

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 622**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

C12Q 1/6869 (2008.01)

G16B 20/10 (2009.01)

G16B 40/30 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.01.2016 PCT/CN2016/070785**

87 Fecha y número de publicación internacional: **21.07.2016 WO16112850**

96 Fecha de presentación y número de la solicitud europea: **13.01.2016 E 16737075 (8)**

97 Fecha y número de publicación de la concesión europea: **24.07.2024 EP 3245299**

54 Título: **Uso de anomalías de tamaño y número en el ADN plasmático para detectar un cáncer**

30 Prioridad:

13.01.2015 US 201562102867 P
03.02.2015 US 201562111534 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.11.2024

73 Titular/es:

THE CHINESE UNIVERSITY OF HONG KONG
(100.0%)
Office of Research and Knowledge Transfer
ServicesRoom 301, Pi Ch'iu BuildingShatin, New
Territories
Hong Kong 999077, HK

72 Inventor/es:

LO, YUK-MING DENNIS;
CHIU, ROSSA WAI KWUN;
CHAN, KWAN CHEE y
JIANG, PEIYONG

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 989 622 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

Uso de anomalías de tamaño y número en el ADN plasmático para detectar un cáncer

5 **Antecedentes**

El análisis del ADN extracelular circulante se ha utilizado cada vez más para la detección y el seguimiento de cánceres (1-3). Diferentes características moleculares asociadas al cáncer, incluidas anomalías en el número de copias (4-7), cambios de metilación (8-11), mutaciones de un solo nucleótido (4, 12-15), secuencias víricas procedentes del cáncer (16, 17) y reordenamientos cromosómicos (18, 19), se pueden detectar en el plasma de pacientes con diversos tipos de cáncer. A pesar de la rápida expansión de las aplicaciones clínicas, muchas características moleculares fundamentales del ADN circulante en pacientes con cáncer siguen sin estar claras, limitando así el uso clínico más eficaz de dichos análisis.

En particular, estudios previos sobre el tamaño del ADN circulante en pacientes con cáncer dieron resultados incoherentes. Los estudios han demostrado que la integridad general (una medida de tamaño) del ADN circulante aumentaría en pacientes con cáncer en comparación con sujetos sin una condición maligna (20-23). Usando PCR con diferentes tamaños de amplicón, se demostró que la proporción de ADN más largo sería mayor en pacientes con cáncer. Se demostró que esta anomalía en la integridad del ADN era reversible después del tratamiento y la persistencia de dichos cambios se asoció con un mal pronóstico (20, 24). Por otra parte, también hay evidencia aparentemente contradictoria de que el ADN circulante procedente de tejidos tumorales podría ser más corto que el procedente de células no malignas. Por ejemplo, se ha demostrado que la proporción de moléculas de ADN que portan mutaciones asociadas al cáncer sería mayor cuando esas mutaciones se detectaran utilizando PCR con amplicones más cortos (12, 25).

Además, estudiar el perfil de tamaño del ADN procedente del tumor en el plasma de los pacientes con CHC es una tarea compleja porque el ADN en plasma procedente del tumor no se puede distinguir fácilmente del ADN de fondo en plasma no procedente del tumor. La detección de mutaciones específicas del cáncer ofrece un medio genotípico para distinguir el ADN plasmático tumoral del no tumoral. Sin embargo, hay relativamente pocas mutaciones específicas del cáncer en todo el genoma (29-32). Por consiguiente, puede ser difícil identificar con precisión el ADN procedente del tumor en el plasma, en particular con el propósito de generar una visión amplia, detallada y a la vez rentable de la distribución del tamaño del ADN procedente del tumor.

Dichas dificultades obstaculizan la obtención de mediciones precisas en muestras que posiblemente contengan mezclas de ADN tumoral y no tumoral.

El documento WO-2009/051842 divulga un método para detectar un cáncer mediante la medición del número de copias genómicas y la longitud de la cadena en ADN extracelular.

