molécula de ácido nucleico bc. El término "proyección en 5'" o "secuencia de proyección en 5'" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bc que se extiende

en dirección 5' más allá del extremo terminal 3' de la hebra complementaria de la molécula de ácido nucleico bc. El término "proyección en 3'" o "secuencia de proyección en 3'" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bc que se extiende en dirección 3' más allá del extremo terminal 5' de la hebra complementaria de la molécula de ácido nucleico bc.

Un "espaciador" puede consistir en un solo nucleótido repetido (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más del mismo nucleótido seguidos), o en una secuencia de 2, 3, 4, 5, 6, 7, 8, 9, 10, o más nucleótidos repetidos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más veces. Un espaciador puede comprender o consistir en una secuencia específica, tal como una secuencia que no hibrida con ninguna secuencia diana en una muestra. Un espaciador puede comprender o consistir en una secuencia de nucleótidos seleccionados al azar.

Las frases "sustancialmente similares" y "sustancialmente idénticos" en el contexto de al menos dos ácidos nucleicos significan generalmente que un polinucleótido incluye una secuencia que tiene una identidad de secuencia de al menos aproximadamente el 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o incluso el 99,5 %, en comparación con un polinucleótido o polipéptido de referencia (por ejemplo, de tipo natural). La identidad de secuencia puede determinarse usando programas conocidos tales como BLAST, ALIGN, y CLUSTAL, usando parámetros convencionales. (Véanse, por ejemplo, Altshul *et al.* (1990) *J. Mol. Biol.* 215:403-410; Henikoff *et al.* (1989) *Proc. Natl. Acad. Sci.* 89:10915; Karin *et al.* (1993) *Proc. Natl. Acad. Sci.* 90:5873; y Higgins *et al.* (1988) *Gen* 73:237). El software para realizar análisis con BLAST está disponible públicamente a través del Centro Nacional de Información Biotecnológica. También pueden realizarse búsquedas en bases de datos usando FASTA (Person *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448.) En algunas realizaciones, moléculas de ácido nucleico sustancialmente idénticas se hibridan entre sí en condiciones rigurosas (por ejemplo, dentro de un intervalo de rigurosidad media a alta).

La "síntesis" de ácidos nucleicos en el presente documento se refiere a cualquier método *in vitro* para crear una nueva hebra de polinucleótido o alargar un polinucleótido existente (es decir, ADN o ARN) de manera dependiente de molde. La síntesis, según la divulgación, puede incluir amplificación, que aumenta el número de copias de una secuencia molde de polinucleótido mediante el uso de una polimerasa. La síntesis de polinucleótidos (por ejemplo, amplificación) da como resultado la incorporación de nucleótidos en un polinucleótido (por ejemplo, extensión a partir de un cebador), formando de ese modo una nueva molécula de polinucleótido complementaria al molde de polinucleótido. La molécula de polinucleótido formada y su molde pueden usarse como moldes para sintetizar moléculas de polinucleótido adicionales. "Síntesis de ADN", tal como se usa en el presente documento, incluye, pero no se limita a, la reacción en cadena de la polimerasa (PCR) y puede incluir el uso de nucleótidos marcados, por ejemplo, para sondas y cebadores de oligonucleótidos, o para la secuenciación de polinucleótidos.

El término "etiqueta" se refiere a un resto detectable que puede ser uno o más átomo(s) o molécula(s), o una colección de átomos y moléculas. Una etiqueta puede proporcionar un motivo distintivo óptico, electroquímico, magnético, o electrostático (por ejemplo, inductivo o capacitivo).

El término "nucleótido marcado con etiqueta" en el presente documento se refiere a un nucleótido que incluye una etiqueta (o especie de etiqueta) que se acopla en cualquier ubicación del nucleótido, incluyendo, pero sin limitarse a, un fosfato (por ejemplo, el fosfato terminal), un resto de azúcar o base nitrogenada del nucleótido. Las etiquetas pueden ser uno o más átomo(s) o molécula(s), o una colección de átomos y moléculas. Una etiqueta puede proporcionar un motivo distintivo óptico, electroquímico, magnético, o electrostático (por ejemplo, inductivo o capacitivo).

El término "dúplex de ADN diana" en el presente documento se refiere a una molécula de ADN bicatenario que deriva de un polinucleótido de muestra que es ADN, por ejemplo, ADN genómico o libre de células ("ADNlc"), y/o ARN.

Tal como se usa en el presente documento, el término "polinucleótido diana" se refiere a una molécula de ácido nucleico o polinucleótido de una población de moléculas de ácido nucleico que tiene una secuencia diana con la que está diseñado que se hibriden uno o más oligonucleótidos. En algunas realizaciones, una secuencia diana identifica de manera única una secuencia derivada de una muestra, tal como una secuencia genómica, mitocondrial, bacteriana, viral, o de ARN (por ejemplo, ARNm, miARN, miARN primario, o pre-miARN) particular. En algunas realizaciones, una secuencia diana es una secuencia común compartida por múltiples polinucleótidos diana diferentes, tales como una secuencia adaptadora común unida a diferentes polinucleótidos diana. "Polinucleótido diana" puede usarse para referirse a una molécula de ácido nucleico bicatenario que incluye una secuencia diana en una o ambas hebras, o a una molécula de ácido nucleico monocatenario que incluye una secuencia diana, y puede derivar de cualquier fuente o proceso para aislar o generar moléculas de ácido nucleico. Un polinucleótido diana puede incluir una o más secuencias diana (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más), que pueden ser iguales o diferentes. En general, los diferentes polinucleótidos diana incluyen secuencias diferentes, tales como uno o más nucleótidos diferentes o una o más secuencias diana diferentes.

El término "molécula de ADN molde" en el presente documento se refiere a una hebra de un ácido nucleico a partir de la cual se sintetiza una hebra de ácido nucleico complementaria mediante una ADN polimerasa, por ejemplo, en

una reacción de extensión de cebador.

El término "de manera dependiente de molde" se refiere a un proceso que implica la extensión dependiente de molde de una molécula cebadora (por ejemplo, la síntesis de ADN por la ADN polimerasa). El término "de manera dependiente de molde" se refiere normalmente a la síntesis de polinucleótidos de ARN o ADN, en la que la secuencia de la hebra de polinucleótido recién sintetizada está dictada por las reglas bien conocidas de emparejamiento de bases complementarias (véanse, por ejemplo, Watson, J. D. *et al.*, En: *Molecular Biology of the Gene*, 4ª ed., W. A. Benjamin, Inc., Menlo Park, California (1987)).

Una "muestra" puede incluir, pero no se limita a, sangre, plasma, saliva, orina, semen, líquido amniótico, ovocitos, piel, cabello, heces, hisopos bucales o lisado de Papanicolaou de un individuo.

Una "gradilla para tubos" se refiere a un soporte para tubos con una pluralidad de ranuras para sostener los tubos de muestra. Normalmente, la gradilla para tubos está configurada para sostener los tubos de muestra en posición vertical.

"Información codificada" o "código de identificación" se refiere a la información que puede recuperarse para identificar una muestra, la fuente de una muestra y/o información sobre una muestra (por ejemplo, un paciente del que se obtuvo una muestra, una fuente de tejido, etc.). La información codificada puede tener, por ejemplo, la forma de un código de barras unidimensional, bidimensional, o tridimensional.

Una "porción adyacente a una región de interés" se refiere a una secuencia que es inmediatamente proximal a una región de interés. La referencia a una "porción de o adyacente a una región de interés" se refiere a una secuencia que 1) se encuentra completamente dentro de la región de interés, 2) se encuentra completamente fuera de, pero inmediatamente proximal a, la región de interés, o 3) incluye una secuencia contigua desde dentro de, e inmediatamente proximal a, la región de interés. La referencia a una "secuencia que es sustancialmente complementaria a una porción de o adyacente a una región de interés" se refiere a 1) una secuencia que es sustancialmente complementaria a una secuencia completamente dentro de la región de interés, 2) una secuencia sustancialmente complementaria a una secuencia completamente fuera de, pero inmediatamente proximal a, la región de interés, o 3) una secuencia que es sustancialmente complementaria a una secuencia contigua desde dentro de, e inmediatamente proximal a, la región de interés.

El término "promedio" tal como se usa en el presente documento se refiere a una media o una mediana, o cualquier valor usado para aproximar la media o la mediana, a menos que el contexto indique claramente lo contrario.

Se entiende que los aspectos y variaciones de la invención descrita en el presente documento incluyen aspectos y variaciones que "consisten" y/o "consisten esencialmente en".

El término "sustancialmente complementarias" se usa para referirse a dos secuencias de ácidos nucleicos (X e Y) en hebras opuestas, ambas con una longitud de al menos 12 bases y la fracción de complementariedad entre ellas es de al menos 0,75. La fracción de complementariedad se calcula de la siguiente manera: en primer lugar, se calcula la alineación óptima entre X y el complemento inverso de Y mediante el algoritmo de Needleman-Wunsch (Needleman *et al.*, A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, vol. 48 (3), págs. 443-453 (1970)) usando parámetros por defecto (es decir, coincidencia = +1, coincidencia errónea = -1 y hueco = -1). Luego, se contabiliza el número de coincidencias para la alineación óptima. Finalmente, la fracción de complementariedad se define como el número de coincidencias dividido entre la menor longitud de cualquiera de las secuencias, es decir, la fracción de la longitud que es complementaria. El término "sustancialmente complementaria" incluye hebras de ácidos nucleicos completamente complementarias.

Un "mosaico" se refiere a uno o más loci contiguos dentro de una región de interés. Una región de interés puede dividirse en uno o más mosaicos. Los mosaicos pueden ser contiguos, aunque no necesariamente. Por tanto, la región de interés puede incluir opcionalmente subregiones no contiguas. Los mosaicos pueden tener la misma longitud o diferentes longitudes. Un "locus" se refiere a una o más bases contiguas, y está contenido completamente dentro del mosaico.

Preparación de bibliotecas de ADN (con etiqueta de secuencia opcional o integración de código de barras molecular)

En algunas realizaciones de los métodos divulgados en el presente documento, se obtiene o prepara una biblioteca de ADN a partir de ADN libre de células obtenido de un paciente con cáncer. La biblioteca de ADN comprende una población de moléculas de ADN. Las moléculas de ADN generalmente tienen la longitud suficiente para que puedan analizarse con precisión mediante secuenciación o PCR digital. Véanse, por ejemplo, las solicitudes de patente estadounidense n.ºs 2013/0225452 y 2012/0208705.

Al producir una biblioteca de ADN a partir de ADN genómico, puede fragmentarse el ADN genómico, por ejemplo, mediante el uso de cizallamiento hidrodinámico u otra fuerza mecánica, o fragmentarse mediante digestión química

o enzimática, tal como la digestión por restricción. Este proceso de fragmentación permite que las moléculas de ADN presentes en el genoma sean lo suficientemente cortas para su análisis, tal como secuenciación o PCR digital. Sin embargo, en general, el ADN libre de células suele ser lo suficientemente corto como para no requerir fragmentación. El ADN libre de células se origina a partir de ADN genómico. Una porción del ADN libre de células obtenido de una muestra de plasma de una madre embarazada se origina a partir del genoma materno y una porción del ADN libre de células se origina a partir del genoma fetal.

En algunas realizaciones, las moléculas de ADN se someten a modificación adicional, que da como resultado la unión de oligonucleótidos a las moléculas de ADN. Los oligonucleótidos pueden comprender una secuencia adaptadora o un código de barras molecular (o ambos). En algunas realizaciones, la secuencia adaptadora es común para todos los oligonucleótidos en una pluralidad de oligonucleótidos que se usan para formar la biblioteca de ADN. En algunas realizaciones, los códigos de barras moleculares son únicos o tienen baja redundancia. A modo de ejemplo, el oligonucleótido puede unirse a las moléculas de ADN mediante ligamiento. Puede usarse la unión directa de los oligonucleótidos a las moléculas de ADN en la biblioteca de ADN, por ejemplo, cuando el enriquecimiento se produce en un proceso posterior. Por ejemplo, en algunas realizaciones, se prepara una biblioteca de ADN mediante la unión directa de un oligonucleótido que comprende un código de barras molecular y una secuencia adaptadora, seguido por el enriquecimiento (por ejemplo, mediante hibridación) de moléculas de ADN que comprenden una región de interés o una porción de una región de interés.

En algunas realizaciones, la preparación y el enriquecimiento de la biblioteca se producen simultáneamente. Por ejemplo, en algunas realizaciones, se amplifican preferentemente las moléculas de ADN que comprenden una región de interés o una porción de la misma. Esto puede lograrse, por ejemplo, combinando el ADN libre de células (o ADN genómico) con oligonucleótidos que comprenden una secuencia específica de diana, una secuencia adaptadora, y un código de barras molecular, y amplificando las moléculas de ADN. Tal como se mencionó anteriormente, en algunas realizaciones, la secuencia adaptadora es común para todos los oligonucleótidos en una pluralidad de oligonucleótidos, y el código de barras molecular es único o tiene una baja redundancia. La secuencia específica de diana es única para la región de interés seleccionada como diana o una porción de la misma. Por tanto, la amplificación por PCR amplifica selectivamente las moléculas de ADN que comprenden la región de interés o una porción de la misma.

Cuando los métodos incluyen el uso de etiquetas o códigos de barras moleculares, la etiqueta o el código de barras molecular también pueden ligarse a los fragmentos o incluirse dentro de las secuencias adaptadoras ligadas. La unión independiente de la etiqueta o el código de barras molecular, a diferencia de la incorporación de la etiqueta o el código de barras molecular, puede variar según el método de enriquecimiento. Por ejemplo, al usar el enriquecimiento de diana basado en captura híbrida, el adaptador puede incluir el código de barras molecular, al usar el enriquecimiento dirigido por PCR, se usan proyecciones y pares de cebadores específicos de diana que incorporarán los adaptadores de secuenciación y códigos de barras moleculares y específicos de muestra; y al usar el enriquecimiento en secuenciador, el adaptador puede ligarse por separado de la etiqueta o el código de barras molecular.

Enriquecimiento dirigido de una región de interés (o porción de la misma)

La divulgación contempla métodos para enriquecer una secuencia diana en una región de interés. Las técnicas de enriquecimiento se conocen en la técnica. Véanse, por ejemplo, el documento WO2013/112923; Mertes *et al.*, Targeted enrichment of genomic DNA regions for next-generation sequencing, Briefings in Functional Genomics, vol. 10(6), págs. 374-386 (2011)). Las técnicas de enriquecimiento a modo de ejemplo incluyen, pero no se limitan a, captura híbrida, circularización selectiva (también denominada sondas de inversión molecular (MIP)) y amplificación por PCR de regiones de interés seleccionadas como diana. Los métodos de captura híbrida se basan en la hibridación selectiva de las regiones genómicas diana con oligonucleótidos diseñados por el usuario. La hibridación puede ser con oligonucleótidos inmovilizados en microalineamientos de alta o baja densidad (captura en alineamiento), o hibridación en fase de disolución con oligonucleótidos modificados con un ligando (por ejemplo, biotina) que pueden inmovilizarse posteriormente en una superficie sólida, tal como una perla (captura en disolución). El método basado en sondas de inversión molecular (MIP) se basa en la construcción de numerosas sondas de oligonucleótidos lineales monocatenarios, que consisten en un ligador común flanqueado por secuencias específicas de diana. Tras el apareamiento con una secuencia diana, la región con huecos de la sonda se rellena mediante polimerización y ligamiento, lo que da como resultado una sonda circularizada. Entonces, las sondas circularizadas se liberan y amplifican usando cebadores dirigidos a la región de ligador común. Los métodos basados en PCR emplean una amplificación por PCR altamente paralela, donde cada secuencia diana de la muestra tiene un par correspondiente de cebadores únicos y específicos de secuencia. En algunas realizaciones, el enriquecimiento de una secuencia diana se produce durante la secuenciación.

Secuenciación

La divulgación contempla métodos de secuenciación de la biblioteca de secuencias. La secuenciación puede realizarse mediante cualquier método conocido en la técnica. Los métodos de secuenciación incluyen, pero no se limitan a, técnicas basadas en la secuenciación de Maxam-Gilbert, técnicas basadas en la terminación de cadena,

secuenciación al azar ("*shotgun*"), secuenciación por PCR puente, secuenciación de una sola molécula en tiempo real, secuenciación por semiconductores iónicos (secuenciación "*Ion Torrent*"), secuenciación por nanoporos, pirosecuenciación (454), secuenciación por síntesis, secuenciación por ligamiento (secuenciación SOLiD), secuenciación por microscopía electrónica, reacciones de secuenciación didesoxi (método de Sanger),
 5 secuenciación masiva en paralelo, secuenciación Polony, y secuenciación por nanobolas de ADN. En algunas realizaciones, la secuenciación implica hibridar un cebador con el molde para formar un dúplex molde/cebador, poner en contacto el dúplex con una enzima polimerasa en presencia de nucleótidos marcados de manera detectable en condiciones que permitan que la polimerasa añada nucleótidos al cebador de manera dependiente de molde, detectar una señal del nucleótido marcado incorporado, y repetir secuencialmente las etapas de poner en
 10 contacto y detectar al menos una vez, en las que la detección secuencial del nucleótido marcado incorporado determina la secuencia del ácido nucleico. En algunas realizaciones, la secuenciación comprende la obtención de lecturas de extremos emparejados. La precisión o precisión promedio de la información de secuencia puede ser mayor del 80 %, 90 %, 95 %, 99 % o 99,98 %. En algunas realizaciones, la información de secuencia obtenida es mayor de 50 pb, 100 pb o 200 pb. La información de la secuencia puede obtenerse en menos de 1 mes, 2 semanas,
 15 1 semana 1 día, 3 horas, 1 hora, 30 minutos, 10 minutos o 5 minutos. La precisión de secuencia, o la precisión promedio, puede ser mayor del 95 % o el 99 %. La cobertura de secuencia puede ser mayor de 20 veces o menor de 500 veces. Los marcadores detectables a modo de ejemplo incluyen radiomarcadores, marcadores fluorescentes, marcadores enzimáticos, etc. En algunas realizaciones, el marcador detectable puede ser un marcador detectable ópticamente, tal como un marcador fluorescente. Los marcadores fluorescentes a modo de ejemplo incluyen cianina, rodamina, fluoresceína, cumarina, BODIPY, Alexa, o colorantes múltiples conjugados. En algunas realizaciones, el nucleótido se señala si uno o más de sus segmentos de secuencia son sustancialmente similares a uno o más segmentos de secuencia de otro nucleótido dentro de la misma partición.

También se contempla que algunos métodos de secuenciación de la biblioteca de secuencias no impliquen una
 25 etapa previa de enriquecimiento de la diana. Por ejemplo, el uso del enriquecimiento en secuenciador, tal como con un secuenciador de nanoporos, permite el enriquecimiento y la secuenciación "simultáneos" de la biblioteca de secuencias mediante el rechazo en tiempo real de moléculas que no pertenecen a la región de interés. Alternativamente, las secuencias pueden secuenciarse selectiva y preferentemente a partir de la región de interés.

En algunas realizaciones, el método utiliza un método de secuenciación dúplex. Este método se describe, por
 30 ejemplo, en la solicitud provisional estadounidense, en tramitación junto con la presente, 62/452,848, presentada el 31 de enero de 2017, titulada "Methods and Compositions for Enrichment of Target Polynucleotides" y puede usar códigos de barras moleculares tal como se describe a continuación.

35 Códigos de barras moleculares

En algunas realizaciones, se usa una secuencia de identificador, es decir, un código de barras molecular para
 40 identificar moléculas de ADN únicas en una biblioteca de ADN. Véanse, por ejemplo, las solicitudes de patente estadounidense n.ºs 2013/0261019 y 2015/0080266. Véase también las solicitudes provisionales estadounidenses, en tramitación junto con la presente, 62/348,791 presentada el 10 de junio de 2016, 62/364,256 presentada el 19 de julio de 2016, y 62/447,784 presentada el 18 de enero de 2017, todas ellas tituladas "Nucleic Acid Adapters and Uses Thereof". Los códigos de barras moleculares ayudan en la reconstrucción de secuencias de ADN contiguas o
 45 asisten en la determinación de la variación del número de copias. Los marcadores a modo de ejemplo incluyen proteínas de unión a ácidos nucleicos, marcadores ópticos, análogos de nucleótido, secuencias de ácidos nucleicos, y otros conocidos en la técnica.

En algunas realizaciones, el código de barras molecular es un código de barras nanoestructurado. En algunas
 50 realizaciones, el código de barras molecular comprende una secuencia de ácido nucleico que, al unirse a un polinucleótido diana, sirve como identificador de la muestra o secuencia de la que derivó el polinucleótido diana. En algunas realizaciones, los códigos de barras moleculares tienen una longitud de al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, o más nucleótidos. En algunas realizaciones, los códigos de barras moleculares tienen una longitud menor de 10, 9, 8, 7, 6, 5, o 4 nucleótidos. En algunas realizaciones, cada código de barras molecular en una pluralidad de códigos de barras moleculares difiere de todos los demás códigos de barras moleculares en la pluralidad en al menos tres posiciones de nucleótido, tales como al menos 3, 4, 5, 6, 7, 8, 9, 10, o más posiciones.
 55 En algunas realizaciones, los códigos de barras moleculares asociados a algunos polinucleótidos tienen una longitud diferente a la de los códigos de barras moleculares asociados a otros polinucleótidos. En general, los códigos de barras moleculares tienen una longitud suficiente y comprenden secuencias lo suficientemente diferentes como para permitir la identificación de muestras basándose en los códigos de barras moleculares a los que están asociados. En algunas realizaciones, tanto el adaptador directo como el inverso comprenden al menos una de una pluralidad de secuencias de código de barras molecular. En algunas realizaciones, cada adaptador inverso comprende al menos una de una pluralidad de secuencias de código de barras molecular, en el que cada secuencia de código de barras molecular difiere de todas las demás secuencias de código de barras molecular en la pluralidad de secuencias de código de barras molecular.

65 Se han notificado en la bibliografía diversos conjuntos de códigos de barras moleculares. Varios investigadores han usado conjuntos que cumplen las condiciones impuestas por un código de Hamming (Hamady *et al.*, Error-correcting

barcoded primers allow hundreds of samples to be pyrosequenced in multiplex, *Nature Methods*, vol. 5(3), págs. 235-237 (2008)); y Lefrançois *et al.*, Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing, *BMC Genomics*, vol. 10, págs. 1-18 (2009)). Otros autores han usado conjuntos que satisfacen condiciones más complejas que un código de Hamming pero comparten la garantía similar de una cierta distancia de Hamming mínima por pares (Fierer *et al.*, The influence of sex, handedness, and washing on the diversity of hand surface bacteria, *Proc. Nat'l Acad. Sci.*, vol. 105, págs. 17.994-17.999 (2008)); Krishnan *et al.*, Barcodes for DNA sequencing with guaranteed error correction and capability, *Electronics Letters*, vol. 47, págs. 236-237 (2011). Como alternativa a los códigos de barras moleculares basados en la distancia de Hamming, otros autores han seleccionado conjuntos de códigos de barras moleculares que satisfacen una distancia mínima de edición por pares. Los conjuntos de tales códigos de barras moleculares pueden funcionar con errores de inserción, delección o sustitución en la lectura de una secuencia de código de barras.

En algunas realizaciones, cada código de barras molecular de un conjunto es único, es decir, dos códigos de barras moleculares cualesquiera seleccionados de un conjunto dado diferirán en al menos una posición de nucleótido. Además, se contempla que los códigos de barras moleculares tengan ciertas propiedades bioquímicas que se seleccionan basándose en cómo se usará el conjunto. Por ejemplo, ciertos conjuntos de códigos de barras moleculares que se usan en una reacción de RT-PCR no deben tener secuencias complementarias a ninguna secuencia del genoma de un cierto organismo o conjunto de organismos. El requisito de no complementariedad ayuda a garantizar que el uso de una secuencia de código de barras molecular particular no dé como resultado un cebado incorrecto durante las manipulaciones de biología molecular que requieren los cebadores, tales como transcripción inversa o PCR. Ciertos conjuntos satisfacen otras propiedades bioquímicas impuestas por los requisitos asociados al procesamiento de las moléculas de secuencia en las que se incorporan los códigos de barras.