Breve resumen

De acuerdo con la presente invención, se proporciona un método de análisis de una muestra biológica de un organismo como se define en la reivindicación 1. En las reivindicaciones dependientes se definen aspectos adicionales y realizaciones preferidas.

Adicionalmente de acuerdo con la presente invención, se proporciona un programa informático como se define en la reivindicación 12.

Se puede obtener una mejor comprensión de la naturaleza y ventajas de las realizaciones de la presente invención con referencia a la siguiente descripción detallada y los dibujos adjuntos.

Breve descripción de los dibujos

La figura 1 es un diagrama que ilustra un método 100 para identificar regiones cromosómicas que presentan una anomalía.

En la figura 2 se muestra un gráfico Circos 200 que identifica regiones que presentan amplificaciones y eliminaciones en muestras de plasma y tejido de un paciente representativo con carcinoma hepatocelular (CHC).

En la figura 3 se muestran resultados de las anomalías del número de copias (ANC) en plasma para varios sujetos. La figura 4 es una tabla 400 que muestra la detectabilidad de las ANC en plasma de pacientes con CHC, portadores del virus de la hepatitis B (VHB), pacientes con cirrosis hepática y sujetos sanos.

En la figura 5 se muestra una tabla 500 de ANC detectadas en el tumor y el plasma correspondiente de 12 pacientes con CHC.

En la figura 6 se muestra un diagrama que ilustra un método de análisis de una muestra biológica de un organismo para determinar si una muestra biológica presenta un primer tipo de cáncer.

En la figura 7 se muestran brazos cromosómicos que presentan patrones diferentes para diferentes tipos de cánceres en la tabla 700.

En las figuras 8A, 8B y 8C se muestran una tabla 800 de patrones de regiones cromosómicas para diferentes tipos

de cáncer.

En la figura 9 se muestra un diagrama que ilustra un método de análisis de una muestra biológica de un organismo. En la figura 10 se muestran gráficos de las proporciones de fragmentos de ADN plasmático (A) más cortos de 150 pb, (B) de 150 a 180 pb y (C) más largos de 180 pb frente a la fracción de ADN tumoral en plasma.

La figura 11 es una ilustración esquemática del principio del análisis del tamaño del ADN plasmático en pacientes con cáncer.

En la figura 12 se muestran distribuciones de tamaño de ADN plasmático originado a partir del 8q amplificado y el 8p eliminado de un caso representativo H291. (A) Las distribuciones de tamaño del ADN plasmático para 8p (rojo) y 8q (verde). (B) Gráfico de frecuencias acumuladas para el tamaño del ADN plasmático para 8p (rojo) y 8q (verde). (C) La diferencia en frecuencias acumuladas para el caso de CHC H291.

En la figura 13 se muestra la diferencia en las frecuencias acumuladas para el tamaño entre 8q y 8p (ΔS). (A) Gráfico de ΔS frente al tamaño de todos los casos de CHC con diferentes ANC en 8p y 8q en plasma. (B) Los valores de ΔS_{166} entre diferentes grupos.

La figura 14 es un gráfico de los valores de ΔS entre 1q y 1p frente al tamaño para un paciente con CHC representativo.

La figura 15 es un gráfico de los valores de ΔS_{166} entre 1q y 1p para sujetos de control sanos, portadores del VHB, pacientes cirróticos y pacientes con CHC.

La figura 16 es un diagrama que ilustra un método para realizar un análisis de puntuación z a nivel de brazo cromosómico (CAZA, del inglés *chromosome arm-level z-score analysis*) y un análisis de tamaño para analizar una muestra biológica de un organismo.

La figura 17 es un diagrama que ilustra un método para analizar una muestra biológica de un organismo.

En la figura 18 se muestran distribuciones de tamaño de fragmentos de ADN plasmático en pacientes con CHC con diferentes concentraciones fraccionarias de ADN procedente del tumor en plasma.

En la figura 19 se muestran perfiles de tamaño del ADN plasmático para (A) controles sanos, (B) portadores crónicos del VHB y (C) pacientes cirróticos.