Los ejemplos de tecnologías de secuenciación para secuenciar códigos de barras moleculares, así como cualquier secuencia generada basada en nucleótidos, incluyen, pero no se limitan a, técnicas basadas en secuenciación de Maxam-Gilbert, técnicas basadas en terminación de cadena, secuenciación al azar, secuenciación por PCR puente, secuenciación de una sola molécula en tiempo real, secuenciación por semiconductores iónicos (secuenciación "Ion Torrent"), secuenciación por nanoporos, pirosecuenciación (454), secuenciación por síntesis, secuenciación por ligamiento (secuenciación SOLiD), secuenciación por microscopía electrónica, reacciones de secuenciación didesoxi (método de Sanger), secuenciación masiva en paralelo, secuenciación Polony, y secuenciación por nanobolas de ADN.

En algunas realizaciones, se usan códigos de barras moleculares para mejorar la potencia de los algoritmos de identificación del número de copias al reducir la no independencia de la duplicación por PCR. En otra realización, pueden usarse códigos de barras moleculares para mejorar la especificidad de prueba al reducir los errores de secuencia generados durante la amplificación.

Pruebas de enfermedades

Unos aspectos de la invención se refieren a métodos que son útiles para mejorar la detección, la monitorización y el tratamiento de un paciente que presenta una enfermedad. La enfermedad puede ser un cáncer. Puede sospecharse o saberse que el paciente presenta un tumor sólido, o puede ser un sujeto que previamente presentó un tumor sólido. En algunos aspectos, el tumor sólido es un tumor de un tejido u órgano. En otros aspectos, el tumor sólido es una masa metastásica de un cáncer de origen sanguíneo. El presente método también puede aplicarse a la detección y/o monitorización de cánceres de origen sanguíneo.

El genotipado del tejido tumoral en busca de alteraciones genéticas somáticas para obtener información procesable se ha convertido en una práctica habitual en la oncología clínica. Sin embargo, el tejido tumoral es una instantánea única en el tiempo, está sujeto a sesgos de selección resultantes de la heterogeneidad tumoral, y puede ser difícil de obtener. Además, en los casos en los que se extirpa el tumor, el tejido tumoral sólo está disponible una vez que el tumor ha reaparecido y ha avanzado lo suficiente como para ser detectado como una masa. Los fragmentos libres de células de ADN se liberan al torrente sanguíneo por células que experimentan apoptosis o necrosis, y la carga de ADN libre de células (ADNlc) circulante se correlaciona con la estadificación y el pronóstico del tumor. La capacidad de detectar y cuantificar mutaciones tumorales en ADNlc ha demostrado ser eficaz para el seguimiento de la dinámica tumoral en tiempo real, además de servir como biopsia líquida que puede usarse para una variedad de aplicaciones clínicas y de investigación que antes no eran posibles. Sin embargo, los métodos actuales están limitados por la cantidad de sangre que puede extraerse para el análisis y por las proporciones extremadamente bajas de ADNlc tumoral, de aproximadamente $1e^{-4}$. El método proporcionado en el presente documento combina el análisis de múltiples sitios somáticos específicos del paciente, por ejemplo, polimorfismos de un solo nucleótido (SNP), lo que permite la detección de mutaciones somáticas asociadas al cáncer del paciente en proporciones extremadamente bajas de ADNlc tumoral de menos de aproximadamente $1e^{-3}$.

En un aspecto, se proporciona un método para determinar la fracción tumoral de una muestra biológica de un paciente que presenta una enfermedad, por ejemplo, cáncer. La visión general del flujo de trabajo del método se proporciona en la figura 1. El método puede representarse en dos fases. En una primera fase, o fase de

reclutamiento, se identifican mutaciones somáticas específicas del paciente, y se crea un panel distintivo de sondas de captura, que son representativas de las mutaciones somáticas identificadas. En una segunda fase, se realiza la monitorización del estado del cáncer del paciente usando el panel de sondas de captura del paciente para identificar mutaciones somáticas que circulan como ADN libre de células. La segunda fase es no invasiva y requiere cantidades clínicamente viables de un líquido biológico, por ejemplo, una extracción de sangre periférica de 10-20 ml, que puede repetirse con la frecuencia deseada para detectar cambios en el cáncer del paciente. Una cantidad clínicamente viable de líquido biológico, por ejemplo, sangre completa, suele comprender al menos 1000 equivalentes de genoma, al menos 2000 equivalentes de genoma, al menos 3000 equivalentes de genoma, al menos 4000 equivalentes de genoma, al menos 5000 equivalentes de genoma, al menos 6000 equivalentes de genoma, al menos 7000 equivalentes de genoma, al menos 8000 equivalentes de genoma, al menos 9000 equivalentes de genoma, al menos 10.000 equivalentes de genoma, al menos 11.000 equivalentes de genoma, al menos 12.000 equivalentes de genoma o al menos 15.000 equivalentes de genoma. En algunas realizaciones, la segunda fase del método utiliza una muestra de sangre completa de entre 5 ml y 20 ml, que comprende entre 3000 y 15000 equivalentes de genoma.

En primer lugar, se identifica un panel de secuencias que comprenden mutaciones somáticas específicas del tumor de un paciente, de la siguiente manera. Se aísla el ADN genómico del tumor y de tejido normal, es decir, tejido no canceroso, usando cualquier método conocido en la técnica, y se secuencian. Se comparan las secuencias de ADN de las muestras tumoral y no tumoral, y se identifica un conjunto de mutaciones somáticas específicas del tumor del paciente. El conjunto de mutaciones somáticas identificadas sirve como un panel distintivo del paciente que puede secuenciarse en diversos estadios de la enfermedad, es decir, el panel distintivo puede cribarse para determinar la presencia de cáncer en la cirugía tras el diagnóstico; durante el tratamiento del cáncer, por ejemplo, a intervalos durante la quimioterapia o la radioterapia, para monitorizar la eficacia del tratamiento; a intervalos durante la remisión para confirmar la ausencia continuada de la enfermedad; y/o para detectar la recidiva de la enfermedad.

A continuación, se obtiene un conjunto de sondas de captura. Este conjunto de sondas de captura comprende secuencias que pueden hibridarse con secuencias diana específicas del genoma del paciente y que abarcan los sitios que comprenden las mutaciones somáticas específicas de tumor identificadas en el tejido tumoral. En algunas realizaciones, el conjunto de sondas de captura se calibra tal como se indica en la solicitud provisional estadounidense, en tramitación junto con la presente, n.º 62447816, titulada "Balanced Capture Probes and Methods of Use Thereof".

Posteriormente, se determina la fracción tumoral en una muestra de líquido del mismo paciente. La determinación de la fracción tumoral comprende la obtención de ADNlc del paciente y el uso de las sondas de captura diseñadas para el panel de marcadores distintivo específico del paciente (por ejemplo, mutaciones), la captura de secuencias diana de ADNlc que comprenden secuencias tumorales y secuencias normales correspondientes. Las secuencias capturadas se analizan y se enumeran, y se determina la fracción tumoral como la proporción de secuencias que comprenden una mutación somática con respecto al número total de secuencias alélicas mutadas y no mutadas correspondientes. La enumeración de secuencias alélicas mutadas y no mutadas se logra mediante el análisis de las lecturas de secuencia contabilizables obtenidas a partir del proceso de secuenciación. El método no requiere la detección de todas las mutaciones somáticas en el panel distintivo del paciente.

Automatización (LIMS)

En algunas realizaciones, los ensayos tal como se describe en el presente documento están integrados con un sistema de gestión de información de laboratorio (LIMS), tal como se describe a continuación.

Un LIMS, también denominado sistema de gestión de laboratorio (LMS) o sistema de información de laboratorio (LIS), es un sistema para modernizar las funciones de un laboratorio que tradicionalmente se realizaban de manera manual o semimanual. Un sistema LIMS puede incluir, pero no se limita a, un servidor u ordenador principal, una base de datos, un software de gestión, y puede estar acoplado a instrumentación de laboratorio correspondiente para realizar funciones de laboratorio respectivas. Un sistema LIMS asistirá generalmente al personal de laboratorio en el seguimiento, análisis, la clasificación, y encaminamiento de muestras de laboratorio a lo largo de procesos de laboratorio complejos, de manera eficiente y rentable.

Las ventajas de los sistemas LIMS incluyen, pero no se limitan a, mejor gestión de muestras, control de calidad, cadena de custodia, y generación de informes. Un sistema LIMS también permite un control flexible del acceso a la información de laboratorio entre un conjunto diverso de usuarios, tales como médicos, pacientes, analistas, y técnicos.

Un LIMS, tal como se divulga en el presente documento, permite la automatización y la gestión de información de laboratorio, y puede realizarse como un sistema, método, o producto de programa informático. Además, la presente divulgación puede adoptar la forma de una realización completamente de software, una realización completamente de hardware, o una combinación de realizaciones de software y hardware. Incluso más, la presente divulgación puede adoptar la forma de un producto de programa informático contenido en un medio de almacenamiento legible por ordenador, donde el código legible por ordenador está incorporado en el medio de almacenamiento. En otro

aspecto, la presente divulgación puede adoptar la forma de software informático implementado como servicio (SaaS). Puede utilizarse cualquier medio de almacenamiento adecuado, tal como almacenamiento óptico, almacenamiento magnético, discos duros, o CD-ROM.

5 La figura 8 ilustra una visión general del sistema 100 para un laboratorio de diagnóstico automático y sistema de gestión de información de laboratorio (en adelante, "LIMS"). El sistema 100 incluye un sistema 101 de gestión de datos, un sistema 102 de automatización, y un sistema 103 de gestión de resultados analíticos (ARMS). En general, el sistema 101 de gestión de datos es una herramienta de base de datos centralizada para el mantenimiento de la información relativa al sistema LIMS, tal como el mantenimiento de pruebas de laboratorio, diagnósticos, equipos, personal, y similares. En una realización, el sistema 101 de gestión de datos se actualiza de manera dinámica y facilita la gestión de información entre otros componentes del sistema LIMS, tales como el sistema 102 de automatización y ARMS 103.

15 El sistema 102 de automatización generalmente permite la gestión del flujo de trabajo del laboratorio y puede permitir que uno o más usuarios creen e implementen procesos de flujo de trabajo del laboratorio personalizados. Por ejemplo, un sistema 102 de automatización puede proporcionar funcionalidad para permitir al usuario crear un diagrama gráfico para modelar diferentes equipos de laboratorio y diagnósticos, y puede permitir al usuario personalizar la sincronización, la toma de decisiones, y otras variables de prueba del análisis de laboratorio. El sistema 102 de automatización puede proporcionar además funcionalidad que permita al usuario implementar uno o más procesos de flujo de trabajo basados en diagramas generados por el usuario, y tales procesos de flujo de trabajo pueden ser modificados de manera dinámica por el usuario. Además, el sistema 102 de automatización puede incluir componentes de hardware y software para interconectarse con equipos de laboratorio, tales como unidades robóticas, sistemas de transportador, depósitos de muestras, sistemas de control climático (por ejemplo, iluminación y temperatura), sistemas neumáticos, sistemas de audio/vídeo, etc.

25 En una realización, el sistema 102 de automatización puede incluir hardware y/o software para permitir que una o más unidades robóticas realicen movimientos relacionados con pruebas de muestras de laboratorio, tales como mezclado, agitación, calentamiento, enfriamiento, recogida y/o colocación de muestras. Por ejemplo, el sistema 102 de automatización puede generar y enviar comandos a la una o más unidades robóticas para permitir que se muevan las unidades robóticas en un espacio tridimensional. Tales comandos también pueden permitir que la una o más unidades robóticas se interconecten con un sistema neumático para utilizar aire a presión para sujetar y liberar una o más muestras. En una realización, las muestras pueden estar contenidas en un tubo de ensayo, un vial, o un recipiente similar. El sistema 102 de automatización puede configurarse además para generar y enviar comandos a la una o más unidades robóticas para permitir que las unidades robóticas retiren y/o vuelvan a colocar la tapa en la parte superior de un recipiente. Por ejemplo, la una o más unidades robóticas pueden estar equipadas con maquinaria que puede detectar la tapa de un tubo de ensayo y que puede además retirar la tapa de un tubo de ensayo mediante uno o más movimientos robóticos. De igual manera, la una o más unidades robóticas pueden estar equipadas con maquinaria para detectar un tubo de ensayo sin tapa y realizar uno o más movimientos robóticos para colocar y sellar el tubo de ensayo con tapa, por ejemplo.

40 El ARMS 103 proporciona generalmente un sistema para la representación y organización de manera dinámica de la información de laboratorio, incluyendo pero sin limitarse a, información tal como resultados de diagnóstico, métricas de control de calidad, datos de prueba históricos, genotipos de muestras, y similares. Por ejemplo, el ARMS 103 puede facilitar la generación de visualizaciones de datos interactivas para permitir que uno o más usuarios supervisen eficazmente la química, los algoritmos, y los productos del laboratorio. El ARMS 103 también puede permitir que uno o más usuarios realicen funciones analíticas complejas, tales como analizar y manipular restricciones de control de calidad, sintetizar datos de prueba sin procesar, y corregir manualmente los resultados de prueba.

50 En una realización, uno o más componentes del sistema 101 de gestión de datos, el sistema 102 de automatización, y/o el ARMS 103 pueden mantenerse en una ubicación cercana al laboratorio y equipos asociados (por ejemplo, una sala de servidores). En otra realización, uno o más componentes del sistema 101 de gestión de datos, el sistema 102 de automatización, y/o el ARMS 103 pueden mantenerse en una ubicación alejada del laboratorio y equipos asociados (por ejemplo, un sistema "basado en la nube"). En aún otra realización, uno o más componentes del sistema 101 de gestión de datos, el sistema 102 de automatización, y/o el ARMS 103 pueden mantenerse en una combinación de ubicaciones cercanas y alejadas.

60 La figura 9 ilustra una vista detallada de un sistema 200 de ejecución de laboratorio (LES). El LES 200 puede incluir una herramienta 210 de gestión de datos y un proceso 220 de automatización. Además, el LES 200 puede comunicarse con un módulo 240 de LIMS. En una realización, el módulo 240 de LIMS puede incluir al menos un módulo 206 de registro y un sistema 230 de gestión de resultados analíticos (ARMS), que se comentan con más detalle con respecto a la figura 10. La figura 9 representa además el dispositivo 201 de usuario y el módulo 202 de aplicación, que se describirá a continuación. El dispositivo 201 de usuario puede permitir que un usuario interactúe con el LES 200 y facilitar así la interacción del usuario con cada uno de la herramienta 210 de gestión de datos, el proceso 220 de automatización, y el ARMS 230, y/u otros sistemas asociados. El dispositivo 201 de usuario puede comunicarse con el módulo 202 de aplicación para realizar una o más funciones tal como se describe en el presente

documento.

En una realización, el módulo 202 de aplicación puede ser una interfaz de programación de aplicaciones (API) para realizar una o más funciones automatizadas. En otra realización, el módulo 202 de aplicación puede ser una interfaz gráfica de usuario (GUI), mediante la cual un usuario puede dar instrucciones al LES 200 para realizar una o más funciones, tales como cargar una secuencia de comandos (*script*), ejecutar un método de diagnóstico, ejecutar una acción de un instrumento de laboratorio, o similar. El dispositivo 201 de usuario también puede interactuar con el LES 200 mediante la interacción directa con otros componentes del sistema. Por ejemplo, el usuario 201 puede proporcionar un comando directamente al planificador 204 para corregir errores de tiempo de ejecución.

En otra realización, el dispositivo 208 de seguimiento de laboratorio facilita la gestión de la ubicación física de una o más unidades robóticas. Por ejemplo, el dispositivo 208 de seguimiento de laboratorio puede configurarse como una base de datos que almacena información de posición de todos los objetos físicos en un momento dado. El dispositivo 208 de seguimiento de laboratorio también puede recibir información desde otros componentes en el LES 200. Por ejemplo, el usuario 201 puede proporcionar un comando al módulo 208 de seguimiento de laboratorio para corregir un error de seguimiento de placa.

La figura 9 representa además el proceso 220 de automatización, que puede proporcionar la gestión del flujo de trabajo de placas de muestra, muestras, y datos asociados. Por ejemplo, el proceso 220 de automatización puede proporcionar información referente a las placas disponibles al módulo 202 de aplicación, o puede indicar de otro modo la disponibilidad de recursos del sistema para el módulo 202 de aplicación. Como otro ejemplo, el proceso 220 de automatización puede recibir información de notificación, tal como un informe de finalización de un trabajo, desde el módulo 202 de aplicación. El proceso 220 de automatización también puede recibir información de la canalización simiente, que puede introducirse manualmente por un usuario y proporcionarse directamente al proceso 220 de automatización desde el dispositivo 201 de usuario. La información de la canalización simiente puede incluir, por ejemplo, información para instanciar nuevos objetos para su gestión en el sistema LIMS. Por ejemplo, un usuario puede utilizar una GUI para crear muestras de investigación, en la que se introducen las muestras de investigación como información de la canalización simiente en el proceso 220 de automatización.

En otra realización, el proceso 220 de automatización puede recibir información de la canalización simiente de un módulo 206 de registro. En aún otra realización, el proceso 220 de automatización puede recibir información de consulta desde el ARMS 230. Por ejemplo, una consulta sobre los resultados que van a presentarse visualmente. El proceso 220 de automatización puede recibir además información de consulta desde el planificador 204, por ejemplo, una consulta sobre un trabajo pendiente. Además, el proceso 220 de automatización puede proporcionar una herramienta 210 de gestión de datos con información de validación de datos e información sobre consultas de datos.

Además, la figura 9 muestra la herramienta 210 de gestión de datos, que se describirá a continuación. La herramienta 210 de gestión de datos puede configurarse para integrar datos cuantitativos, realizar un seguimiento de códigos de barras de muestras, y gestionar el flujo de trabajo global del LES 200. En una realización, la herramienta 210 de gestión de datos puede recibir información referente a una operación de informe desde el módulo 202 de aplicación. En otra realización, la herramienta 210 de gestión de datos puede recibir una operación de informe desde el módulo 205 de desplazador. Además, la herramienta 210 de gestión de datos puede recibir un comando para corregir errores de seguimiento de placa desde un usuario a través del módulo 207 de dispositivo de seguimiento de laboratorio. En aún otra realización, la herramienta 210 de gestión de datos puede recibir, desde el planificador 204, una consulta referente a datos de estado. En un ejemplo, una consulta de este tipo se refiere a información sobre sellos, giros, o ubicación.

La figura 9 representa además el servidor 203 de secuencias de comandos y el repositorio 207, que se describirán a continuación. En una realización, el servidor 203 de secuencias de comandos puede comunicarse con un repositorio 207 de sistema de control de versiones (VCS) para obtener una o más secuencias de comandos de software para su uso en el funcionamiento del LES 200. El repositorio 207 de VCS puede mantenerse por repositorios conocidos tales como "Github", o cualquier otro servicio de repositorio de VCS apropiado, tal como apreciará un experto habitual en la técnica. En una realización, el servidor 203 de secuencias de comandos puede obtener secuencias de comandos de software desde el repositorio 207 de VCS, y puede enviar además una o más secuencias de comandos de software al módulo 202 de aplicación. El servidor 203 de secuencias de comandos puede configurarse además para implementar secuencias de comandos y gestionar metadatos de secuencias de comandos.

El planificador 204 puede configurarse para automatizar la planificación y ejecutar aplicaciones. Por ejemplo, el planificador 204 puede incluir al menos un módulo de software, tal como un compilador de secuencias de comandos, un planificador, y/o un ejecutor. En una realización, el planificador 204 puede proporcionar uno o más comandos para realizar una acción a un módulo 202 de aplicación o puede proporcionar además una consulta para una función de API a un módulo 202 de aplicación. En otra realización, el planificador 204 puede configurarse para iniciar y/o suministrar una o más consultas para una función de API, y puede configurarse además para iniciar y/o entregar una o más consultas sobre datos de estado. En otra realización, el planificador 204 puede configurarse para iniciar y/o suministrar una o más consultas referentes a un trabajo pendiente. En aún otra realización, el planificador 204 puede

configurarse para recibir un comando para corregir errores de tiempo de ejecución.

5 La aplicación 205 de desplazador puede configurarse para comunicarse con una o más unidades robóticas en un entorno de laboratorio. Por ejemplo, la aplicación 205 de desplazador puede facilitar indicar a la una o más unidades robóticas que realicen uno o más movimientos en el espacio tridimensional. La aplicación 205 de desplazador puede enviar instrucciones a la una o más unidades robóticas referentes a un movimiento, una trayectoria, una dirección, u otra información relacionada con el espacio tridimensional en el que la una o más unidades robóticas pueden realizar cualquier número de movimientos. En otra realización, el planificador 204 puede proporcionar uno o más comandos para realizar un movimiento al módulo 205 de desplazador, tal como por ejemplo, movimientos robóticos descritos con detalle con respecto a la figura 13.

15 Además, el LES 200 puede configurarse para comunicarse con el módulo 209 de fabricación. En una realización, el módulo 209 de fabricación está configurado para proporcionar al LES 200 información relacionada con componentes de muestra, tales como plástico, reactivos, y similares. Por ejemplo, el módulo 209 de fabricación puede asistir en la identificación de los componentes de muestra que se introducen en el LES 200. En otra realización, el módulo 209 de fabricación puede configurarse para declarar y generar etiquetas de código de barras para una o más placas de muestra y tubos de muestra.

20 El LES 200 puede comunicarse además con el módulo 211 SciComp. En una realización, el módulo 211 SciComp puede facilitar la automatización general del sistema LIMS al gestionar el procesamiento de todas las fases principales, incluyendo, pero sin limitarse a, (i) la adquisición de muestras físicas, (ii) la secuenciación, (iii) la generación de datos sin procesar, (iv) el análisis de datos, y (v) la transferencia de datos analizados al ARMS. Por ejemplo, el módulo 211 SciComp puede asistir en el proceso 220 de automatización consultando el proceso 200 de automatización para obtener información sobre el siguiente trabajo a procesar. El módulo 211 SciComp puede incluir además componentes tales como un servidor de secuencias de comandos y/o un planificador para mantener un flujo de trabajo eficiente. En una realización, el módulo 211 SciComp puede realizar las tareas de análisis de datos necesarias del sistema LIMS, y puede ejecutar los algoritmos necesarios para producir automáticamente llamadas de variantes de pacientes a partir de datos sin procesar a datos analizados.

30 Aunque sólo se enumera una instancia de cada módulo en la figura 9 (por ejemplo, un planificador 204 y un desplazador 205), el LES 200 puede incluir una o más instancias de cualquiera de tales módulos. Por ejemplo, puede haber dos o más instancias del planificador 204, cada una de las cuales está asociada a un proceso o dispositivo específico dentro del entorno de laboratorio.

35 La figura 10 ilustra una representación detallada del sistema 300 de gestión de información de laboratorio (LIMS). En una realización, el LIMS 300 incluye un módulo 301 de registro y módulo 302 de gestión de muestras. El módulo 301 de registro puede configurarse para registrar la llegada de una muestra e instanciar la llegada de la muestra en una o más bases de datos. Por ejemplo, el módulo 301 de registro puede configurarse para enviar un primer conjunto de información al ARMS 303. El primer conjunto de información puede incluir, por ejemplo, información relativa a una orden de panel de enfermedades. El módulo 302 de gestión de muestras puede configurarse para comunicarse con el módulo 301 de registro en la organización de una o más muestras que van a sembrarse en el ARMS 303. El módulo 306 de análisis puede recibir una o más salidas del ARMS 303, tales como los resultados relativos a una orden de panel de enfermedades. El LIMS 300 puede incluir además un módulo 308 de validación y un módulo 308 de información biológica. El módulo 308 de validación y el módulo 308 de información biológica pueden configurarse, cada uno, para asistir en el desarrollo de ensayos de muestra para pruebas.

50 Tal como se muestra en la figura 10, el LIMS 300 puede comunicarse además con el LES 310 y SciComp 320, tal como se comentó con respecto a la figura 9. El LIMS 300 puede incluir además un módulo 304 de revisión de llamadas, que puede configurarse para proporcionar técnicas de procesamiento para revisar y modificar datos de procesamiento de llamadas de variantes. El LIMS 300 puede incluir además un módulo 307 de base de datos para almacenar información relacionada con muestras y datos de prueba asociados, tal como se usa dentro del LIMS 300.