En la figura 20 se muestran diagramas de caja de la proporción de fragmentos cortos para sujetos de control sanos, pacientes con CHC con fracción de ADN tumoral de menos de un 2 % en plasma y pacientes con CHC con fracción de ADN tumoral de más de un 6 %.

La figura 21 es una curva de eficacia diagnóstica (ROC, del inglés *operating characteristic curve*) para aplicar $P(<150)$ para diferenciar pacientes con CHC con una fracción de ADN tumoral inferior al 2 % de sujetos de control sanos.

La figura 22 es una curva de eficacia diagnóstica (ROC) para aplicar $P(<150)$ para diferenciar pacientes con CHC con una fracción de ADN tumoral superior al 6 % y sujetos sanos.

En la figura 23 se muestran diagramas de caja de la proporción de fragmentos largos para sujetos de control sanos y pacientes con CHC con fracción de ADN tumoral inferior al 2 % en plasma.

La figura 24 es una curva ROC para utilizar $P(>180)$ para diferenciar pacientes con CHC con una fracción de ADN tumoral inferior al 2 % de sujetos de control sanos.

En la figura 25 se muestran diagramas de caja del tamaño medio de los fragmentos de sujetos de control sanos, pacientes con CHC con una fracción de ADN tumoral inferior al 2 % y pacientes con CHC con una fracción de ADN tumoral superior al 6 %.

La figura 26 es una curva ROC para utilizar el tamaño medio de los fragmentos para diferenciar entre pacientes con CHC con una fracción de ADN tumoral inferior al 2 % y sujetos de control sanos.

La figura 27 es una curva ROC para utilizar el tamaño medio de los fragmentos para diferenciar entre pacientes con CHC con una fracción de ADN tumoral superior al 6 % y sujetos de control sanos.

En la figura 28 se muestra un diagrama de caja de la proporción de fragmentos cortos de ADN plasmático de menos de 150 pb que se alinearon con el cromosoma 1q para pacientes con CHC con una fracción de ADN tumoral superior al 6 % y para sujetos de control sanos.

La figura 29 es una curva ROC para utilizar la proporción de fragmentos cortos de ADN plasmático de menos de 150 pb para diferenciar entre pacientes con CHC con una fracción de ADN tumoral superior al 6 % y sujetos de control sanos.

La figura 30 es un gráfico de ΔS frente al tamaño del tumor de pacientes con CHC.

La figura 31 es un gráfico del porcentaje de fragmentos de ADN de un tamaño determinado frente al tamaño del tumor.

En la figura 32 se muestra un diagrama de bloques de un sistema informático 10 ilustrativo que puede utilizarse con el sistema y los métodos de acuerdo con las realizaciones de la presente divulgación.

TÉRMINOS

La expresión "*muestra biológica*", como se utiliza en el presente documento, se refiere a cualquier muestra que se toma de un sujeto (por ejemplo, un ser humano, tal como una mujer gestante) y contiene una o más moléculas de ácido nucleico de interés. Los ejemplos incluyen plasma, saliva, líquido pleural, sudor, líquido ascítico, bilis, orina, suero, jugo pancreático, heces, líquido de lavado cervical y muestras de frotis cervical.

La expresión "*ácido nucleico*" o "*polinucleótido*" se refiere a un ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN) y a un polímero de los mismos en forma mono o bicatenaria. A menos que esté específicamente limitado, la expresión abarca los ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen

propiedades de unión similares a las del ácido nucleico de referencia y que se metabolizan de manera similar a la de los nucleótidos de origen natural. A menos que se indique de otro modo, una secuencia de ácido nucleico particular también abarca implícitamente variantes modificadas conservativamente de la misma (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, polimorfismos de un solo nucleótido (SNP, del inglés *single nucleotide polymorphism*) y secuencias complementarias así como la secuencia explícitamente indicada. De manera específica, las sustituciones de codones degenerados se pueden lograr mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con restos de base mixta y/o de desoxiinosina (Batzner M. A. *et al.*, Nucleic Acid Res 1991; 19:5081; Ohtsuka E. *et al.*, J Biol Chem 1985; 260:2605-2608; y Rossolini G. M. *et al.*, Mol Cell Probes 1994; 8:91-98). La expresión ácido nucleico se utiliza indistintamente con gen, ADNc, ARNm, ARN pequeño no codificante, microARN (miARN), ARN asociado a Piwi y ARN en horquilla corto (ARNhc) codificado por un gen o locus.