55 El ARMS 303 puede configurarse además como una base de datos que contiene genotipos para las muestras. Por ejemplo, el ARMS 303 puede configurarse para procesar, mantener, y suministrar información sobre datos de genotipado basándose en uno o más archivos de formato de llamadas de variantes (VCF). Tal como apreciará un experto habitual en la técnica, un archivo VCF es un formato de archivo de texto normalizado para representar y almacenar variaciones en la secuencia genética. En una realización, el ARMS 303 puede proporcionar una consulta de resultados a un proceso de automatización en el LES 320. Por ejemplo, puede usarse una consulta de resultados para determinar qué resultados pueden presentarse visualmente.

60 En otra realización, el ARMS 303 incluye la funcionalidad para generar una GUI, donde la GUI proporciona al usuario datos en tiempo real correspondientes a diagnósticos y análisis de laboratorio para una o más muestras. La GUI puede permitir al usuario realizar una pluralidad de funciones, incluyendo pero sin limitarse a, la monitorización y el ajuste del control de calidad (CC), la generación del historial de muestras, el etiquetado manual de muestras, y la capacidad de aprobar o rechazar manualmente una muestra. El ARMS 303 puede incluir una funcionalidad para

generar informes de diagnóstico personalizados, incluyendo la generación de gráficos, tablas, hojas de cálculo, representaciones gráficas, diagramas, y/u otras visualizaciones para permitir una interpretación eficiente de los datos.

5 La figura 11 ilustra un sistema 400 informático de propósito general en el que pueden implementarse uno o más sistemas, tal como se describe en el presente documento. El sistema 400 puede incluir, pero no se limita a, componentes conocidos tales como la unidad 401 central de procesamiento (CPU), el almacenamiento 402, la memoria 403, el adaptador 404 de red, la fuente 405 de alimentación, controladores 406 de entrada/salida (E/S), el bus 407 eléctrico, uno o más elementos 408 de presentación visual, uno o más dispositivos 409 de entrada de usuario, y otros dispositivos 410 externos. Los expertos en la técnica entenderán que el sistema 400 puede contener otros componentes bien conocidos que pueden añadirse, por ejemplo, a través de ranuras 412 de expansión, o mediante cualquier otro método conocido por los expertos en la técnica. Tales componentes pueden incluir, pero no se limitan a, componentes de redundancia de hardware (por ejemplo, fuentes de alimentación o unidades de copia de seguridad de datos duales), componentes de refrigeración (por ejemplo, ventiladores o sistemas de refrigeración por agua), memoria adicional y hardware de procesamiento, y similares.

El sistema 400 puede estar, por ejemplo, en forma de un ordenador cliente-servidor que puede conectarse a, y/o facilitar el funcionamiento de, una pluralidad de estaciones de trabajo o sistemas informáticos similares a través de una red. En otra realización, el sistema 400 puede conectarse a una o más estaciones de trabajo a través de una red de internet o intranet, y facilitar así la comunicación con un mayor número de estaciones de trabajo o sistemas informáticos similares. Incluso más, el sistema 400 puede incluir, por ejemplo, una estación de trabajo principal o un ordenador principal de propósito general para permitir interacción directa de un usuario con un servidor central. Alternativamente, el usuario puede interactuar con el sistema 400 a través de una o más estaciones 413 de trabajo remotas o locales. Tal como apreciará un experto habitual en la técnica, puede haber cualquier número práctico de estaciones de trabajo remotas para comunicarse con el sistema 400.

La CPU 401 puede incluir uno o más procesadores, por ejemplo procesadores Intel® Core™ i7, procesadores de la serie AMD FX™, u otros procesadores, tal como apreciarán los expertos en la técnica. La CPU 401 puede comunicarse además con un sistema operativo, tal como el sistema operativo Windows NT® de Microsoft Corporation, el sistema operativo Linux, o un sistema operativo tipo Unix. Sin embargo, un experto habitual en la técnica apreciará que también pueden utilizarse sistemas operativos similares. El almacenamiento 402 puede incluir uno o más tipos de almacenamiento, tal como conoce un experto habitual en la técnica, tales como una unidad de disco duro (HDD), una unidad de estado sólido (SSD), unidades híbridas, y similares. En un ejemplo, el almacenamiento 402 se utiliza para mantener datos de manera persistente para el almacenamiento a largo plazo. La memoria 403 puede incluir uno o más tipos de memoria tal como conoce un experto habitual en la técnica, tales como memoria de acceso aleatorio (RAM), memoria de sólo lectura (ROM), disco duro o cinta, memoria óptica, o unidad de disco duro extraíble. La memoria 403 puede utilizarse para acceder a la memoria a corto plazo, tal como por ejemplo, para cargar aplicaciones de software o gestionar procesos temporales del sistema.

Tal como apreciará un experto habitual en la técnica, el almacenamiento 402 y/o la memoria 403 pueden almacenar uno o más programas de software informático. Tales programas de software informático pueden incluir lógica, código, y/u otras instrucciones para permitir que el procesador 401 realice las tareas, operaciones, y otras funciones tal como se describe en el presente documento, así como tareas y funciones adicionales tal como apreciaría un experto habitual en la técnica. El sistema 402 operativo también puede funcionar en actuación conjunta con firmware, tal como se conoce bien en la técnica, para permitir que el procesador 401 coordine y ejecute diversas funciones y programas de software informático, tal como se describe en el presente documento. Tal firmware puede residir dentro del almacenamiento 402 y/o la memoria 403.

Además, los controladores 406 de E/S pueden incluir uno o más dispositivos para recibir, transmitir, procesar, y/o interpretar información procedente de una fuente externa, tal como conoce bien un experto habitual en la técnica. En una realización, los controladores 406 de E/S pueden incluir una funcionalidad para facilitar la conexión a uno o más dispositivos 409 de usuario, tales como uno o más teclados, ratones, micrófonos, almohadillas ópticas (*trackpads*), paneles táctiles, o similares. Por ejemplo, los controladores 406 de E/S pueden incluir un controlador de bus serie, un controlador de bus serie universal (USB), un controlador FireWire, y similares, para la conexión a cualquier dispositivo de usuario adecuado. Los controladores 406 de E/S también pueden permitir la comunicación con uno o más dispositivos inalámbricos a través de tecnología tal como, por ejemplo, la comunicación de campo cercano (NFC) o Bluetooth™. En una realización, los controladores 406 de E/S pueden incluir conjuntos de circuitos u otra funcionalidad para la conexión a otros dispositivos 410 externos tales como tarjetas de módem, tarjetas de interfaz de red, tarjetas de sonido, dispositivos de impresión, dispositivos de presentación visual externos, o similares. Además, los controladores 406 de E/S pueden incluir controladores para una variedad de dispositivos 408 de presentación visual conocidos por los expertos habituales en la técnica. Tales dispositivos de presentación visual pueden transmitir información visualmente a uno o más usuarios en forma de píxeles, y tales píxeles pueden estar dispuestos de manera lógica en el dispositivo de presentación visual para que el usuario pueda percibir la información reproducida en el dispositivo de presentación visual. Tales dispositivos de presentación visual pueden estar en forma de un dispositivo de pantalla táctil, dispositivo de presentación visual sin pantalla táctil tradicional, o cualquier otra forma de dispositivo de presentación visual, tal como apreciará un experto habitual en la técnica.

Además, la CPU 401 puede comunicarse además con controladores 406 de E/S para reproducir una interfaz gráfica de usuario (GUI), por ejemplo, en uno o más dispositivos 408 de presentación visual. En un ejemplo, la CPU 401 puede acceder al almacenamiento 402 y/o la memoria 403 para ejecutar uno o más programas y/o componentes de software que permitan al usuario interactuar con el sistema, tal como se describe en el presente documento. En una realización, una GUI, tal como se describe en el presente documento, incluye uno o más iconos u otros elementos gráficos con los que el usuario puede interactuar y realizar diversas funciones. Por ejemplo, la GUI 407 puede presentarse visualmente en un dispositivo 408 de presentación visual de pantalla táctil, mediante lo cual el usuario interactúa con la GUI a través de la pantalla táctil, tocando el usuario la pantalla físicamente, por ejemplo, con los dedos. Como otro ejemplo, la GUI puede presentarse visualmente en un elemento de presentación visual no táctil tradicional, mediante lo cual el usuario interactúa con la GUI mediante el teclado, el ratón, y otros componentes 409 de E/S convencionales. La GUI puede residir en el almacenamiento 402 y/o la memoria 403, al menos en parte, como un conjunto de instrucciones de software, tal como apreciará un experto habitual en la técnica. Además, la GUI no se limita a los métodos de interacción descritos anteriormente, ya que un experto habitual en la técnica puede apreciar cualquier variedad de medios para interactuar con una GUI, tales como métodos de interacción con un sistema informático basados en la voz u otros métodos adaptados a personas con discapacidad.

Además, el adaptador 404 de red puede permitir que el dispositivo 400 se comunique con la red 411. El adaptador 404 de red puede ser un controlador de interfaz de red, tal como un adaptador de red, una tarjeta de interfaz de red, un adaptador LAN, o similar. Tal como apreciará un experto habitual en la técnica, el adaptador 404 de red puede permitir la comunicación con una o más redes 411, tales como por ejemplo, una red de área local (LAN), una red de área metropolitana (MAN), una red de área amplia (WAN), una red en la nube (IAN), o Internet.

Una o más estaciones 413 de trabajo pueden incluir, por ejemplo, componentes conocidos tales como CPU, almacenamiento, memoria, adaptador de red, fuente de alimentación, controladores de E/S, bus eléctrico, uno o más elementos de presentación visual, uno o más dispositivos de entrada de usuario, y otros dispositivos externos. Tales componentes pueden ser iguales, similares, o comparables a los descritos con respecto al sistema 400 anteriormente. Los expertos habituales en la técnica entenderán que una o más estaciones 413 de trabajo pueden contener otros componentes bien conocidos, incluyendo pero sin limitarse a, componentes de redundancia de hardware, componentes de refrigeración, hardware de procesamiento/memoria adicional, y similares.

La figura 12 ilustra un proceso 500 de laboratorio a modo de ejemplo facilitado, por ejemplo, por el proceso 220 de automatización en la figura 9. En una realización, el proceso 220 de automatización proporciona al usuario la capacidad de crear procesos de flujo de trabajo de laboratorio para mantener colas de muestras para diagnóstico y análisis. Por ejemplo, un usuario puede crear uno o más objetos gráficos en un elemento de presentación visual de GUI, donde los objetos pueden representar uno o más estados, decisiones, entradas, salidas, u otras condiciones del laboratorio para modelar un proceso de laboratorio. Puede crearse un proceso de laboratorio resultante basado en el uno o más objetos gráficos creados por el usuario, tales como por ejemplo, un proceso como el que se representa gráficamente en la figura 12.

En una realización, el proceso 500 incluye un objeto 501 de agrupación de entrada, que puede representar, por ejemplo, una o más placas de reacción en cadena de la polimerasa (PCR). Las muestras de la agrupación de entrada pueden programarse para someterse a una o más pruebas, diagnósticos, u otros procesos 502 de laboratorio. Por ejemplo, las muestras dentro de la una o más placas de PCR pueden someterse a un proceso para la amplificación de ADN. La flecha 510 puede representar la transferencia de una placa 501 de PCR al proceso 502 de amplificación, por ejemplo. La flecha 520 puede representar una salida satisfactoria del proceso 502 de amplificación, tal como por ejemplo, una placa de PCR amplificada. La agrupación 503 de salida puede representar, por ejemplo, una o más placas de PCR amplificadas. Por tanto, la flecha 520 puede representar la transferencia de una placa de PCR amplificada a un objeto 503 de agrupación de salida. Aunque en el proceso 500 sólo se representan una entrada, un proceso, y una salida, se apreciará que cualquier número de entradas, salidas, procesos, transferencias, u otras funciones de laboratorio pueden representarse mediante tal diagrama gráfico, y que la divulgación no se limita al proceso a modo de ejemplo representado gráficamente en la figura 12.

La figura 13 ilustra un proceso 600 de transferencia a modo de ejemplo para controlar una unidad robótica que transporta uno o más tubos de muestra. El proceso 600 de transferencia puede facilitarse al menos en parte mediante el proceso 220 de automatización tal como se describe en las figuras 13 y 14. Por ejemplo, el proceso 220 de automatización puede comunicarse con procesos de hardware y software asociados a uno o más sistemas robóticos, de visión, y/o neumáticos para realizar procesos 600 de transferencia. En una realización, el proceso 600 de transferencia se utiliza para transferir al menos una muestra desde una ubicación de origen a una ubicación de destino mediante el uso de al menos una unidad robótica acoplada con sistemas de visión y neumáticos.

El proceso 600 de transferencia puede comenzar en la etapa 601, donde el proceso de automatización puede recibir una petición de transferencia de muestra. Tal petición puede ser, por ejemplo, una petición manual introducida por un usuario o una petición automatizada iniciada por un proceso de flujo de trabajo preprogramado. En una realización, la petición incluye información que identifica al menos un código de barras de muestra correspondiente a

una muestra actual, y puede incluir además información que identifica una ubicación de destino para transferir la muestra asociada al código de barras desde una ubicación de origen a una ubicación de destino.

5 En la etapa 602, el proceso de automatización puede enviar información de transferencia a un sistema de visión para identificar la ubicación espacial de la muestra identificada. En una realización, el sistema de visión realiza un proceso de coincidencia de visión en la etapa 603 para identificar si existe un código de barras coincidente dentro del área de observación del sistema de visión. Si se encuentra un código de barras coincidente, el sistema de visión puede enviar la información de ubicación espacial correspondiente al sistema robótico en la etapa 604. Tal información de ubicación espacial puede corresponder a la información de ubicación de muestra descubierta por el sistema de visión cuando identifica el código de barras coincidente en la etapa 603. La información de ubicación espacial puede estar en un formato legible para la unidad robótica con el fin de permitir que la unidad robótica identifique una ubicación tridimensional en el espacio correspondiente a la muestra física identificada.

15 En la etapa 605, la unidad robótica puede recibir y procesar la información de ubicación espacial y, además, sujetar la muestra identificada. Por ejemplo, la unidad robótica puede usar la información de ubicación espacial para mover un brazo robótico hasta una ubicación correspondiente a la posición directamente por encima de la muestra identificada. Entonces, puede hacerse descender el brazo robótico hasta una ubicación cercana a la muestra, y el brazo puede sujetar la muestra utilizando, por ejemplo, un sistema neumático. En un ejemplo, la muestra está contenida en un tubo de ensayo, que sujeta un brazo robótico, donde un sistema neumático genera un vacío para agarrar el tubo de ensayo.

25 En la etapa 606, el brazo robótico puede elevarse mientras sujeta la muestra, y el brazo robótico puede moverse hasta una ubicación correspondiente a la ubicación de destino recibida en la petición de transferencia de muestra. En la etapa 607, el brazo robótico puede hacer descender la muestra sobre una ubicación correspondiente a la ubicación deseada, y puede liberar la muestra del agarre robótico realizando uno o más procesos neumáticos a través del sistema neumático. Por ejemplo, el sistema neumático puede liberar el agarre de la muestra liberando el vacío y expulsando brevemente aire cerca de la muestra.

30 La figura 14 ilustra un sistema 700 robótico para gestionar procesos de laboratorio automáticos. En una realización, el sistema 700 robótico incluye un brazo 701 robótico para facilitar el movimiento de una o más muestras. Por ejemplo, el brazo 701 robótico puede configurarse para sujetar un tubo de ensayo que contiene una muestra y transportar el tubo de ensayo desde una primera ubicación hasta una segunda ubicación. En otro ejemplo, el brazo 701 robótico puede configurarse para sujetar una gradilla de muestras y transportar la gradilla de muestras desde una primera ubicación hasta una segunda ubicación. Una gradilla de muestras puede contener una o más muestras y puede almacenarse, por ejemplo, en un depósito 705 de gradilla de muestras. En una realización, el depósito 705 de gradillas de muestras puede contener una o más gradillas de muestras y puede facilitar el almacenamiento y la recuperación eficientes de una o más gradillas de muestras.

40 En una realización, el brazo 701 robótico puede fijarse además a la base 702 de brazo robótico y puede configurarse para rotar en un movimiento de 360 grados alrededor del entorno de laboratorio. Por ejemplo, el brazo 701 robótico puede extenderse desde una primera posición, tal como la posición representada gráficamente en la figura 14, hasta una segunda posición, tal como una posición que se extiende hacia un primer aparato 703a de manipulación de líquidos. Además, el brazo 701 robótico puede, por ejemplo, retraerse desde la posición extendida en el primer aparato 703a de manipulación de líquidos y volver a la posición tal como se representa gráficamente en la figura 14. Además, el brazo 701 robótico puede retraerse desde la posición en el primer aparato 703a de manipulación de líquidos, y luego extenderse hasta una posición dentro de un segundo aparato 703b de manipulación de líquidos. En una realización, el brazo robótico puede realizar diversos movimientos dentro del aparato 703a de manipulación de líquidos y el aparato 703b de manipulación de líquidos con el fin de facilitar diversos procedimientos de prueba de muestra.

50 En otra realización, el brazo 701 robótico puede configurarse para transportar una o más muestras y/o gradillas de muestras desde el depósito 705 de gradillas de muestras hasta el aparato 703a de manipulación de líquidos o el aparato 703b de manipulación de líquidos. El brazo 701 robótico puede configurarse además para devolver una o más muestras y/o gradillas de muestras desde el aparato 703a de manipulación de líquidos o el aparato 703b de manipulación de líquidos hasta el depósito 705 de gradillas de muestras, por ejemplo. Además, aunque sólo se representan gráficamente dos aparatos 703a y 703b de manipulación de líquidos en la figura 14, se apreciará que pueden implementarse aparatos de manipulación de líquidos adicionales dentro del entorno de laboratorio, y que el brazo 701 robótico puede extenderse hacia otras áreas similares dentro del alcance del brazo 701 robótico.

60 En aún otra realización, el brazo 701 robótico puede estar rodeado por uno o más sensores 704. Los sensores 704 pueden, por ejemplo, detectar movimientos específicos dentro de un área que rodea el brazo 701 robótico, tal como un área de detección de movimiento predefinida. En una realización, el área de detección de movimiento puede estar definida por una región esférica o semiesférica centrada en o cerca de un punto de acoplamiento del brazo 701 robótico a la base 702 de brazo robótico. En otra realización, el área de detección de movimiento puede estar definida por una región esférica o semiesférica centrada en o cerca de un punto específico en un espacio definido por un usuario. Por ejemplo, el área de detección de movimiento puede configurarse y actualizarse de manera

dinámica por un usuario y puede definir áreas tridimensionales personalizadas en el espacio que rodea el brazo 701 robótico.

Los sensores 704 pueden, por ejemplo, proporcionar señales a uno o más sistemas de software dentro del entorno de laboratorio para impedir que el brazo 701 robótico se mueva a áreas específicas dentro del entorno de laboratorio. En un ejemplo, los sensores 704 pueden configurarse para detectar movimientos asociados a un usuario u otro objeto dentro de un área de detección de movimiento especificada cerca del brazo 701 robótico. Si los sensores 704 detectan tales movimientos, los sensores 704 pueden enviar una o más señales de alarma a los sistemas de software asociados al brazo 701 robótico para detener todos los movimientos del brazo 701 robótico. Los sensores 704 pueden configurarse para, por ejemplo, enviar señales a sistemas de software asociados al brazo 701 robótico para reanudar los movimientos del brazo 701 robótico tras detectar los sensores 704 que cualquiera de tal usuario, objeto, u otro evento que provocó las señales de alarma ya no se encuentra dentro del área de detección de movimiento. En otra realización, los sensores 704 y el brazo 701 robótico pueden permanecer deshabilitados después de la señal de alarma hasta que se inicie y complete un proceso de reinicio predefinido por el usuario. Tras completarse tal proceso de reinicio por el usuario, el brazo 701 robótico y los sensores 704 podrán, por ejemplo, reanudar sus operaciones normales.

En otra realización, el sistema 700 robótico incluye un aparato 706 de manipulación de líquidos adicional que tiene una unidad robótica configurada para la extracción de ADN automatizada. El aparato 706 de manipulación de líquidos puede configurarse para manipular múltiples tamaños de tubo y/o múltiples tipos de muestra. Por ejemplo, aparatos 706 de manipulación de líquidos puede configurarse para manipular o bien un tamaño de tubo de 4 mm o bien un tamaño de tubo de 6 mm. En otro ejemplo, el aparato 706 de manipulación de líquidos puede configurarse para manipular o bien una muestra de sangre o bien una muestra de saliva. En otra realización, el sistema 700 robótico incluye un frigorífico 707 robótico que puede configurarse para almacenar y recuperar placas de muestra de uno o más tamaños diferentes. El frigorífico 707 robótico puede configurarse adicionalmente, por ejemplo, para permitir la anulación humana y permitir el acceso manual al contenido dentro del frigorífico 707 robótico.

La figura 15 ilustra una vista en ángulo de un brazo 800 robótico, por ejemplo, un brazo robótico tal como el brazo 701 robótico en la figura 14. Tal como se representa gráficamente en la figura 15, el brazo 800 robótico incluye una porción 801 de manipulación de muestras, una primera sección 802 de sólidos, una segunda sección 803 de sólidos, una tercera sección 804 de sólidos, y una base 805 de brazo robótico. En una realización, la porción 801 de manipulación de muestras está conectada a la primera sección 802 de sólidos. En otra realización, la primera sección 802 de sólidos está conectada en un extremo a la porción 801 de manipulación de muestras y está conectada en el otro extremo a la segunda sección 803 de sólidos. En aún otra realización, la segunda sección 803 de sólidos está conectada en un extremo a la primera sección 802 de sólidos y está conectada en el otro extremo a la tercera sección 804 de sólidos. En aún otra realización, la tercera sección 804 de sólidos está conectada en un extremo a la segunda sección 803 de sólidos y está conectada en el otro extremo a la base 805 de brazo robótico.

Dispositivo de recogida de tubos

Los dispositivos de clasificación de tubos se han descrito previamente. Por ejemplo, véase la solicitud de patente estadounidense n.º 15/388.193 presentada el 22 de diciembre de 2017, titulada "Robotic System for Sorting Sample Tubes".

Los dispositivos de clasificación de tubos incluyen un brazo robótico para recoger un tubo de una primera ubicación en una primera gradilla de tubos de muestra y clasificar el tubo a una segunda ubicación en la primera gradilla de tubos de muestra o en una segunda gradilla de tubos de muestra. La carga inicial de tubos en la primera gradilla de tubos puede ser manual o puede utilizar un segundo brazo robótico, por ejemplo, que incluya una interfaz entre otro sistema robótico y el dispositivo de clasificación de tubos. El uso de un sistema robótico, tal como se divulga en el presente documento, mejora la velocidad de clasificación, permitiendo, por ejemplo, clasificar más de 10, 20, 30, o 40 tubos por minuto. Los tubos pueden ser tubos de muestra, es decir, tubos que contienen una muestra de paciente, o un tubo de reactivo de ensayo, es decir, un tubo que contiene un reactivo útil en un ensayo tal como se describe en el presente documento. En una realización, el tubo es un tubo de reactivo de ensayo que contiene una sonda o un cebador.

El uso del dispositivo de recogida de tubos permite la rápida personalización del panel de sondas. La selección de una pluralidad de sondas de un panel de sondas preparadas puede realizarse rápidamente. La selección de la pluralidad de sondas puede realizarse en unos minutos, y los tubos que contienen cada una de las sondas se seleccionan, clasifican y envían a la estación que combinará una alícuota de las sondas de tal manera que la pluralidad de sondas sean únicas para la muestra que se somete a prueba.

El dispositivo de recogida de tubos permite la rápida creación de grandes paneles de sondas y elimina la limitación de la preparación manual de los paneles. Una vez determinadas las sondas seleccionadas, el dispositivo de recogida de tubos permite la rápida colocación de los tubos correctos en una gradilla para la preparación del panel distintivo.

Métodos de ensayo

En general, los métodos de ensayo del presente documento comprenden las siguientes etapas:

- 5 A) Cribar un espécimen, por ejemplo una muestra, para identificar un motivo distintivo genético que se compone de marcadores de segregación.
- B) Diseñar un reactivo personalizado/único, por ejemplo sondas, para detectar dicho motivo distintivo/marcadores.
- 10 C) Cribar un espécimen "desconocido" usando el reactivo personalizado/único para determinar en qué medida está presente el motivo distintivo.