El término "*gen*" significa el segmento de ADN implicado en la producción de una cadena polipeptídica. Puede incluir regiones anteriores y posteriores a la región codificante (líder y remolque) así como las secuencias intervinientes (intrones) entre los segmentos de codificación individuales (exones).

Como se utiliza en el presente documento, el término "*locus*" o su forma plural "*loci*" es una ubicación o dirección de cualquier longitud de nucleótidos (o pares de bases) que tiene una variación a través de los genomas.

La expresión "*marcador secuenciado*" (también denominado lectura de secuencia) se refiere a una secuencia obtenida de toda o parte de una molécula de ácido nucleico, por ejemplo, un fragmento de ADN. En una realización, sólo se secuencia un extremo del fragmento, por ejemplo, aproximadamente 30 pb. A continuación, el marcador secuenciado se puede alinear con un genoma de referencia. Como alternativa, ambos extremos del fragmento se pueden secuenciar para generar dos marcadores secuenciados, que pueden proporcionar una mayor precisión en la alineación y también proporcionar una longitud del fragmento. En otra realización más, un fragmento de ADN lineal puede circularizarse, por ejemplo, mediante unión, y la parte que abarca el sitio de unión se puede secuenciar.

La expresión concentración fraccionaria de ADN tumoral se utiliza indistintamente con las expresiones proporción de ADN tumoral y fracción de ADN tumoral, y se refiere a la proporción de moléculas de ADN que están presentes en una muestra procedente de un tumor.

La expresión "*perfil de tamaño*" generalmente se refiere a los tamaños de los fragmentos de ADN en una muestra biológica. Un perfil de tamaño puede ser un histograma que proporciona una distribución de una cantidad de fragmentos de ADN en una variedad de tamaños. Se pueden utilizar varios parámetros estadísticos (también denominados parámetros de tamaño o simplemente parámetros) para distinguir un perfil de tamaño de otro. Un parámetro es el porcentaje de fragmentos de ADN de un determinado tamaño o intervalo de tamaños en relación con todos los fragmentos de ADN o en relación con fragmentos de ADN de otro tamaño o intervalo.

El término "*parámetro*", como se utiliza en el presente documento, significa un valor numérico que caracteriza un conjunto de datos cuantitativo y/o una relación numérica entre conjuntos de datos cuantitativos. Por ejemplo, un parámetro es una proporción (o función de una proporción) entre una primera cantidad de una primera secuencia de ácido nucleico y una segunda cantidad de una segunda secuencia de ácido nucleico.

El término "*clasificación*", como se utiliza en el presente documento, se refiere a cualquier número(s) u otro carácter(es) (incluidas palabras) que se asocian a una propiedad particular de una muestra. Por ejemplo, un símbolo "+" podría significar que una muestra está clasificada como que tiene eliminaciones o amplificaciones (por ejemplo, duplicaciones). Las expresiones "*valor de corte*" y "*umbral*" se refieren a un número predeterminado utilizado en una operación. Por ejemplo, un tamaño de corte puede referirse a un tamaño por encima del cual se excluyen los fragmentos. Un valor liminar puede ser un valor por encima o por debajo del cual se aplica una determinada clasificación. Cualquiera de estas expresiones puede utilizarse en cualquiera de estos contextos.

La expresión "*nivel de cáncer*" puede referirse a si existe cáncer, un estadio de un cáncer, un tamaño del tumor, cuántas eliminaciones o amplificaciones de una región cromosómica están implicadas (por ejemplo, duplicadas o triplicadas) y/u otra medida de la gravedad de un cáncer. El nivel de cáncer podría ser un número u otros caracteres. El nivel podría ser cero. El nivel de cáncer también incluye afecciones premalignas o precancerosas asociadas con eliminaciones o amplificaciones.