15 Los métodos que se describen en el presente documento detectan la presencia de un motivo distintivo genético único. Los métodos encuentran uso, por ejemplo, en aplicaciones relacionadas con el cáncer. Sin embargo, se entenderá que también es posible una prueba prenatal no invasiva cuando se conoce(n) el/los perfil(es) genómico(s) paterno y/o materno. Por ejemplo, cuando cada progenitor es portador de una enfermedad, de tal manera que si el feto heredase una copia del alelo portador de cada progenitor, se vería afectado, entonces puede realizarse un análisis basado en la información genómica parental conocida, de manera análoga al ensayo basado en el cáncer.

20 Los ensayos descritos en el presente documento también encuentran uso en el análisis forense de ADN. Por ejemplo, para identificar a una persona en particular, por ejemplo, un sujeto individual o sospechoso, en una muestra de prueba que comprende una mezcla de ADN de múltiples fuentes, es decir, cuando se encuentra más de un contribuyente en una muestra biológica, se realiza entonces un ensayo basado en el ADN de la persona en particular usando una combinación única de sondas específicas para el individuo.

25

Fase I - Panel distintivo de marcadores/mutaciones y sondas de captura

Panel distintivo de mutaciones/marcadores

30 En algunas realizaciones, la secuenciación del ácido nucleico de la muestra se realiza usando secuenciación del genoma completo (WGS). En algunas realizaciones, se realiza una secuenciación dirigida y puede ser secuenciación o bien de ADN o bien de ARN. La secuenciación dirigida puede realizarse con respecto a un subconjunto del genoma completo. En algunas realizaciones, la secuenciación dirigida se realiza con respecto a intrones, exones, secuencias no codificantes o una combinación de los mismos. En otras realizaciones, se realiza una secuenciación del exoma completo (WES) dirigida del ADN de la muestra. El ADN se secuencia usando una plataforma de secuenciación de nueva generación (NGS), que consiste en una secuenciación masiva en paralelo. Las tecnologías de NGS proporcionan información de secuencia de alto rendimiento y proporcionan información cuantitativa digital, ya que cada lectura de secuencia que se alinea con la secuencia de interés es contabilizable. En ciertas realizaciones, los moldes de ADN amplificados de manera clonal o moléculas individuales de ADN se secuencian de manera masiva en paralelo dentro de una celda de flujo (por ejemplo, tal como se describe en Volkerding *et al.*, Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 11:31-46 [2010]; y el documento WO 2014/015084). Además de la información de secuenciación de alto rendimiento, la NGS proporciona información cuantitativa, ya que cada lectura de secuencia es contabilizable y representa un molde de ADN clonal individual o una sola molécula de ADN. Las tecnologías de secuenciación de NGS incluyen la pirosecuenciación, la secuenciación por síntesis con terminadores con colorante reversibles, la secuenciación por ligamiento de sondas de oligonucleótidos y la secuenciación por semiconductores iónicos. El ADN de muestras individuales puede secuenciarse individualmente (es decir, secuenciación singleplex) o el ADN de múltiples muestras puede agruparse y secuenciarse como moléculas genómicas indexadas (es decir, secuenciación múltiplex) en una sola tanda de secuenciación, para generar hasta varios cientos de millones de lecturas de secuencias de ADN. Las plataformas disponibles comercialmente incluyen, por ejemplo, plataformas para secuenciación por síntesis, secuenciación por semiconductores iónicos, pirosecuenciación, secuenciación por terminador con colorante reversible, secuenciación por ligamiento, secuenciación de una sola molécula, secuenciación por hibridación, y secuenciación por nanoporos. Están disponibles plataformas para secuenciación por síntesis, por ejemplo, de Illumina, 454 Life Sciences, Helicos Biosciences, y Qiagen. Las plataformas de Illumina pueden incluir, por ejemplo, la plataforma Solexa de Illumina y el sistema Genome Analyzer de Illumina, y se describen en Gudmundsson *et al* (Nat. Genet. 2009 41:1122-6), Out *et al* (Hum. Mutat. 2009 30:1703-12) y Turner (Nat. Methods 2009 6:315-6), las publicaciones de solicitudes de patente estadounidense n.º US20080160580 y US20080286795, las patentes estadounidenses n.ºs 6.306.597, 7.115.400, y 7.232.656. Las plataformas de 454 Life Science incluyen, por ejemplo, GS Flex y GS Junior, y se describen en la patente estadounidense n.º 7.323.305. Las plataformas de Helicos Biosciences incluyen la plataforma True Single Molecule Sequencing. Ion Torrent, un sistema alternativo de NGS, está disponible de ThermoScientific y es una tecnología basada en semiconductores que detecta los iones de hidrógeno que se liberan durante la polimerización de ácidos nucleicos. Cualquier método de detección que permita la detección de marcadores segregables puede usarse con el ensayo proporcionado en el presente documento.

65 En algunas realizaciones, puede obtenerse ADN de tejido fresco tal como el obtenido de una biopsia de un tumor primario no resecado y/o de una masa metastásica. En otras realizaciones, el ADN puede obtenerse de tejido fijado

con formalina e incorporado en parafina (FFPE). El tejido FFPE ha demostrado ser un sustrato adecuado para la secuenciación NGS y el análisis, y abre los especímenes clínicos y de archivo a enfoques de secuenciación de alto rendimiento para el análisis de todo el espectro de mutaciones del ADN (Duncavage *et al.* J Mol Diagn 13:325-333 [2011]). En aún otras realizaciones, el ADN tumoral puede obtenerse de tejido congelado en un banco. En algunas realizaciones, la muestra es una muestra de sangre que comprende ADN fetal libre de células.

Se obtiene tejido normal, es decir, tejido no tumoral del mismo paciente, y puede ser una muestra de biopsia de tejido normal fresco, una muestra FFPE normal, o una muestra congelada normal. En algunas realizaciones, la muestra normal es una muestra de tejido coincidente, es decir, una muestra obtenida de la porción no tumoral del mismo tejido del que se obtuvo la muestra tumoral. En otras realizaciones, la muestra normal puede obtenerse de un tejido diferente del que se obtuvo la muestra tumoral, es decir, una muestra de tejido no coincidente. En aún otras realizaciones, la muestra normal puede obtenerse de células sanguíneas normales.

El ADN se extrae de ambas muestras de tejido tumoral y de tejido normal y se secuencian de manera masiva en paralelo usando uno cualquiera de los métodos de secuenciación de nueva generación (NGS), tal como se describió anteriormente. Se conocen bien métodos para aislar ácidos nucleicos de fuentes biológicas y diferirán dependiendo de la naturaleza de la fuente. Un experto en la técnica puede aislar fácilmente el/los ácido(s) nucleico(s) de una fuente según sea necesario para el método descrito en el presente documento. En algunos casos, puede resultar ventajoso fragmentar las moléculas de ácido nucleico en la muestra de ácidos nucleicos. La fragmentación puede ser aleatoria o puede ser específica, tal como se logra, por ejemplo, usando digestión con endonucleasas de restricción. Los métodos para la fragmentación aleatoria se conocen bien en la técnica e incluyen, por ejemplo, la digestión con ADNasa limitada, el tratamiento alcalino y el cizallamiento físico. En una realización, se obtienen ácidos nucleicos de la muestra como ADNlc, que no se somete a fragmentación.

En algunas realizaciones ilustrativas, el ADN de muestra obtenido de tejido tumoral y tejido normal es ADN genómico celular, que se somete a fragmentación para dar fragmentos de aproximadamente 300 pares de bases o más, aproximadamente 400 pares de bases o más, o aproximadamente 500 pares de bases o más, y a los que pueden aplicarse fácilmente métodos NGS.

En algunas realizaciones, se realiza la secuenciación del genoma completo (WGS) del ADN tumoral y normal.

En otras realizaciones, se realiza la secuenciación del exoma completo (WES) del ADN tumoral y normal. La WES comprende seleccionar secuencias de ADN que codifican para proteínas y secuenciar ese ADN usando cualquier tecnología de secuenciación de ADN de alto rendimiento. Los métodos que pueden usarse para seleccionar como diana el ADN del exoma incluyen el uso de la reacción en cadena de la polimerasa (PCR), sondas de inversión molecular (MIP), captura híbrida, y captura en disolución. La utilidad de los enfoques genómicos dirigidos está bien establecida, y los métodos disponibles comercialmente para la WES incluyen el alineamiento de captura de Roche NimbleGen (Roche NimbleGen Inc., Madison, WI), SureSelect de Agilent (Agilent Technologies, Santa Clara, CA), y la PCR en emulsión de RainDance Technologies (RainDance Technologies, Lexington, MA), el panel de investigación del exoma xGen® de IDT y otros.

En aún otras realizaciones, la secuenciación dirigida puede centrarse en un conjunto selecto de genes, regiones génicas, o amplicones que tienen asociaciones conocidas al cáncer o una enfermedad hereditaria. En algunas realizaciones, pueden usarse paneles fijos o disponibles comercialmente que contienen genes asociados al cáncer conocidos. En otras realizaciones, pueden usarse paneles personalizados de diseño de genes asociados al cáncer conocidos.

Se comparan las secuencias obtenidas de las muestras de tejido tumoral y normal del paciente, y se identifican las mutaciones somáticas que están presentes sólo en el ADN tumoral y se anotan para su uso en la creación de un panel distintivo de marcadores, es decir, mutaciones somáticas, que son específicas para el paciente. Se identifican una pluralidad de mutaciones somáticas que son únicas para el paciente para crear un panel distintivo específico del paciente de marcadores, es decir, mutaciones somáticas. En algunas realizaciones, el conjunto de mutaciones somáticas identificadas a partir de la comparación del ADN normal y tumoral del paciente puede incluir más de 50, hasta 100, hasta 200, hasta 300, hasta 400, hasta 500, hasta 600, hasta 700, hasta 800, hasta 900, hasta 1000, hasta 1500, hasta 2000, hasta 2500, hasta 3000, hasta 4000, hasta 5000, hasta 6000, hasta 7000, hasta 8000, hasta 9000, hasta 10.000, hasta 11.000, hasta 12.000, hasta 13.000, hasta 14.000, hasta 15.000, o más de 15.000 mutaciones. En otras realizaciones, el conjunto de mutaciones únicas identificadas a partir de la comparación del ADN normal y tumoral de los pacientes incluye entre 50 y 15.000 mutaciones, entre 100 y 15.000 mutaciones, entre 500 y 13.000 mutaciones, entre 1000 y 10.000 mutaciones, entre 2000 y 8000 mutaciones, o entre 4000 y 6000 mutaciones.

Pueden identificarse marcadores, por ejemplo mutaciones, comparando el ADN de una muestra con una secuencia de referencia, por ejemplo el genoma humano de referencia. Las mutaciones somáticas que pueden identificarse comparando secuencias de tejido normal y tumoral incluyen cambios de un solo par de bases, por ejemplo polimorfismos de un solo nucleótido (SNP), polimorfismos de múltiples nucleótidos, inserciones, deleciones, e inversiones. La identificación de mutaciones somáticas se realiza alineando las lecturas de secuencia con un

genoma de referencia, por ejemplo hg18. En algunas realizaciones, las lecturas de secuencia comprenden aproximadamente 20 pb, aproximadamente 25 pb, aproximadamente 30 pb, aproximadamente 35 pb, aproximadamente 40 pb, aproximadamente 45 pb, aproximadamente 50 pb, aproximadamente 55 pb, aproximadamente 60 pb, aproximadamente 65 pb, aproximadamente 70 pb, aproximadamente 75 pb, aproximadamente 80 pb, aproximadamente 85 pb, aproximadamente 90 pb, aproximadamente 95 pb, aproximadamente 100 pb, aproximadamente 110 pb, aproximadamente 120 pb, aproximadamente 130 pb, aproximadamente 140 pb, aproximadamente 150 pb, aproximadamente 200 pb, aproximadamente 250 pb, aproximadamente 300 pb, aproximadamente 350 pb, aproximadamente 400 pb, aproximadamente 450 pb, aproximadamente 500 pb, o más de 500 pb.

Posteriormente, se selecciona un subconjunto o la totalidad de las mutaciones somáticas específicas del paciente identificadas para obtener un panel distintivo de mutaciones que son específicas del paciente, y que puede interrogarse repetidamente. EL panel distintivo puede interrogarse, por ejemplo, en diferentes momentos para monitorizar la regresión de la enfermedad, en diferentes momentos para monitorizar la progresión de la enfermedad, en diferentes momentos para detectar una recaída tras la remisión. El panel de mutaciones también puede interrogarse en momentos tras, coincidiendo con y/o antes de la cirugía, tras, durante y/o antes de quimioterapia, tras, durante y/o antes de radioterapia. El panel de marcadores, por ejemplo mutaciones, también puede interrogarse en momentos antes de, coincidiendo con, y/o tras una prueba de obtención de imágenes, tal como una exploración por TEP, una exploración por TEP/TC, IRM, o una radiografía.

En algunas realizaciones, el panel distintivo de marcadores, por ejemplo mutaciones, que puede interrogarse para el paciente, puede ser un conjunto único de mutaciones somáticas que se identifican específicamente en el genoma del paciente. El panel distintivo de marcadores es un conjunto de una mezcla de mutaciones somáticas que se sabe que están asociadas a la enfermedad del paciente, y mutaciones somáticas presentes en el genoma del paciente que no se sabe que estén asociadas a la enfermedad del paciente. En aún otras realizaciones, el panel distintivo de marcadores puede ser un conjunto de mutaciones somáticas que actualmente se sabe que están asociadas a la enfermedad del paciente. Los paneles distintivos pueden ser conjuntos de mutaciones somáticas que se sabe que están asociadas a varias enfermedades, tales como cáncer, enfermedades neurodegenerativas, enfermedades infecciosas, enfermedades autoinmunitarias, anemia y fibrosis quística.

En algunas realizaciones, el panel distintivo es un conjunto de mutaciones somáticas que se han identificado en un tumor de un paciente con cáncer. En algunas realizaciones, el panel distintivo puede ser un conjunto de mutaciones somáticas que se sabe que están asociadas a uno cualquiera de una variedad de tumores sólidos, incluyendo tumores metastásicos de cánceres de origen sanguíneo, por ejemplo linfomas. En otras realizaciones, el panel distintivo es un conjunto de mutaciones que se han identificado en el paciente con cáncer, pero que no son mutaciones somáticas que se sabe que están asociadas al tipo de cáncer del paciente que está sometiéndose a prueba. En aún otras realizaciones, el panel distintivo es un conjunto de una mezcla de mutaciones somáticas que se sabe que están asociadas al tipo de cáncer que presenta el paciente y mutaciones somáticas que no se sabe que estén asociadas al tipo de cáncer del paciente.

Los tipos de cánceres que pueden detectarse y/o monitorizarse según el método proporcionado incluyen, pero no se limitan a, cáncer suprarrenal, cáncer anal, cáncer de las vías biliares, cáncer de vejiga, cáncer de hueso, tumores cerebrales/del SNC en adultos, tumores cerebrales/del SNC en niños, cáncer de mama, cáncer de mama en hombres, cáncer en adolescentes, cáncer en niños, cáncer en adultos jóvenes, cáncer de origen primario desconocido, enfermedad de Castleman, cáncer de cuello uterino, cáncer de colon/recto, cáncer de endometrio, cáncer de esófago, tumores de la familia de Ewing, cáncer ocular, cáncer de vesícula biliar, tumores carcinoides gastrointestinales, tumor del estroma gastrointestinal (GIST), enfermedad trofoblástica gestacional, enfermedad de Hodgkin, sarcoma de Kaposi, cáncer de riñón, cáncer de laringe e hipofaringe, leucemia, leucemia linfocítica aguda (LLA) en adultos, leucemia mieloide aguda (LMA), leucemia linfocítica crónica (LLC), leucemia mieloide crónica (LMC), leucemia mielomonocítica crónica (LMMC), leucemia en niños, cáncer de hígado, cáncer de pulmón, cáncer de pulmón no microcítico, cáncer de pulmón microcítico, tumor carcinoide de pulmón, linfoma, linfoma de la piel, mesotelioma maligno, mieloma múltiple, síndrome mielodisplásico, cáncer de cavidad nasal y seno paranasal, cáncer de nasofaringe, neuroblastoma, linfoma no Hodgkin, linfoma no Hodgkin en niños, cáncer de cavidad oral y orofaringe, osteosarcoma, cáncer de ovario, cáncer de páncreas, cáncer de pene, tumores de hipófisis, cáncer de próstata, retinoblastoma, rabdomiosarcoma, cáncer de glándulas salivales, sarcoma: cáncer de tejidos blandos en adultos, cáncer de piel, cáncer de piel: células basales y escamosas, cáncer de piel: melanoma, cáncer de piel: células de Merkel, cáncer de intestino delgado, cáncer de estómago, cáncer de testículo, cáncer de timo, cáncer de tiroides, sarcoma uterino, cáncer vaginal, cáncer de vulva, macroglobulinemia de Waldenström, y tumor de Wilms.

60 Sondas de captura

El panel distintivo está representado por un conjunto de sondas de captura de oligonucleótidos, cada una diseñada para hibridar, al menos parcialmente, con una secuencia diana que se ha identificado que comprende la mutación identificada en la muestra tumoral del paciente o en la secuencia parental. En algunas realizaciones, el panel distintivo comprende sondas de captura que comprenden las mutaciones somáticas identificadas en el tumor del paciente, y sondas de captura de la secuencia no mutada correspondiente, es decir, secuencia diana normal. En

algunas realizaciones, la sonda de captura está diseñada para hibridarse selectivamente con la secuencia diana. La sonda de captura puede ser complementaria en al menos el 70 %, 75 %, 80 %, 90 %, 95 %, o más del 95 % a una secuencia diana. En algunas realizaciones, la sonda de captura es complementaria al 100 % a una secuencia diana. En algunas realizaciones, las sondas de captura son sondas de ADN. En otras realizaciones, las sondas de captura pueden ser ARN (Gnirke *et al.* "Solution hybrid selection with ultra-long oligonucleotides for massively parallel sequencing", publicado en Nature Biotechnology, febrero de 2009;27(2):182-9. doi: 10.1038/nbt.1523. Publicación electrónica: 1 de febrero de 2009).

La sonda de captura generalmente tiene la longitud suficiente como para abarcar la secuencia de la mutación somática, o la secuencia normal correspondiente comprendida en la secuencia genómica seleccionada como diana por la sonda de captura. La longitud y la composición de una sonda de captura pueden depender de muchos factores, incluyendo la temperatura de la reacción de apareamiento, la fuente y la composición de bases del oligonucleótido, y la razón estimada de la sonda con respecto a la secuencia genómica diana. Adicionalmente, la longitud de la sonda de captura depende de la longitud de la secuencia diana que está diseñada que capture. El método proporcionado utiliza ADN libre de células (ADNlc), incluyendo ADN tumoral circulante (ADNtc), como la fuente de las secuencias diana que van a capturarse. Por consiguiente, dado que el ADNlc está altamente fragmentado, con un promedio de aproximadamente 170 pb, la sonda de captura puede tener, por ejemplo, entre 100 y 300 pb, entre 150 y 250 pb, o entre 175 y 200 pb. Actualmente, los métodos conocidos en la técnica describen sondas que suelen tener una longitud de más de 120 bases. En una realización actual, si el alelo tiene una o pocas bases, entonces las sondas de captura pueden tener menos de aproximadamente 110 bases, menos de aproximadamente 100 bases, menos de aproximadamente 90 bases, menos de aproximadamente 80 bases, menos de aproximadamente 70 bases, menos de aproximadamente 60 bases, menos de aproximadamente 50 bases, menos de aproximadamente 40 bases, menos de aproximadamente 30 bases, y menos de aproximadamente 25 bases, y esto es suficiente para garantizar un enriquecimiento equitativo de todos los alelos. Cuando la mezcla de ADN que va a enriquecerse usando la tecnología de captura híbrida es una mezcla que comprende ADNlc aislado de sangre, la longitud promedio del ADN es bastante corta, normalmente de menos de 200 bases. El uso de sondas más cortas dan como resultado una mayor probabilidad de que las sondas de captura híbridas capturen los fragmentos de ADN deseados. Variaciones mayores pueden requerir sondas más largas. En una realización, las variaciones de interés tienen una longitud de una base (un SNP) a unas pocas bases. En una realización, pueden enriquecerse preferentemente regiones seleccionadas como diana del genoma usando sondas de captura híbridas, en las que estas sondas de captura híbridas tienen una longitud menor de 90 bases y pueden tener menos de 80 bases, menos de 70 bases, menos de 60 bases, menos de 50 bases, menos de 40 bases, menos de 30 bases, o menos de 25 bases. En una realización, para aumentar la probabilidad de secuenciar el alelo deseado, puede reducirse la longitud de la sonda que está diseñada para hibridarse con las regiones que flanquean la ubicación del alelo polimórfico, desde más de 90 bases hasta aproximadamente 80 bases, o hasta aproximadamente 70 bases, o hasta aproximadamente 60 bases, o hasta aproximadamente 50 bases, o hasta aproximadamente 40 bases, o hasta aproximadamente 30 bases, o hasta aproximadamente 25 bases.

En una realización, las sondas de captura híbridas pueden diseñarse de tal manera que la región de la sonda de captura con ADN que es complementario al ADN que se encuentra en las regiones que flanquean el alelo polimórfico no sea inmediatamente adyacente al sitio polimórfico. En su lugar, la sonda de captura puede diseñarse de tal manera que la región de la sonda de captura que está diseñada para hibridarse con el ADN que flanquea el sitio polimórfico de la diana esté separada de la porción de la sonda de captura que estará en contacto de van der Waals con el sitio polimórfico en una pequeña distancia cuya longitud es equivalente a una base o un pequeño número de bases. En una realización, la sonda de captura híbrida está diseñada para hibridarse con una región que flanquea el alelo polimórfico, pero no lo atraviesa; esto puede denominarse sonda de captura flanqueante. La longitud de la sonda de captura flanqueante puede ser menor de aproximadamente 120 bases, menor de aproximadamente 110 bases, menor de aproximadamente 100 bases, menor de aproximadamente 90 bases, menor de aproximadamente 80 bases, menor de aproximadamente 70 bases, menor de aproximadamente 60 bases, menor de aproximadamente 50 bases, menor de aproximadamente 40 bases, menor de aproximadamente 30 bases, o menor de aproximadamente 25 bases. La región del genoma que se selecciona como diana por la sonda de captura flanqueante puede estar separada por el locus polimórfico en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-20, o más de 20 pares de bases.

Para pequeñas inserciones o deleciones, una o más sondas que se superpongan a la mutación pueden ser suficientes para capturar y secuenciar los fragmentos que comprenden la mutación. La hibridación puede ser menos eficiente si se compara con la eficiencia de captura limitante de sonda, que suele estar diseñada para la secuencia del genoma de referencia. Para garantizar la captura de los fragmentos que comprenden la mutación, podrían diseñarse dos sondas: una que coincida con el alelo normal y otra que coincida con el alelo mutante. Una sonda más larga puede mejorar la hibridación. Múltiples sondas superpuestas pueden mejorar la captura. Finalmente, colocar una sonda inmediatamente adyacente a la mutación, pero sin superponerse a la misma, puede permitir una eficiencia de captura relativamente similar de los alelos normal y mutante.

Para las repeticiones cortas en tándem (STR), es improbable que una sonda que se superponga a estos sitios altamente variables capture correctamente el fragmento. Para mejorar la captura, podría colocarse una sonda adyacente al sitio variable, pero sin superponerse al mismo. El fragmento podría secuenciarse entonces de manera

normal para revelar la longitud y la composición del STR.

Para delecciones grandes, una serie de sondas superpuestas, un enfoque común usado actualmente en sistemas de captura de exones, puede ser eficaz. Sin embargo, con este enfoque puede resultar difícil determinar si un individuo es heterocigoto o no. Según el método proporcionado, las sondas personalizadas están diseñadas para garantizar la captura del conjunto único de mutaciones somáticas identificadas en el tumor del paciente.

Una prueba de detección de enfermedades basada en captura dirigida podría combinarse con otra prueba de diagnóstico prenatal no invasiva basada en captura dirigida, por ejemplo, para aneuploidía.

Las sondas de captura pueden modificarse para comprender restos de purificación que sirven para aislar el dúplex de captura de las secuencias de ADNlc no hibridadas ni dirigidas mediante la unión a una pareja de unión del resto de purificación. Las parejas de unión adecuadas para su uso en la invención incluyen, pero no se limitan a, antígenos/anticuerpos (por ejemplo, digoxigenina/anti-digoxigenina, dinitrofenilo (DNP)/anti-DNP, dansil-X-antidansilo, fluoresceína/anti-fluoresceína, amarillo Lucifer/anti-amarillo Lucifer y rodamina anti-rodamina); biotina/avidina (o biotina/estreptavidina); proteína de unión a calmodulina (CBP)/calmodulina; hormona/receptor hormonal; lectina/hidrato de carbono; péptido/receptor de membrana celular; proteína A/anticuerpo; hapteno/anti-hapteno; enzima/cofactor; y enzima/sustrato. Otros pares de unión adecuados incluyen polipéptidos tales como el péptido FLAG (Hopp *et al.*, *BioTechnology*, 6:1204-1210 (1988)); el péptido epitópico KT3 (Martin *et al.*, *Science*, 255:192-194 (1992)); péptido epitópico de la tubulina (Skinner *et al.*, *J. Biol. Chem.*, 266: 15163-15166 (1991)); y la etiqueta peptídica de la proteína 10 del gen T7 (Lutz-Freyermuth *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:6393-6397 (1990)) y los anticuerpos para cada uno de los mismos. Los ejemplos no limitativos adicionales de parejas de unión incluyen agonistas y antagonistas de receptores de membrana celular, toxinas y venenos, epítomos virales, hormonas tales como esteroides, receptores hormonales, péptidos, enzimas y otros polipéptidos catalíticos, sustratos enzimáticos, cofactores, fármacos, incluyendo fármacos de molécula orgánica pequeña, opiáceos, receptores de opiáceos, lectinas, azúcares, sacáridos, incluyendo polisacáridos, proteínas, y anticuerpos, incluyendo anticuerpos monoclonales y fragmentos de anticuerpos sintéticos, células, membranas celulares y restos en las mismas, incluyendo receptores de membrana celular, y orgánulos. En algunas realizaciones, la primera pareja de unión es un resto reactivo y la segunda pareja de unión es una superficie reactiva que reacciona con el resto reactivo, tal como se describe en el presente documento con respecto a otros aspectos de la invención. En algunas realizaciones, los cebadores de oligonucleótidos se unen a la superficie sólida antes de iniciar la reacción de extensión. Se conocen en la técnica métodos para la adición de parejas de unión para capturar las sondas de oligonucleótidos e incluyen la adición durante la síntesis (por ejemplo, mediante el uso de un nucleótido modificado que comprende la pareja de unión) o después de la misma. Además, las sondas de captura pueden anclarse a una superficie sólida, por ejemplo, una perla magnética, lo que facilita el aislamiento de las secuencias capturadas.

La síntesis de sondas de captura puede realizarse mediante métodos conocidos en la técnica. Por ejemplo, puede obtenerse la síntesis paralela de sondas de captura usando un microalineamiento de Agilent que permite la síntesis de oligonucleótidos largos, por ejemplo, de 200 meros. Las sondas de captura pueden modificarse durante la síntesis o tras la síntesis para comprender restos que permitan aislar el dúplex formado por la sonda de captura mientras se hibrida con la secuencia diana, tal como se describe en otra parte del presente documento.

Fase II - Detección y monitorización de tumores mediante el análisis de ADNlc

Captura de mutaciones somáticas en ADNlc

En la segunda fase del método, las muestras usadas para determinar la fracción tumoral del paciente incluyen muestras que contienen ácidos nucleicos que están "libres de células" (por ejemplo, ADN libre de células). Pueden obtenerse ácidos nucleicos libres de células, incluido el ADN libre de células (ADNlc), mediante diversos métodos conocidos en la técnica a partir de muestras biológicas que incluyen, pero no se limitan a, plasma, suero y orina (véanse, por ejemplo, Fan *et al.*, *Proc Natl Acad Sci* 105:16266-16271 [2008]; Koide *et al.*, *Prenatal Diagnosis* 25:604-607 [2005]; Chen *et al.*, *Nature Med.* 2: 1033-1035 [1996]; Lo *et al.*, *Lancet* 350: 485-487 [1997]; Botezatu *et al.*, *Clin Chem.* 46: 1078-1084, 2000; y Su *et al.*, *J. Mol. Diagn.* 6: 101-107 [2004]). Otras muestras de líquidos biológicos incluyen, pero no se limitan a, sangre, sudor, lágrimas, esputo, flujo ótico, linfa, saliva, líquido cefalorraquídeo, lavados, suspensión de médula ósea, flujo vaginal, lavado transcervical, líquido cerebral, ascitis, leche, secreciones de las vías respiratorias, el tracto intestinal y aparato genitourinario, líquido amniótico, leche, y muestras de leucoforesis. En algunas realizaciones, la muestra es una muestra que puede obtenerse fácilmente mediante procedimientos no invasivos, por ejemplo, sangre, plasma, suero, sudor, lágrimas, esputo, orina, flujo ótico, saliva o heces. En ciertas realizaciones, la muestra es una muestra de sangre periférica, o las fracciones de plasma y/o suero de una muestra de sangre periférica. En otras realizaciones, la muestra biológica es un hisopo o frotis, un espécimen de biopsia, o un cultivo celular. En otra realización, la muestra es una mezcla de dos o más muestras biológicas, por ejemplo, una muestra biológica puede comprender dos o más de una muestra de líquido biológico, una muestra de tejido, y una muestra de cultivo celular. Tal como se usa en el presente documento, los términos "sangre", "plasma" y "suero" abarcan expresamente fracciones o porciones procesadas de las mismas.

En ciertas realizaciones, pueden obtenerse muestras de fuentes incluyendo, pero sin limitarse a, muestras de

diferentes individuos, muestras de diferentes fases de desarrollo del mismo individuo o de diferentes individuos, muestras de diferentes individuos enfermos (por ejemplo, individuos con cáncer o que se sospecha que presentan un trastorno genético), individuos normales, muestras obtenidas en diferentes estadios de una enfermedad en un individuo, muestras obtenidas de un individuo sometido a diferentes tratamientos para una enfermedad, muestras de individuos sometidos a diferentes factores ambientales, muestras de individuos con predisposición a una patología, muestras de individuos expuestos a un agente de una enfermedad infecciosa (por ejemplo, VIH), y similares. En una realización ilustrativa, pero no limitativa, la muestra es una muestra de sangre obtenida de un paciente con cáncer.

En diversas realizaciones, el ADNlc presente en la muestra puede enriquecerse de manera específica o inespecífica antes de su uso (por ejemplo, antes de la captura y secuenciación). El enriquecimiento inespecífico del ADN de la muestra se refiere a la amplificación de genoma completo de los fragmentos de ADN genómico de la muestra, que puede usarse para aumentar el nivel de ADN de la muestra antes de la captura y secuenciación. El enriquecimiento inespecífico puede ser el enriquecimiento selectivo de exomas. Se conocen en la técnica métodos para la amplificación de genoma completo. La PCR cebada con oligonucleótidos degenerados (DOP), la técnica de PCR con extensión de cebadores (PEP) y la amplificación por desplazamiento múltiple (MDA) son ejemplos de métodos de amplificación de genoma completo. En algunas realizaciones, la muestra no está enriquecida en ADNlc.

Tal como se describe en otra parte del presente documento, el ADNlc está presente como fragmentos con un promedio de aproximadamente 170 pb. Por tanto, no se requiere una fragmentación adicional del ADNlc. En algunas realizaciones, se obtiene suficiente ADN libre de células de una muestra de sangre de 10 ml para determinar con certeza la presencia o ausencia de cáncer en un paciente. Las muestras de sangre usadas en el método proporcionado pueden ser de aproximadamente 5 ml, aproximadamente 10 ml, aproximadamente 15 ml, aproximadamente 20 ml, aproximadamente 25 ml o más de 25 ml. Normalmente, 20 ml de plasma sanguíneo contienen entre 5000 y 10.000 equivalentes de genoma y proporcionan ADNlc más que suficiente para determinar la fracción tumoral según el método proporcionado. En algunas realizaciones, se obtiene suficiente ADNlc a partir de 10 ml a 20 ml de sangre para determinar la fracción tumoral.

Para separar el ADN libre de células a partir de células en una muestra, pueden usarse diversos métodos incluyendo, pero sin limitarse a, fraccionamiento, centrifugación (por ejemplo, centrifugación en gradiente de densidad), precipitación específica de ADN, o clasificación celular de alto rendimiento y/u otros métodos de separación. Existen kits disponibles comercialmente para la separación manual y automatizada de ADNlc (Roche Diagnostics, Indianápolis, Ind.; Qiagen, Germantown, MD).

El ADNlc puede someterse a reparación de extremos y, opcionalmente, puede añadirse una cola de dA, y se ligan adaptadores bicatenarios que comprenden secuencias complementarias a los cebadores de amplificación y secuenciación a los extremos de las moléculas de ADNlc para permitir la secuenciación NGS, por ejemplo, usando una plataforma de Illumina. Adicionalmente, cada uno de los adaptadores bicatenarios comprende además una secuencia de código de barras no aleatoria que permite diferenciar moléculas de ADNlc individuales. En algunas realizaciones, las secuencias de código de barras son secuencias aleatorias. En otras realizaciones, las secuencias de código de barras son secuencias de código de barras no aleatorias. Las secuencias de código de barras no aleatorias ofrecen una ventaja significativa con respecto a las secuencias de código de barras aleatorias, ya que las secuencias de código de barras no aleatorias permiten la identificación inequívoca de las lecturas de secuenciación que se describen a continuación. Las secuencias de código de barras no aleatorias están diseñadas específicamente para mantener el equilibrio de bases tanto dentro como entre todos los códigos de barras. Adicionalmente, en algunas realizaciones, los códigos de barras no aleatorios pueden comprender un nucleótido T en el extremo 3', complementario al nucleótido A de las moléculas de ADNlc con cola de dA. En realizaciones que usan una proyección de nucleótido T en el extremo 3' del código de barras, pueden diseñarse códigos de barras de tres longitudes diferentes para evitar que una única base se disperse por toda la celda de flujo del secuenciador. Las secuencias de código de barras no aleatorias pueden estar presentes en adaptadores como secuencias de 13, 14, y 15 pb; 10, 11, y 12 pb; 11, 12, y 13 pb; 13, 14, y 15 pb; 14, 15, y 16 pb; 15, 16, y 17 pb, y similares. En algunas realizaciones, la secuencia de código de barras más corta puede ser de 8 pb y la secuencia de código de barras más larga puede ser de 100 pb. En la tabla 1 se proporciona un conjunto a modo de ejemplo de códigos de barras que tienen tres longitudes diferentes. Se proporcionan códigos de barras adicionales, por ejemplo, en la solicitud provisional estadounidense n.º 62/348.791, presentada el 10 de junio de 2016, y la solicitud provisional estadounidense n.º 62/364.256, presentada el 19 de julio de 2016.

TABLA 1

	Secuencia de código de barras no aleatoria
SEQ ID NO:1	AATGCCATGGCTT
SEQ ID NO:2	CAGTAGCTCTGAT
SEQ ID NO:3	GACGTATACGCTT
SEQ ID NO:4	TACGGACTCGTAT
SEQ ID NO:5	AACGTTTCGAGTCCT
SEQ ID NO:6	CACGTTACGATGAT
SEQ ID NO:7	GCTCCTAGACGTAT
SEQ ID NO:8	TATCGAGCTAGCCT
SEQ ID NO:9	ACGTAGCTGATCAGT
SEQ ID NO:10	CAGGACTAGCTTACT
SEQ ID NO:11	GCATCGCTAGTAGAT
SEQ ID NO:12	TACGTAGTACGCAGT

5 Cada secuencia del panel que está presente en la muestra de ADNlc se selecciona como diana por una o más sondas de captura descritas en otra parte del presente documento, y se aísla para un análisis posterior.

Secuenciación y análisis

10 Los fragmentos de ADNlc con código de barras aislados de la muestra de líquido del paciente, por ejemplo, muestra de sangre, se amplifican, por ejemplo por PCR, y se capturan usando las sondas híbridas. La captura de los fragmentos con código de barras comprende obtener hebras individuales de ADNlc con código de barras e hibridar el ADNlc con código de barras con diferentes sondas híbridas. Cada una de las diferentes sondas híbridas se hibrida con una secuencia diana de ADNlc con código de barras monocatenario para formar un dúplex diana-sonda híbrida. El dúplex se aísla del ADNlc no hibridado mediante la unión del resto de unión de purificación comprendido en la sonda híbrida con la pareja de unión del resto de purificación correspondiente. Tal como se describe en otra parte del presente documento, la pareja de unión del resto de purificación correspondiente puede inmovilizarse sobre una superficie sólida, por ejemplo, una perla magnética, que facilita la separación del dúplex de captura de las moléculas de ADNlc no hibridadas en disolución. El ADNlc con código de barras del dúplex se libera y se somete a secuenciación usando un instrumento de NGS.

20 En algunas realizaciones, las secuencias capturadas pueden analizarse usando la tecnología de secuenciación por síntesis de Illumina, que usa desoxirribonucleótidos terminadores reversibles fluorescentes. Las lecturas generadas por el proceso de secuenciación se alinean con una secuencia de referencia y se asocian a una secuencia del panel de secuencias somáticas específico para el paciente. El mapeo de las lecturas de secuencia puede lograrse comparando la secuencia de las lecturas con la secuencia del genoma de referencia para determinar la información genética específica y, opcionalmente, el origen cromosómico de la molécula de ácido nucleico secuenciada (por ejemplo, ADN libre de células). Están disponibles varios algoritmos informáticos para alinear secuencias, incluyendo, sin limitación, BLAST (Altschul *et al.*, 1990), BLITZ (MPsrch) (Sturrock y Collins, 1993), FASTA (Person y Lipman, 1988), BOWTIE (Langmead *et al.*, Genome Biology 10:R25.1-R25.10 [2009]) o ELAND (Illumina, Inc., San Diego, California, EE.UU.). En una realización, los datos de secuenciación se procesan mediante análisis de alineación bioinformática para el sistema Genome Analyzer de Illumina, que usa el software de alineación eficiente a gran escala de bases de datos de nucleótidos (ELAND). El software adicional incluye SAMtools (SAMtools, Bioinformatics, 2009, 25(16):2078-9) y el procedimiento de compresión por clasificación de bloques de Burroughs-Wheeler, que implica la clasificación o preprocesamiento de bloques para hacer que la compresión sea más eficiente.

40 La tasa de error en la secuenciación mediante métodos de NGS es de aproximadamente 1 de cada 500 bases, lo que genera numerosos errores de secuenciación. La alta tasa de error se vuelve problemática, especialmente al intentar identificar mutaciones somáticas en mezclas de secuencias de ADN que comprenden sólo una pequeña fracción de especies mutadas. Además, los métodos de NGS suelen utilizar ADN monocatenario como fuente principal de material de secuenciación. Cualquier error durante la etapa de amplificación de la molécula de ADN previa a la secuenciación se perpetúa y se vuelve indistinguible como un error externo dependiente de la tecnología. Se producen errores químicos con una frecuencia de aproximadamente 1000 bases. La combinación de errores de

secuenciación y químicos complica el límite de detección (LOD).

5 Por consiguiente, en algunas realizaciones, se realiza la secuenciación bicatenaria del ADNlc. Tal como se describe en otra parte del presente documento, el ADNlc puede someterse a reparación de extremos y, opcionalmente, puede añadirse una cola de dA, y se ligan adaptadores bicatenarios que comprenden secuencias complementarias a los cebadores de amplificación y secuenciación a los extremos de las moléculas de ADNlc para permitir la secuenciación NGS, por ejemplo, usando una plataforma de Illumina.

10 Dado que cada adaptador contiene un código de barras diferente, cada fragmento de ADN se marca con dos secuencias de código de barras distintas (designadas arbitrariamente como A, con la secuencia de código de barras complementaria A', y B, con la secuencia de código de barras complementaria B'. Véase la figura). Tras el ligamiento de los adaptadores con código de barras a la molécula de ADNlc bc, las hebras con código de barras individuales se amplifican por PCR a partir de los sitios de cebadores asimétricos en las colas de adaptador y se someten a secuenciación de extremos emparejados usando, por ejemplo, una plataforma de Illumina. Se producen dos especies de productos de PCR de pares de acoplamiento a partir de cada fragmento de ADN. Los derivados de una hebra tendrán la secuencia de etiqueta A adyacente a la secuencia de celda de flujo S1 y la secuencia de etiqueta B adyacente a la secuencia de celda de flujo S2. Por tanto, la primera hebra contendrá las secuencias de código de barras $A_{R1}|B_{R2}$ y la segunda hebra contendrá secuencias de código de barras $B_{R1}|A_{R2}$.

20 Las lecturas de secuencia que comparten un conjunto único de etiquetas se agrupan en familias emparejadas con miembros que tienen identificadores de código de barras de hebra en cualquiera de la orientación $A_{R1}|B_{R2}$ o $B_{R1}|A_{R2}$. Cada par de familias refleja la amplificación de un fragmento de ADN bicatenario. La secuenciación bicatenaria permite diferenciar mutaciones somáticas verdaderas, por ejemplo, sustituciones, mutaciones que surgen de daño químico, por ejemplo, durante la preparación de muestras, y los errores de secuenciación, y corregir de ese modo los errores químicos y de secuenciación para maximizar la especificidad del ensayo (figura 9). Sólo las mutaciones verdaderas presentes en ambas hebras de un fragmento de ADN aparecerán en todos los miembros de un par de familias. La fracción tumoral puede calcularse entonces como la proporción de diferentes secuencias de ADNlc, comprendiendo cada una al menos una mutación somática, es decir, secuencias de ADNtc, en relación con el número total de ADNlc diferentes, es decir, ADNtc y sus secuencias normales correspondientes. A diferencia del enfoque monocatenario, el método actual corrige errores de secuenciación aleatorios.

35 En algunas realizaciones, se identifican al menos 10 mutaciones somáticas diferentes, al menos 20 mutaciones somáticas diferentes, al menos 30 mutaciones somáticas diferentes, al menos 40 mutaciones somáticas diferentes, al menos 50 mutaciones somáticas diferentes, al menos 60 mutaciones somáticas diferentes, al menos 70 mutaciones somáticas diferentes, al menos 80 mutaciones somáticas diferentes, al menos 100 mutaciones somáticas diferentes, al menos 150 mutaciones somáticas diferentes, al menos 200 mutaciones somáticas diferentes, al menos 250 mutaciones somáticas diferentes, al menos 300 mutaciones somáticas diferentes, al menos 400 mutaciones somáticas diferentes, al menos 500 mutaciones somáticas diferentes, o más de 500 mutantes diferentes en la muestra de ADNlc del paciente. El número de marcadores de segregación que teóricamente pueden identificarse en una muestra puede ser hasta el número de marcadores de segregación únicos entre los especímenes de ácido nucleico que están sometidos a prueba. Por ejemplo, existen aproximadamente 3 millones de marcadores de segregación entre dos individuos humanos, mientras que el número para una muestra tumoral puede diferir del de una muestra no tumoral coincidente del paciente, puede tener 10, 100, 1000, 10.000, o más marcadores. En algunas realizaciones, los diferentes mutantes se identifican a una profundidad molecular de al menos 1000, al menos 2000, al menos 3000, al menos 4000, al menos 5000, al menos 6000, al menos 7000, al menos 8000, al menos 9000, al menos 10.000 o más. En otras realizaciones, las diferentes mutaciones se identifican a una profundidad de secuenciación de hasta la cantidad total de moléculas de ácido nucleico presentes en la muestra.

50 En algunas realizaciones, las mutaciones son mutaciones de la línea germinal. En algunas realizaciones, las mutaciones son mutaciones somáticas.

Ejemplos

55 Ejemplo 1

Simulación por ordenador de la determinación de la fracción tumoral

60 Este ejemplo es una simulación por ordenador de un experimento diseñado para determinar la fracción tumoral de acuerdo con los métodos proporcionados en el presente documento.

En este ejemplo, se supone que se interrogó un panel de 100 sitios somáticos a una profundidad de 5000 veces en una muestra virtual, es decir, *in silico*, que tenía una fracción tumoral de $1e^{-4}$. Se realizaron 100.000 repeticiones del proceso, y se puntuó el número de moléculas tumorales en cada caso.

65 Al seleccionar como diana 100 sitios somáticos y secuenciarlos a una profundidad de $5e^3$, dada una fracción tumoral

de $1e^{-4}$, se esperaría observar aproximadamente 50 moléculas tumorales en un fondo de 500.000 moléculas normales. Cada repetición comienza con cero lecturas tumorales. Se puntúa entonces cada sitio como que produce cierto número de lecturas tumorales (potencialmente cero o más) mediante el uso de una extracción aleatoria a partir de una distribución binomial con una probabilidad de éxito igual a $1e^{-4}$ y un número de ensayos igual a 5.000.

Por ejemplo, la primera ejecución de la simulación comenzaría con cero lecturas tumorales. Luego, para el sitio 1 se realiza una extracción aleatoria a partir de una distribución binomial con los criterios anteriores, y esta cantidad se añade al recuento tumoral. Para el sitio 2, se realiza entonces otra extracción aleatoria de la distribución binomial, y esa cantidad también se añade al recuento tumoral. Esto se repite para cada sitio hasta que todos los sitios se "sometan a prueba". Este proceso se repite 100.000 veces. Se muestran los resultados en la figura 2, que es un histograma de lecturas tumorales por permutación a lo largo de las 100.000 permutaciones.

La sensibilidad del método determinada mediante esta simulación muestra que se puede esperar la detección de al menos 20 moléculas tumorales, es decir, moléculas que comprenden sitios somáticos, en un fondo total de 500.000 moléculas normales, es decir, sitios que no se detectaron que contenían una mutación somática del panel. La simulación que se muestra en la figura 2 se basa en la detección de mutaciones somáticas en ADN libre de células de una muestra de plasma de una extracción de sangre de aproximadamente 20 ml, que se estima que contiene entre 3000 y 5000 copias del genoma. Al seleccionar como diana 100 sitios somáticos y secuenciar los sitios a una profundidad de $5e^3$, dada una fracción tumoral de $1e^{-4}$, se esperaría observar aproximadamente 50 moléculas tumorales en un fondo de 500.000 moléculas normales. Dadas estas condiciones, la distribución del número de moléculas tumorales observadas en 100.000 simulaciones del proceso, es decir el experimento, detectó al menos 20 moléculas tumorales en todas las simulaciones. Por tanto, el método ofrece una sensibilidad extremadamente alta.

La figura 3 ilustra este mismo procedimiento de simulación para determinar el número mínimo de lecturas que se observarán en diferentes condiciones, por ejemplo, diferente número de sitios que se someten a seguimiento, y diferentes niveles de fracción tumoral. El gráfico muestra diferentes líneas que representan el diferente número de sitios sometidos a seguimiento, y el número mínimo de lecturas observadas en una fracción tumoral dada. El valor mínimo para el número de lecturas tumorales es de 20 para cada línea del gráfico. Sin embargo, diferentes condiciones producirán gráficos diferentes donde el valor mínimo puede ser mayor (o menor) de 20; un histograma generado por una simulación, con parámetros especificados (fracción tumoral, número de sitios), tendrá un aspecto similar a los histogramas de la figura 3, pero desplazado a lo largo de cualquier eje. Por ejemplo, al realizar un seguimiento de 1000 sitios con una fracción tumoral de 1 de cada 10.000, se observan aproximadamente 600 sitios tumorales. Extrapolando, si se realiza un seguimiento de $3e^6$ sitios (es decir, el número de diferencias genéticas entre dos humanos), entonces se espera que el ensayo pueda detectar la presencia de una muestra humana entre ADN al azar si el humano está presente en una proporción de $3e^{-8}$ (que es $1/(n.^{\circ} \text{ de sitios} * 100)$). Cuanto mayor sea el número de sitios, menor será la fracción tumoral necesaria para detectar la presencia de mutaciones somáticas, es decir, oncogénicas. Los métodos actuales usados por otros investigadores suelen detectar mutaciones somáticas a fracciones tumorales de 1 de cada 100 o 1 de cada 1000. El método proporcionado en el presente documento detecta mutaciones somáticas a fracciones tumorales que son de 10 a 1000 veces menores que las fracciones requeridas por métodos de otros investigadores.

Aunque este ejemplo se ha redactado para un tumor, es igualmente aplicable a una mezcla de dos poblaciones de secuencias de ácidos nucleicos de más de una fuente, por ejemplo, dos humanos diferentes. Por tanto, es aplicable para distinguir entre ácidos nucleicos fetales y maternos, distinguir entre un huésped y un agente infeccioso, distinguir entre cultivos, determinar la presencia de la secuencia de ácidos nucleicos de un único individuo en una mezcla de secuencias de ácidos nucleicos.

Ejemplo 2

Detección de mutaciones somáticas en mezclas de secuencias homocigotas y heterocigotas derivadas de una mezcla de dos muestras

El propósito de este experimento fue evaluar la capacidad del método proporcionado para detectar mezclas muy pequeñas de ADN, del orden de 1 molécula foránea por 1 millón de moléculas "de fondo", mezclando ADN de dos muestras de donantes en proporciones conocidas y secuenciándolo tal como se describe.

Método

1. Extraer el ADN genómico de dos muestras de donantes. Preparar una biblioteca para cada donante que contenga códigos de barras moleculares que permitan la secuenciación dúplex, y secuencias de código de barras de muestra que indiquen a partir de qué muestra se originaron las lecturas de secuenciación posteriores.
2. Someter a secuenciación de genoma completo las dos muestras.
3. Seleccionar ~10.000 sitios que sean homocigotos en la muestra 1 y heterocigotos en la muestra 2. La muestra 2 se consideró la muestra de "cáncer".

4. Diseñar y sintetizar sondas de captura híbridas biotiniladas que seleccionan como diana esos 10.000 sitios para enriquecer en sitios que contengan marcadores de segregación.

5. Realizar mezclas de ADN de la muestra 2 con ADN de la muestra 1 en una variedad de proporciones de mezcla:

a. $1e^{-2}$

b. $1e^{-3}$

c. $1e^{-4}$

d. $1e^{-5}$

e. $1e^{-6}$

f. 0 (control negativo)

6. Secuenciar muestras mezcladas usando un flujo de trabajo de secuenciación dúplex y enriquecerlas usando sondas de captura híbridas que se diseñaron en la etapa 4.

7. Analizar datos de secuencia:

a. Alinear con el genoma humano

b. Combinar códigos de barras moleculares en lecturas/moléculas

c. Eliminar por filtrado los sitios con ruido de manera heurística, teniendo en cuenta los códigos de barras moleculares y las lecturas observadas en la muestra de control negativo.

d. Hacer un recuento/contar las lecturas de cáncer observadas y las lecturas sin cáncer observadas

e. Calcular la fracción tumoral estimada.

8. Comparar las estimaciones con las mezclas esperadas.

Se muestran los resultados en las figuras 4, 5 y 6. Los resultados proporcionados en la figura 4 muestran que se obtuvo una detección de tan sólo 5 moléculas tumorales por cada 1.000.000 de moléculas secuenciadas. En el gráfico, cada punto es la fracción tumoral que se esperaba observar (eje X), y la fracción tumoral observada (eje Y). El punto de datos $1e^{-9}$ representa la fracción nula. Cabe destacar que las barras de error en las fracciones tumorales, por ejemplo, en $5e^{-6}$, no se superponen con las barras de error de la fracción nula. Por tanto, el método detecta y cuantifica con precisión el tumor en una mezcla que tiene una fracción de tan sólo 5 moléculas por cada 1.000.000 de moléculas al secuenciar 10.000 sitios en el panel. **Nota:** La fracción tumoral esperada es de $0,5 * \text{proporción de mezcla experimental}$, debido a que se eligieron principalmente sitios heterocigotos en la muestra de prueba inicial (etapa 3).

Se logró una detección estadísticamente significativa para mezclas de hasta 5 partes por millón, en comparación con un control negativo. El poder estadístico para detectar mezclas está impulsado por la profundidad de secuenciación de plasma y del número de sitios sometidos a seguimiento.

La figura 5 muestra en (A) el gráfico proporcionado en la figura 4, y el número de lecturas normales (B) y lecturas de cáncer (C) obtenidas para las condiciones descritas para el gráfico anterior.

La figura 6 muestra una tabla con los resultados obtenidos para las condiciones 5:1.000.000 y nula. Se detectaron 18 lecturas de cáncer de un total de 3.540.529 lecturas normales, lo que facilita una proporción estimada de cáncer de $18/3.540.529 = 5,0840e^{-6}$, con un intervalo de confianza binomial del 95 % en torno a la estimación. La significación estadística radica en que la condición positiva y la condición negativa no se superponen.

Ejemplo 3

MUESTRAS CLÍNICAS

Este ejemplo divulga la prueba en tres muestras de pacientes diferentes.

Se obtuvieron muestras de una empresa de adquisición de muestras comercial. Las muestras fueron las siguientes:

Tabla 2

Paciente	Tipo y estadio	Muestras disponibles	Mutaciones somáticas identificadas
1067	De mama en estadio 3c	Tumor FFPE, tumor fresco congelado, normal, plasma Pre-Op (preoperatorio)	562
1071	Carcinoma de pulmón de células escamosas en estadio 3a	Tumor FFPE, tumor fresco congelado, normal, plasma Pre-Op	3575
2956	Adenocarcinoma de pulmón en estadio 1a	Tumor FFPE, normal, plasma Pre-Op, plasma Post-Op (posoperatorio)	47

- 5 Se caracterizaron las muestras tal como se describe en el presente documento. Se sometieron muestras tumorales y normales a secuenciación de genoma completo y se identificaron las mutaciones somáticas. A continuación, se diseñaron sondas de captura híbridas que seleccionaron como diana todas las mutaciones somáticas identificadas para la muestra respectiva. Por ejemplo, el paciente 1067 tenía 562 sondas diseñadas para cada una de las 562 mutaciones somáticas identificadas. Todas las muestras (tumorales, normales, ADNlc de plasma) de cada paciente
- 10 volvieron a secuenciarse usando las sondas diseñadas y un flujo de trabajo de secuenciación dúplex para eliminar el ruido biológico y sistémico introducido en el flujo de trabajo. Se muestran los resultados en las figuras 7A-C. Tal como puede observarse, cada uno de los cánceres en estadio 3 fue detectable en el tumor y plasma Pre-Op, lo que indica que es posible diseñar sondas capaces de detectar las mutaciones en el ácido nucleico.
- 15 La muestra en estadio 1 (muestra 2956) no proporcionó una señal en el plasma Pre-Op, lo que indica que el nivel de ADNlc tumoral estaba por debajo del límite de detección. Esto indicó que había marcadores insuficientes como para permitir la detección del cáncer en esta muestra de plasma. Sin embargo, la identificación de más marcadores de segregación puede proporcionar una señal detectable.

20

REIVINDICACIONES

1. Método de preparación de una fracción enriquecida en ácido desoxirribonucleico (ADN) de un paciente con cáncer, comprendiendo el método:
 - (a) extraer ADN de una muestra de líquido de plasma sanguíneo o suero sanguíneo del paciente con cáncer para obtener fragmentos de ADN libre de células (ADNlc);
 - (b) producir, a partir de los fragmentos de ADNlc extraídos de (a), una fracción de ADN con fragmentos de ADN tumoral circulante (ADNtc), en el que la fracción de ADN con fragmentos de ADNtc se produce mediante
 - (i) poner en contacto los fragmentos de ADNlc de (a) con una pluralidad de oligonucleótidos, en el que cada oligonucleótido comprende una secuencia de ácido nucleico que puede hibridarse a un fragmento de ADN que comprende una de un subconjunto de un conjunto de mutaciones somáticas específicas de tumor, en el que el conjunto de mutaciones somáticas específicas de tumor se determina mediante:
 - (A) secuenciar ADN genómico de una muestra tumoral obtenida del paciente con cáncer y secuenciar ADN genómico de una muestra no tumoral obtenida del paciente con cáncer, obteniéndose de ese modo secuencias de ADN genómico de la muestra tumoral y secuencias de ADN genómico de la muestra no tumoral;
 - (B) alinear las secuencias de ADN genómico de la muestra tumoral con un genoma humano de referencia que no procede del paciente y alinear las secuencias de ADN genómico de la muestra no tumoral con el genoma de referencia que no procede del paciente, obteniéndose de ese modo un conjunto de secuencias de mutación somática específicas de tumor que están presentes en las secuencias de ADN genómico de la muestra tumoral pero no están presentes en las secuencias de ADN genómico de la muestra no tumoral, y
 - (ii) enriquecer selectivamente los fragmentos de ADNlc extraídos de (a) con la pluralidad de oligonucleótidos,
- en el que la fracción de ADN después de (b) está enriquecida con fragmentos de ADN que comprenden una o más del subconjunto de mutaciones somáticas específicas de tumor.
2. Método según la reivindicación 1, que comprende además analizar los fragmentos de ADN enriquecidos en (b).
 3. Método según la reivindicación 2, en el que analizar los fragmentos de ADN en la fracción producida en (c) comprende secuenciar uno o más de los fragmentos de ADN.
 4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que el método comprende una sensibilidad de detección de aproximadamente 20 a aproximadamente 50 fragmentos de ADNtc que comprenden una o más del subconjunto del conjunto de mutaciones somáticas en la muestra de líquido por un fondo total de aproximadamente 500.000 fragmentos de ADNlc.
 5. Método según una cualquiera de las reivindicaciones 1 a 4, en el que la pluralidad de oligonucleótidos puede detectar al menos 10 mutaciones somáticas específicas de tumor diferentes.
 6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que la extracción de ADN de la muestra de líquido de plasma sanguíneo o suero sanguíneo del paciente con cáncer para obtener los fragmentos de ADN libre de células (ADNlc) comprende separar el ADNlc de la muestra de plasma sanguíneo o suero sanguíneo.
 7. Método según una cualquiera de las reivindicaciones 1 a 6, que comprende además determinar una fracción de los fragmentos de ADNlc que son fragmentos de ADNtc.
 8. Método según la reivindicación 1, en el que se sabe que las mutaciones somáticas están asociadas a uno cualquiera de una variedad de tumores sólidos.
 9. Método para preparar una fracción enriquecida en ácido desoxirribonucleico (ADN) de un paciente con cáncer, que comprende:
 - (a) secuenciar ADN genómico de una muestra tumoral obtenida del paciente con cáncer y secuenciar ADN genómico de una muestra no tumoral obtenida del paciente con cáncer, obteniéndose de ese modo

secuencias de ADN genómico de la muestra tumoral y secuencias de ADN genómico de la muestra no tumoral;

5 (b) alinear las secuencias de ADN genómico de la muestra tumoral con un genoma humano de referencia que no procede del paciente y alinear las secuencias de ADN genómico de la muestra no tumoral con el genoma de referencia que no procede del paciente, obteniéndose de ese modo un conjunto de secuencias de mutación somática específicas de tumor que están presentes en las secuencias de ADN genómico de la muestra tumoral pero no están presentes en las secuencias de ADN genómico de la muestra no tumoral;

10 (c) usar una muestra de líquido obtenida del paciente con cáncer, en el que la muestra de líquido contiene ADN;

15 (d) poner en contacto el ADN de la muestra de líquido con una pluralidad de sondas de oligonucleótidos, en el que cada sonda de oligonucleótidos individual en la pluralidad de sondas de oligonucleótidos comprende una secuencia de ácido nucleico que puede hibridarse a un fragmento de ADN libre de células (ADNlc) que comprende una de un subconjunto del conjunto de secuencias de mutación somática específicas de tumor; y

20 (e) enriquecer fragmentos de ADN que se hibridan a una o más de la pluralidad de sondas de oligonucleótidos, preparándose de ese modo una fracción de ADN que está enriquecida para fragmentos de ADN que comprenden el subconjunto del conjunto de mutaciones somáticas específicas de tumor.

10. Método según la reivindicación 9 que comprende además analizar la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor, en el que analizar comprende secuenciar uno o más de los fragmentos de ADN en la fracción.

11. Método según la reivindicación 10, en el que la detección de una secuencia de uno o más de los fragmentos de ADN indica una recidiva del cáncer del paciente.

30 12. Método según una cualquiera de las reivindicaciones 9 a 11, en el que la muestra de líquido es plasma sanguíneo o suero sanguíneo libre de células, y en el que el ADN en la muestra de líquido es ADNlc que comprende tanto ADN tumoral circulante (ADNtc) como ADNlc que no deriva de un tumor.

35 13. Método según la reivindicación 12, que comprende además:
determinar una cantidad total de ADNlc en la muestra de líquido;

40 determinar una cantidad total de ADNtc en la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor; y

comparar la cantidad total de ADNtc en la fracción de ADN enriquecida para fragmentos de ADN que comprenden una o más mutaciones somáticas específicas de tumor con la cantidad total de ADNlc en la muestra de líquido para determinar una fracción tumoral para el paciente;

45 en el que la cantidad de ADNtc refleja la carga tumoral del paciente.

14. Método según una cualquiera de las reivindicaciones 9 a 13, en el que las sondas de oligonucleótidos pueden hibridarse a fragmentos de ADNlc que comprenden colectivamente al menos 10 mutaciones somáticas específicas de tumor diferentes.

50 15. Método según la reivindicación 9, en el que se sabe que las mutaciones somáticas están asociadas a uno cualquiera de una variedad de tumores sólidos.

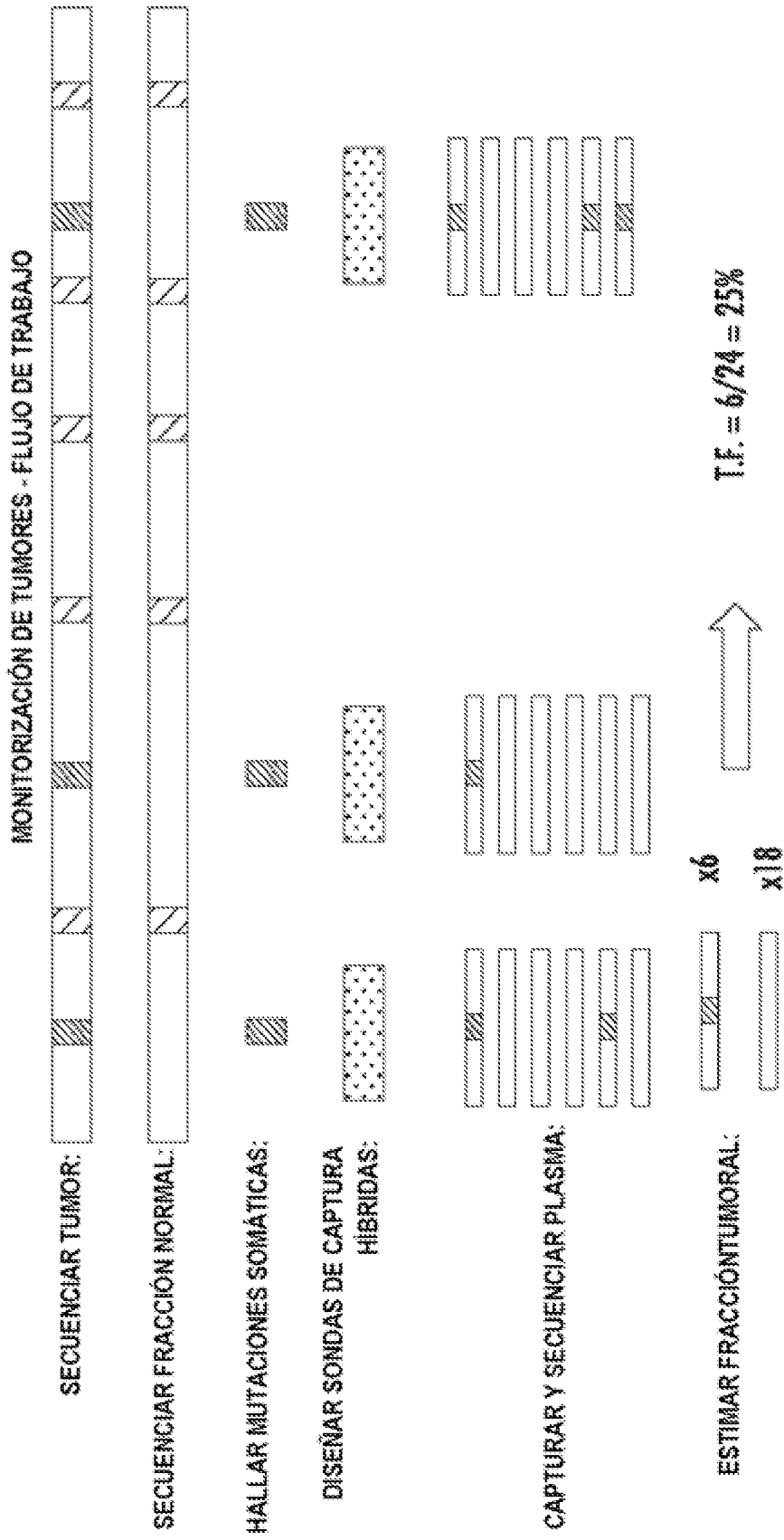


FIG. 7

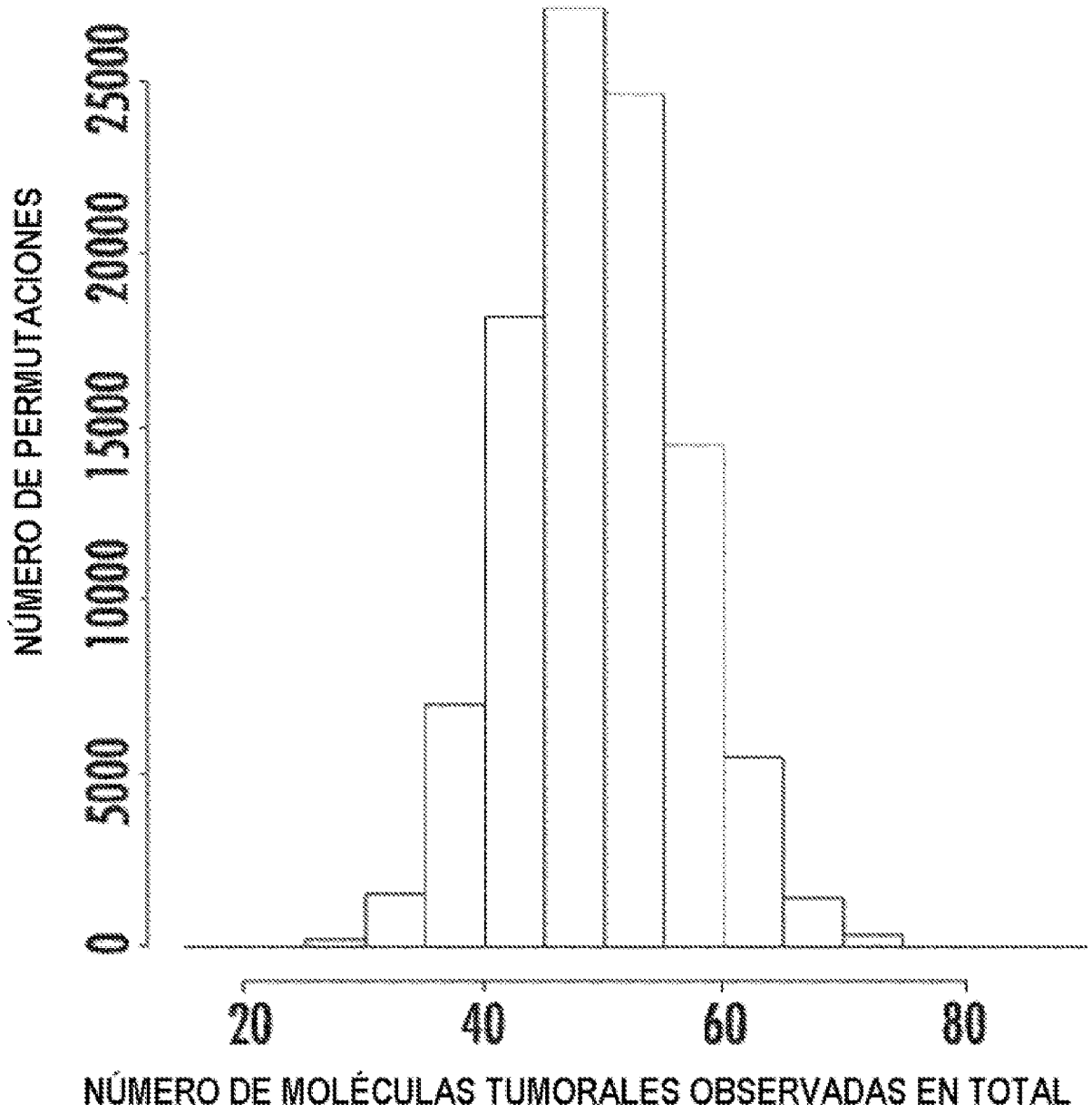


FIG. 2

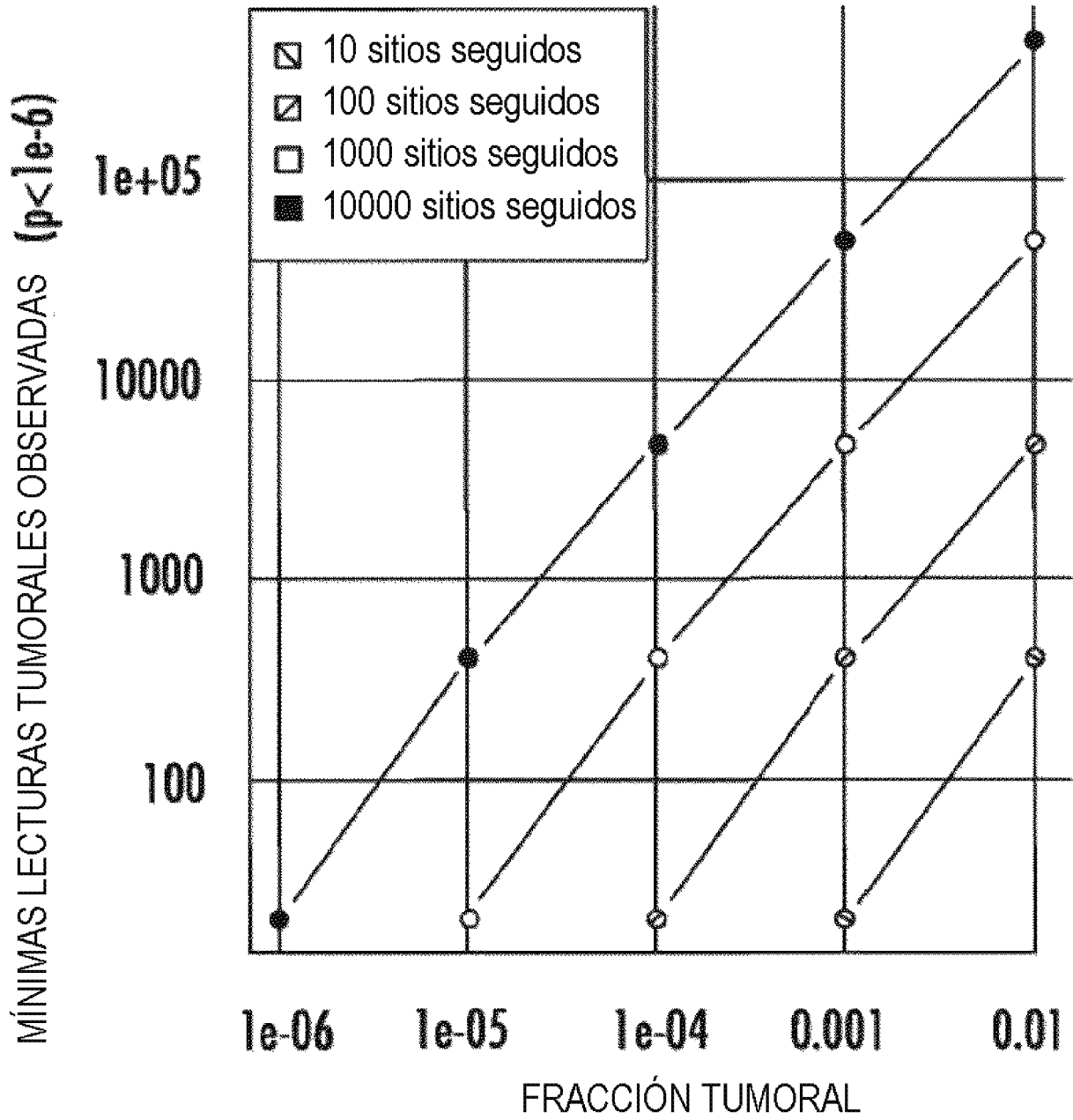


FIG. 3

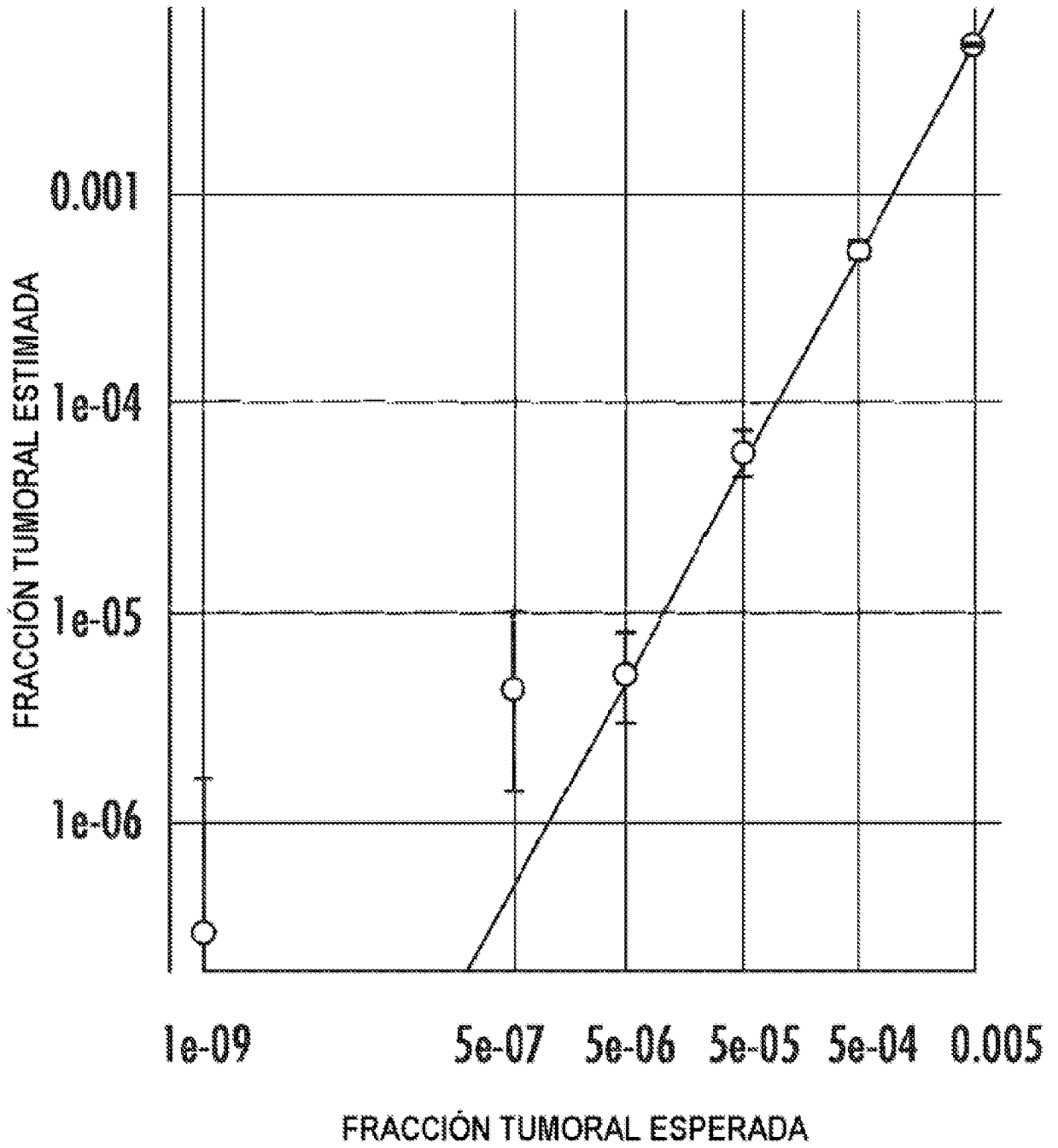


FIG. 4

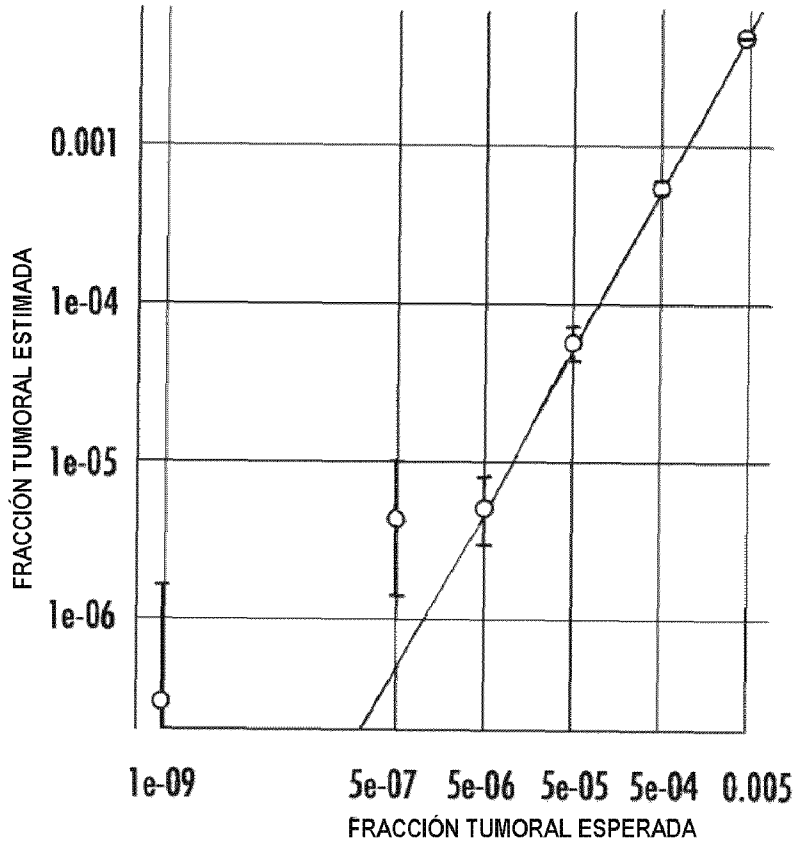


FIG. 5A

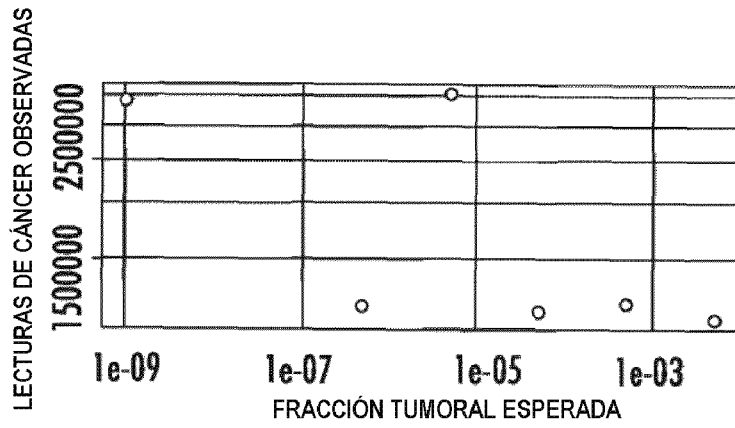


FIG. 5B

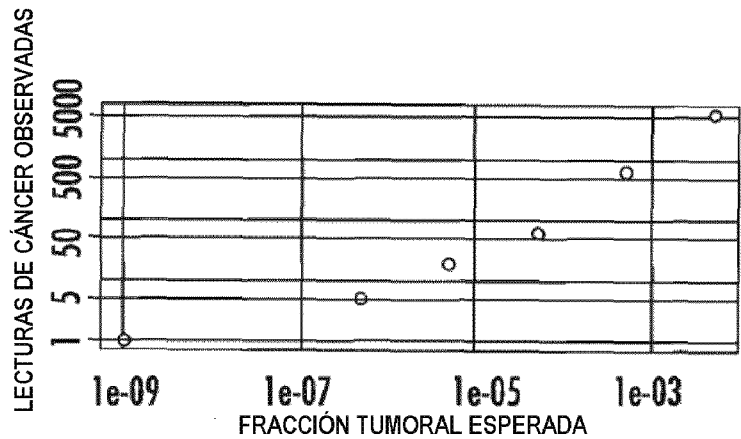


FIG. 5C

condición	normal		mezcla	inferior	superior
mix_0_rep1	3406585	1	2.9355e-07	7.4320e-09	1.6355e-06
mix_0_rep2	3110983	1	3.2144e-07	8.1382e-09	1.7910e-06
mix_1e5_rep1	3540529	18	5.0840e-06	3.0131e-06	8.0348e-06
mix_1e5_rep2	3444781	30	8.7088e-06	5.8758e-06	1.2432e-05

FIG. 6

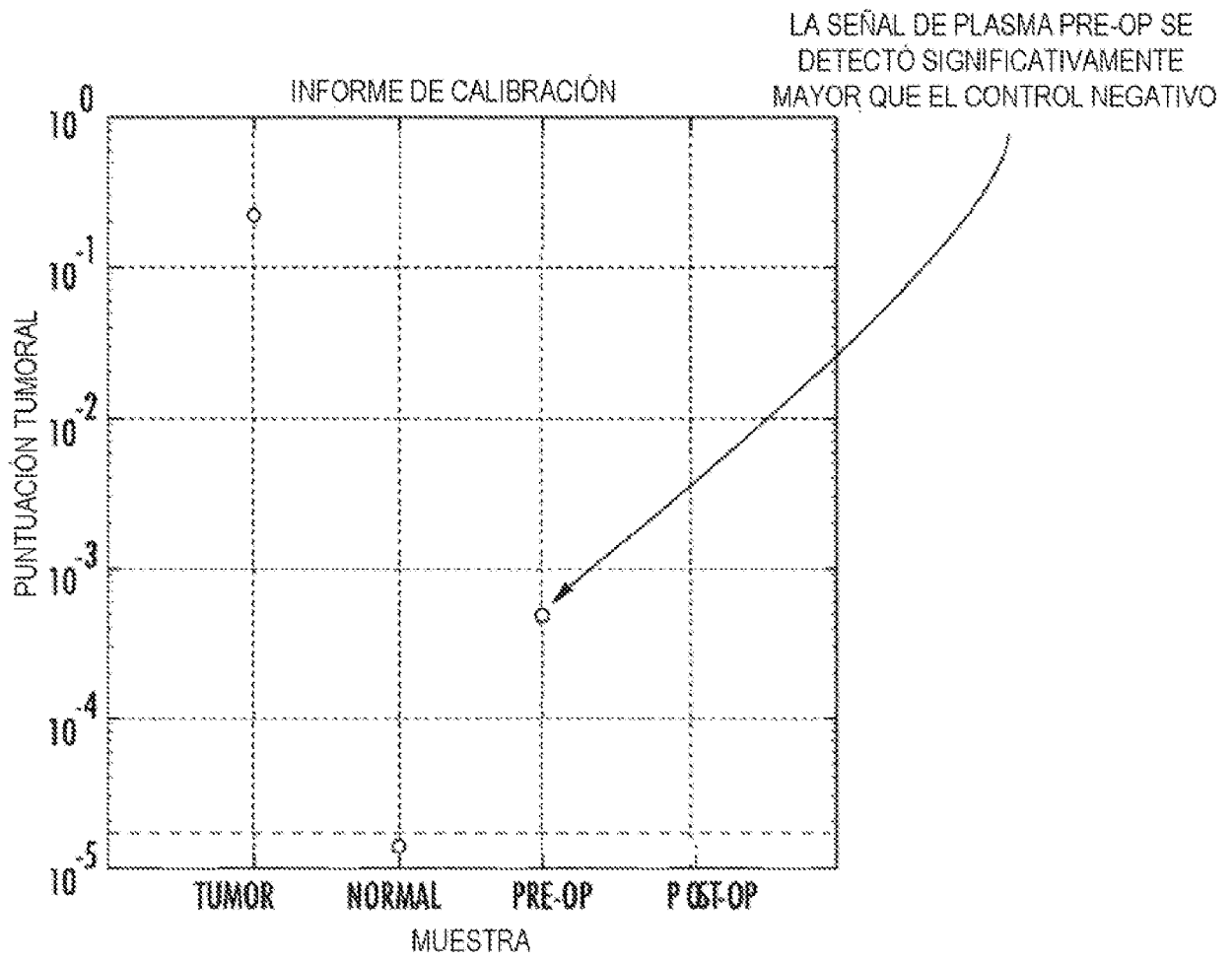


FIG. 7A

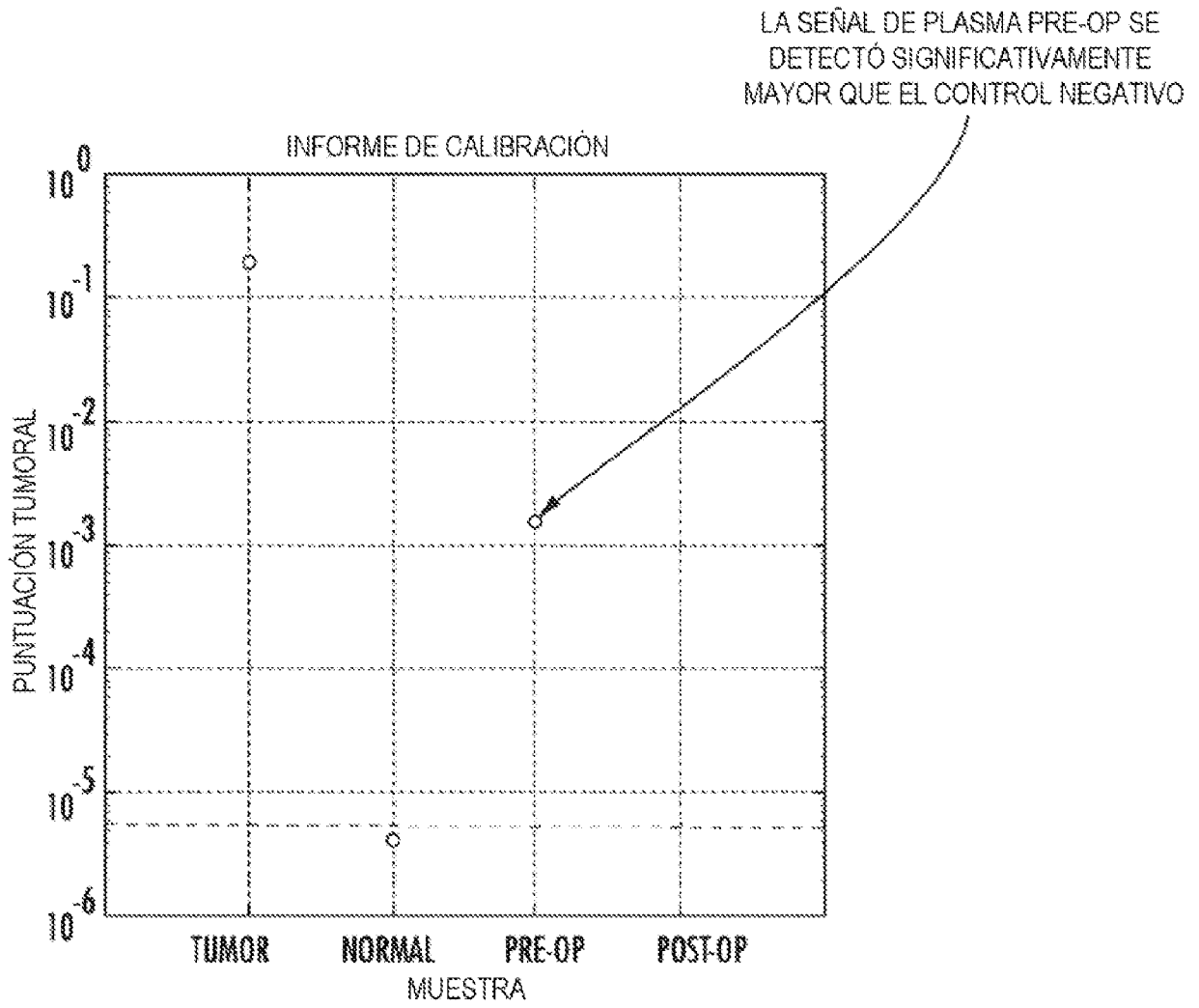


FIG. 7B

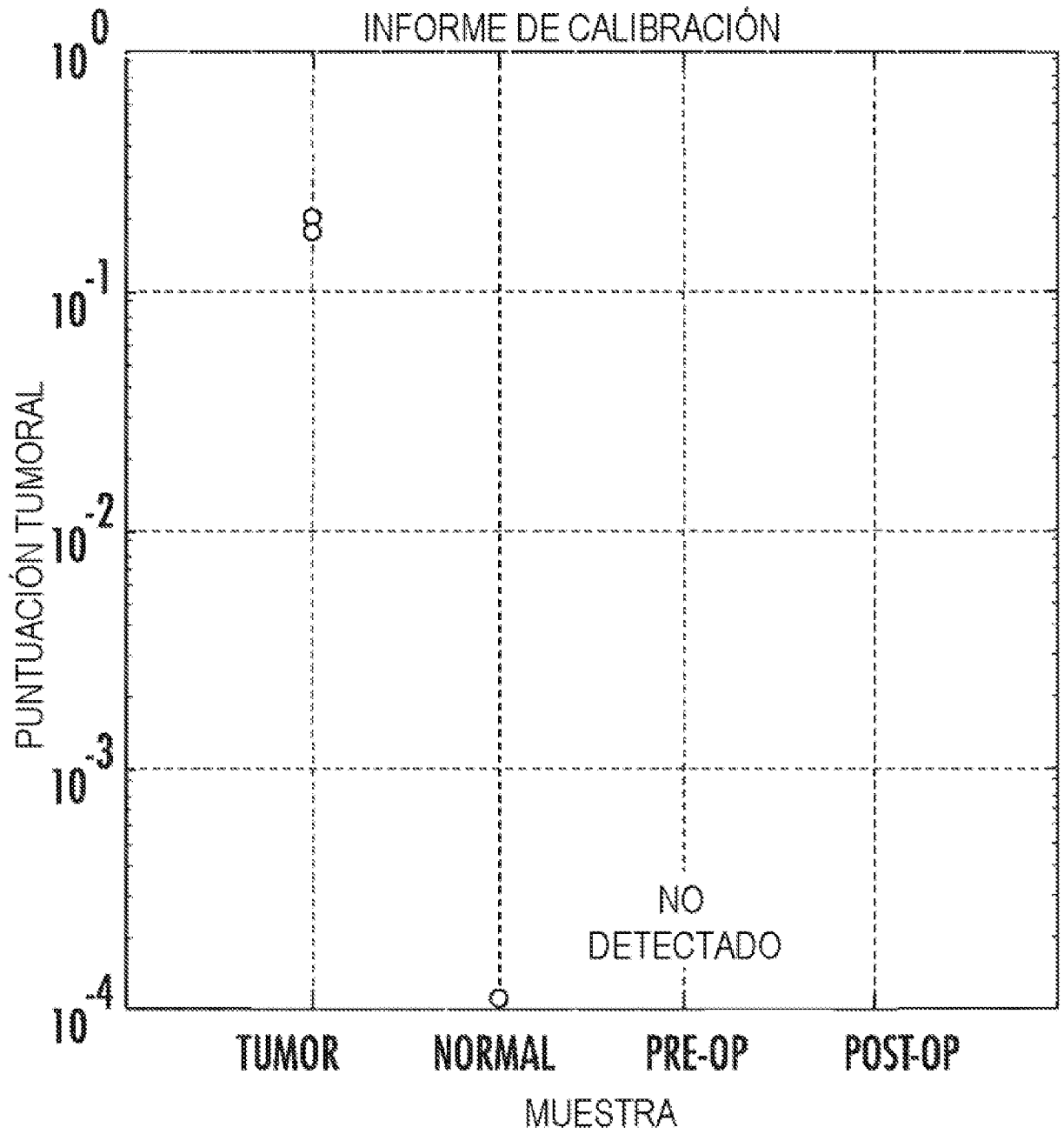


FIG. 7C

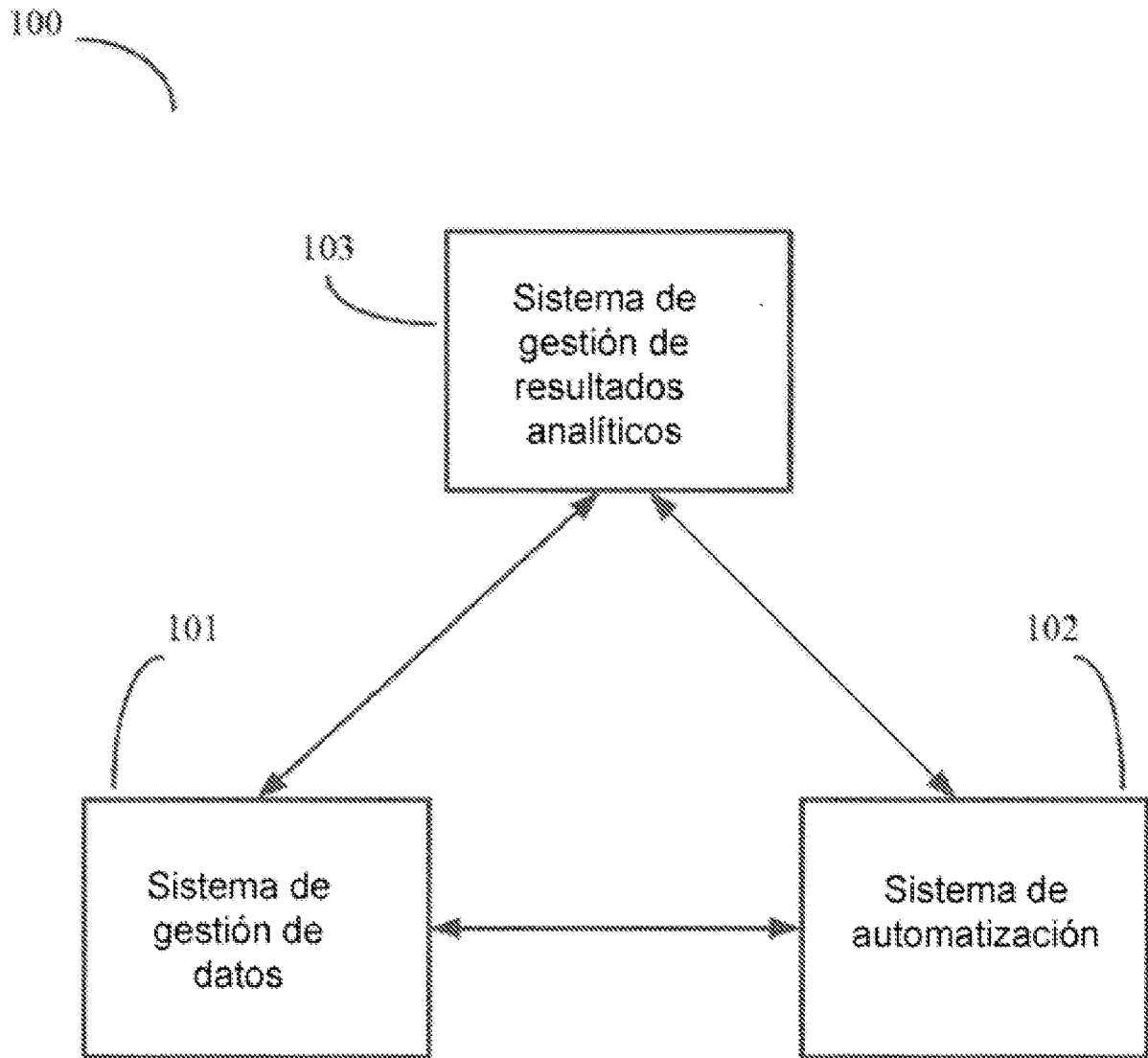


FIG. 8

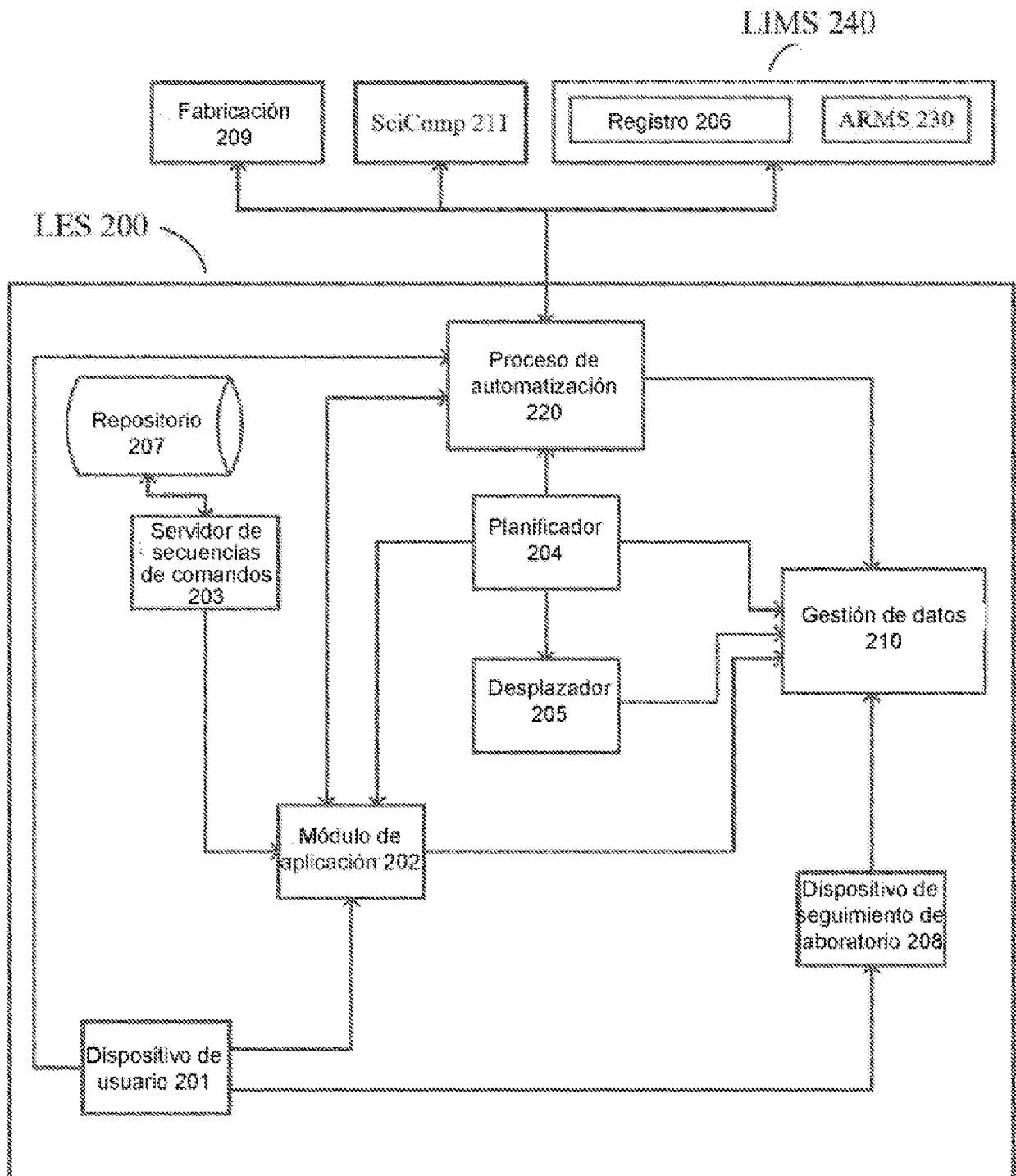


FIG. 9

LIMS 300

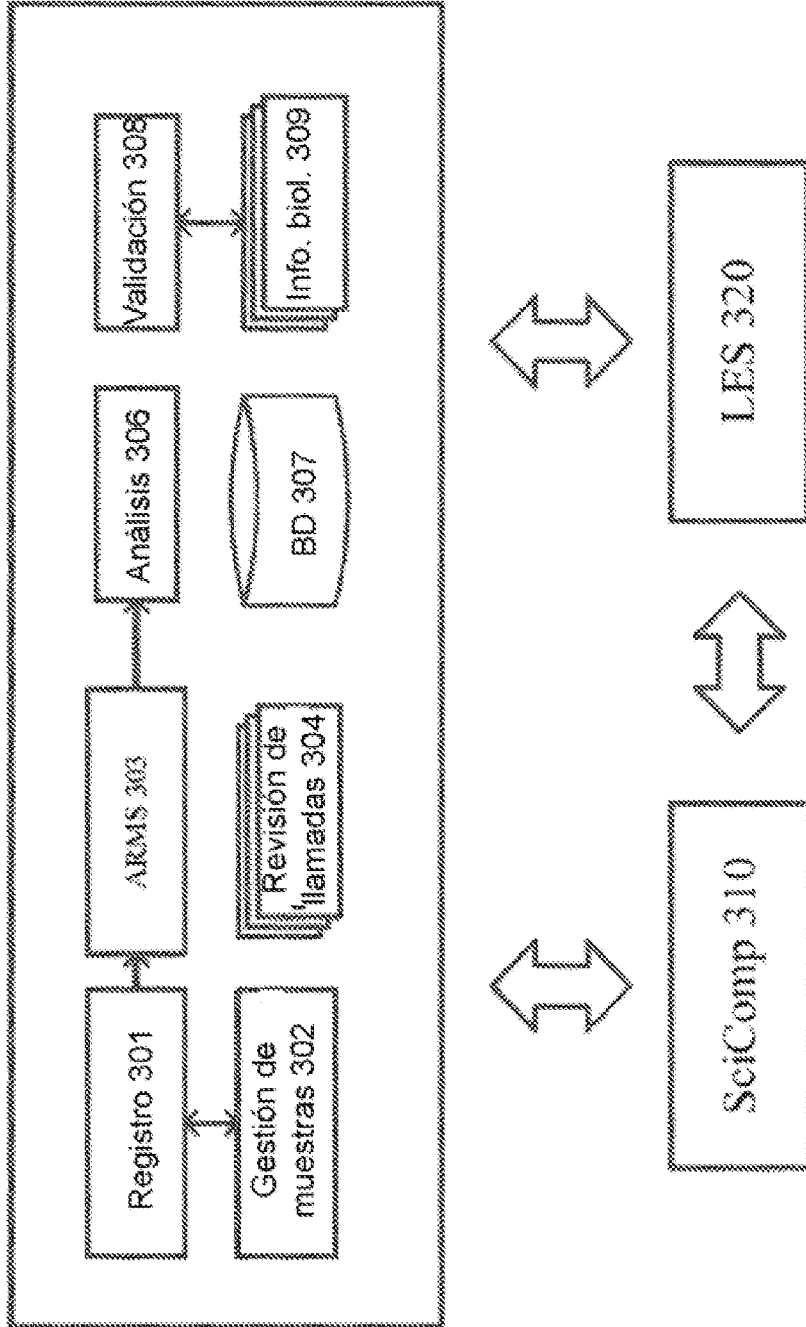


FIG. 10

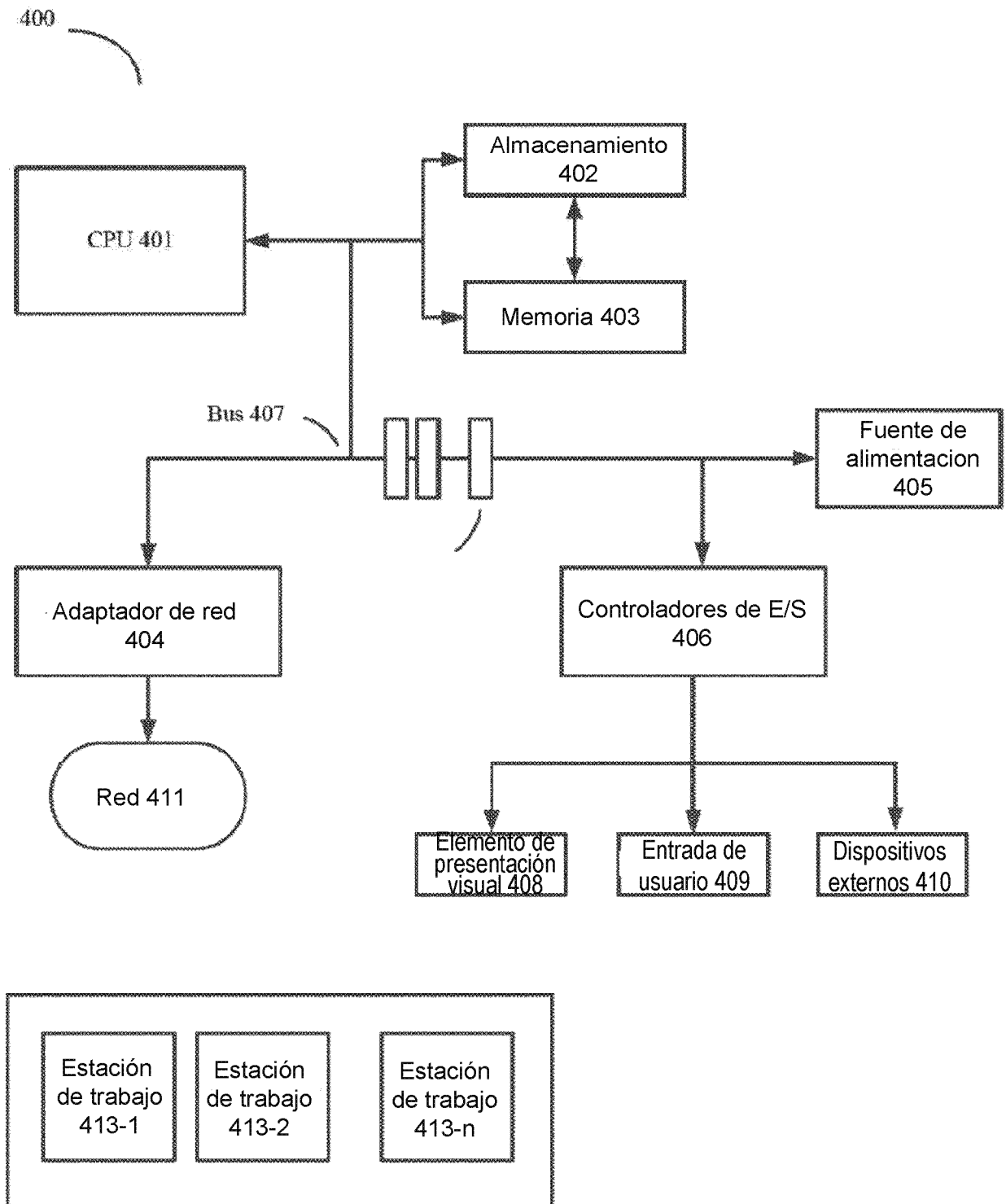


FIG. 11

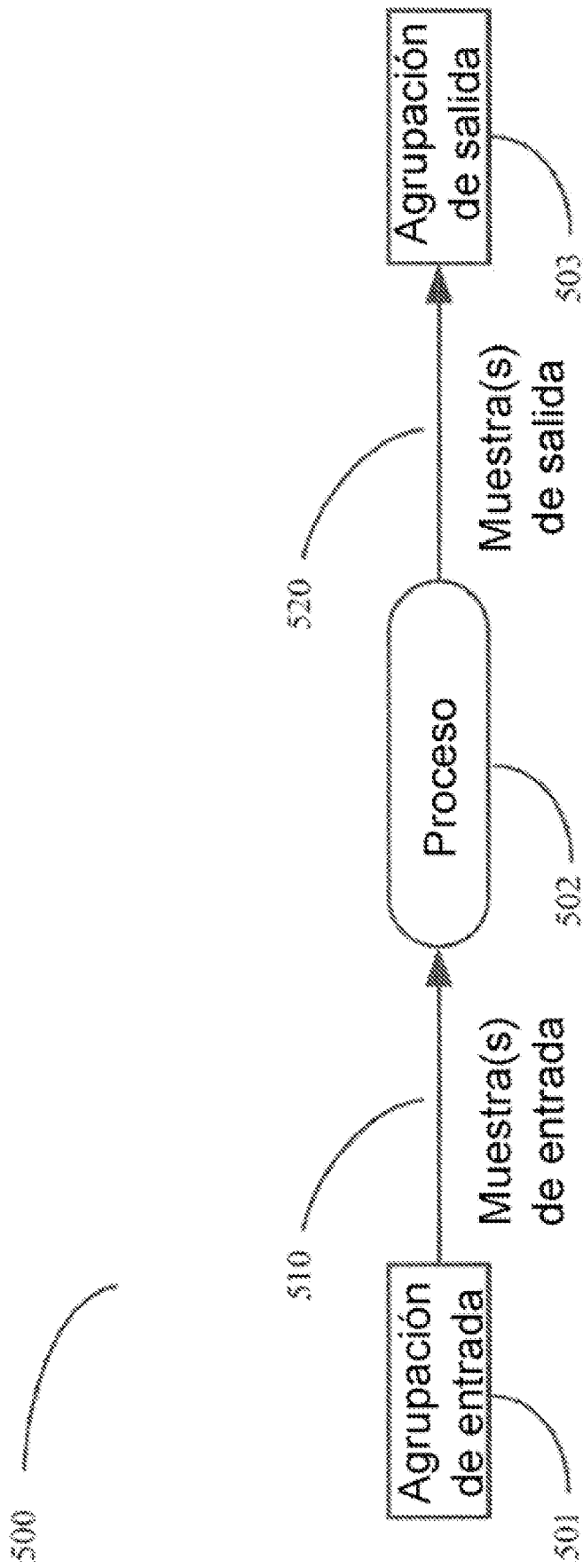


FIG. 12

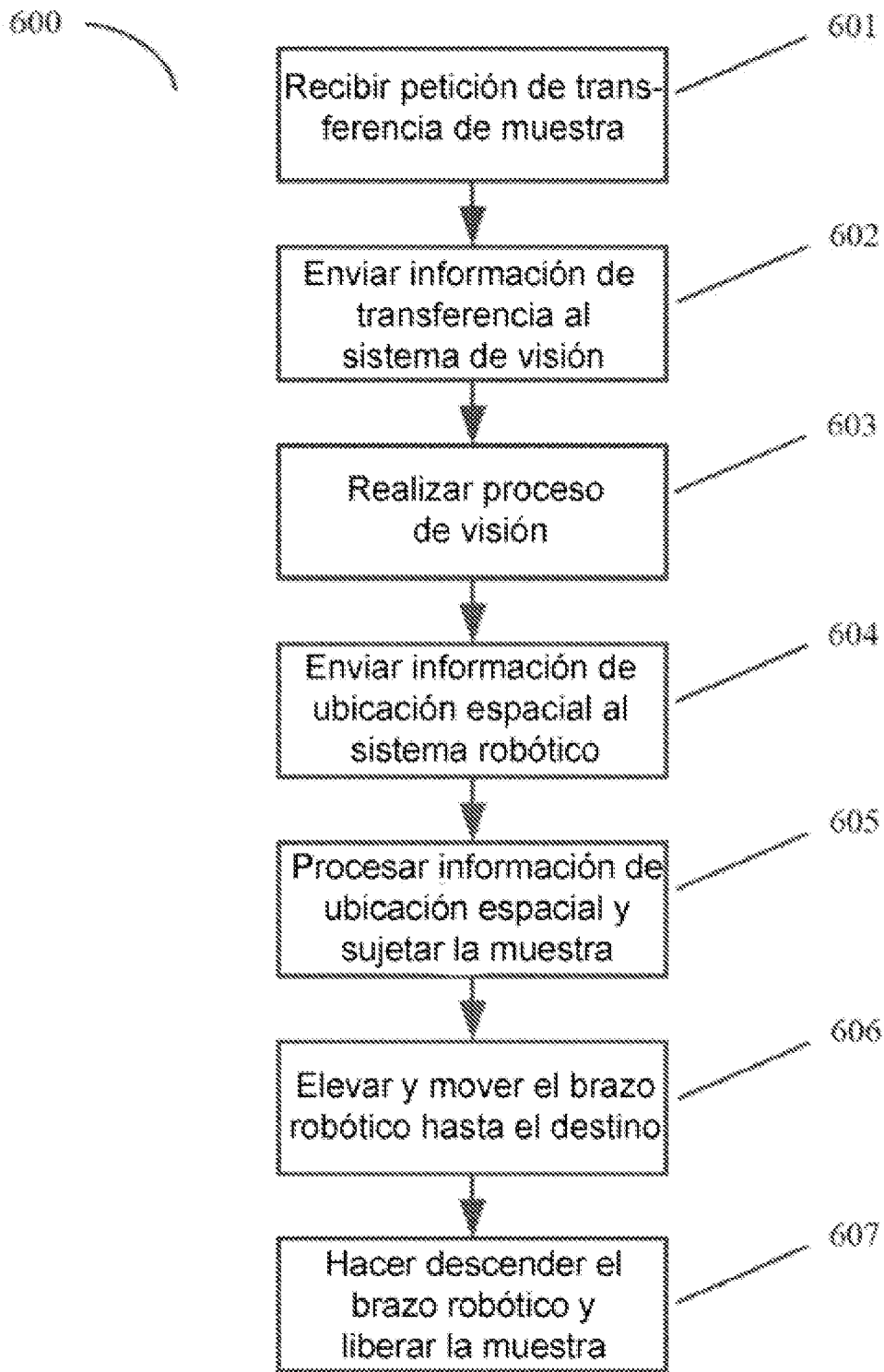


FIG. 13

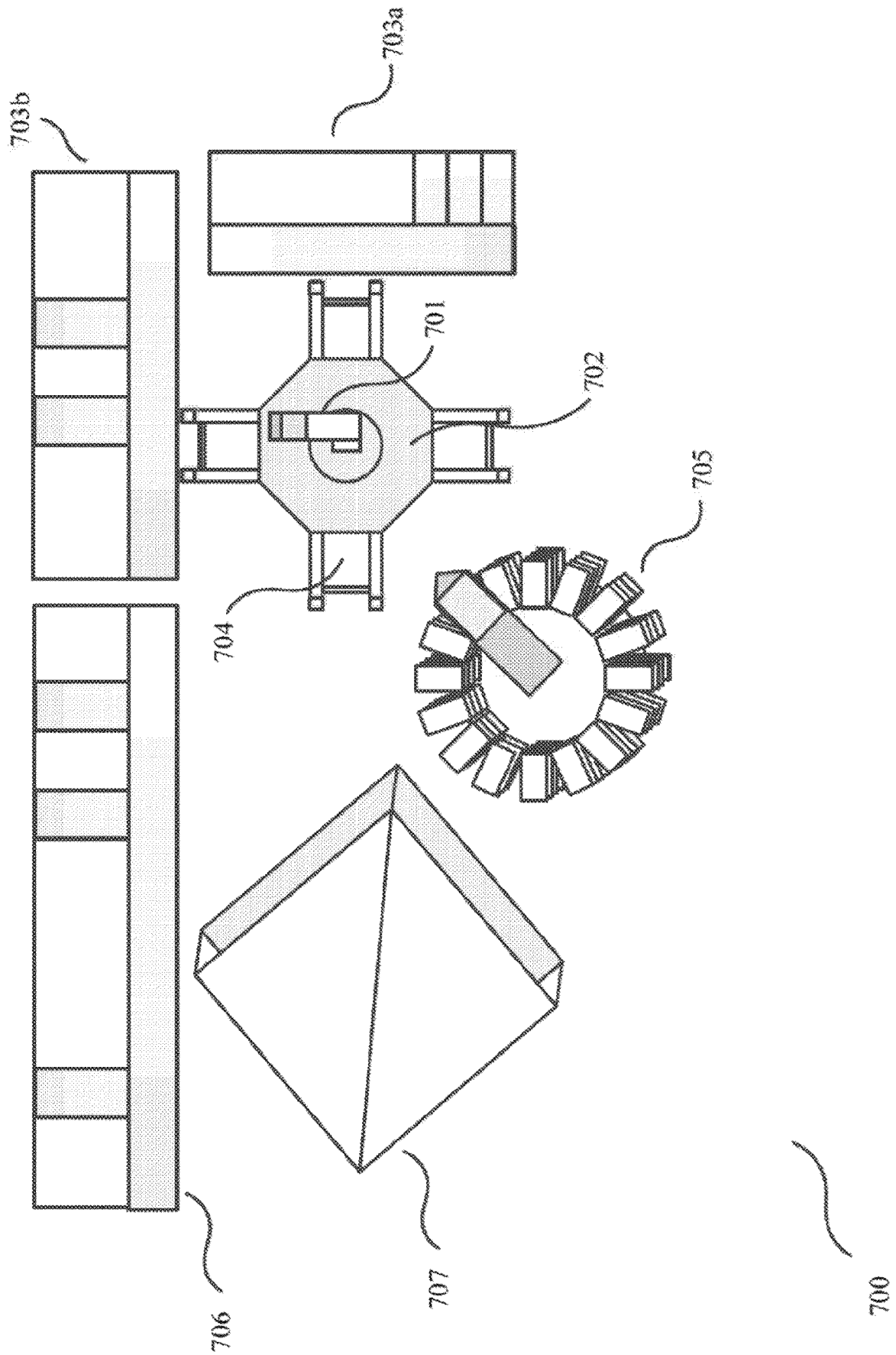
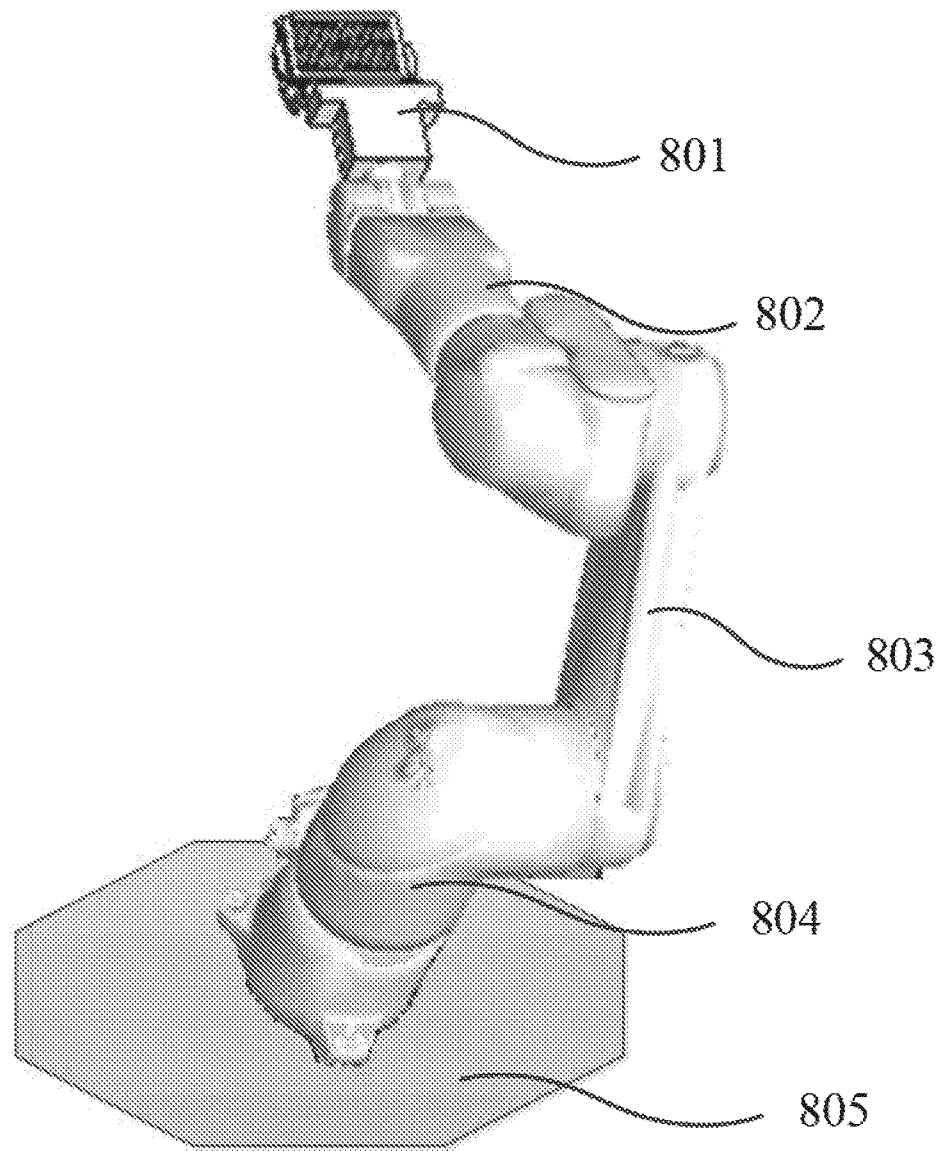


FIG. 14



800

FIG. 15