Una "región subcromosómica" es una región que es más pequeña que un cromosoma. Ejemplos de regiones subcromosómicas son 100 kb, 200 kb, 500 kb, 1 Mb, 2 Mb, 5 Mb o 10 Mb. Otro ejemplo de una región subcromosómica es aquella que corresponde a una o más bandas, o subbandas, o uno de los brazos de un cromosoma. Las bandas o subbandas son características observadas en el análisis citogenético. Se puede hacer referencia a una región subcromosómica por sus coordenadas genómicas en relación con una secuencia de referencia del genoma humano.

Descripción detallada

Los cánceres a menudo tienen regiones con anomalías en el número de copias (amplificaciones o eliminaciones) en relación con el genoma normal de la persona. Las técnicas pueden contar fragmentos de ADN extracelular en una muestra (por ejemplo, plasma o suero) que incluyen fragmentos de ADN tumoral y fragmentos de ADN no tumoral. El recuento puede identificar regiones que están sobrerrepresentadas (indicativo de amplificación) o infrarrepresentadas (indicativo de eliminación). Pero, como dichas técnicas basadas en el recuento son de naturaleza estadística, pueden producirse indicaciones incorrectas. Las realizaciones pueden identificar una región como posiblemente portadora de una anomalía en el número de copias (también denominada anomalía) mediante un análisis basado en el recuento y confirmar si la región tiene la anomalía mediante un análisis basado en el tamaño. Dicha confirmación proporciona una precisión adicional a la hora de identificar regiones con anomalías.

Las regiones que presentan anomalías se pueden utilizar para identificar la existencia de cáncer en el organismo del que se obtuvo la muestra. Pero, la existencia de cáncer no conlleva un tipo de cáncer. A fin de abordar este problema, las realizaciones pueden utilizar patrones de referencia de anomalías en regiones a partir de muestras con cánceres conocidos. Es posible determinar un patrón de prueba que indique qué regiones son anómalas para una muestra determinada que se esté analizando, y el patrón de prueba se puede comparar con los patrones de referencia para determinar un tipo de cáncer. Se puede determinar una cantidad de regiones del patrón de prueba que presentan una misma eliminación o amplificación que un patrón de referencia correspondiente a un tipo particular de cáncer, y la cantidad se puede comparar con un umbral para determinar una clasificación de si el tipo particular de cáncer está presente. Una vez que se identifica que una región tiene una anomalía y corresponde a un tipo particular de cáncer, se puede tener mayor confianza al analizar la región en busca de ADN tumoral. Por ejemplo, la región se puede utilizar para medir una fracción de ADN tumoral en la muestra.

Además, varios estudios han mostrado resultados incoherentes en cuanto a la longitud de los fragmentos de ADN extracelular tumoral: algunos muestran fragmentos más largos de ADN tumoral y otros muestran fragmentos más cortos de ADN tumoral. El análisis a continuación muestra que ambas pueden ser correctas, pero para diferentes fracciones de ADN tumoral. Las realizaciones pueden utilizar diferentes umbrales de tamaño en un análisis de tamaño basado en una fracción de ADN tumoral medida, que puede determinarse mediante el recuento de fragmentos de ADN en una región identificada como que tiene una anomalía. Por consiguiente, algunas implementaciones pueden conciliar estas aparentes incoherencias a través de, por ejemplo: (a) perfiles de tamaño de alta resolución de todo el genoma del ADN plasmático mediante secuenciación masiva en paralelo; y (b) un enfoque eficaz para distinguir el ADN procedente de tumores del ADN de fondo no tumoral en el plasma de pacientes con cáncer (por ejemplo, utilizando regiones identificadas como que tienen una anomalía).

I. INTRODUCCIÓN

Se ha vuelto posible medir las longitudes de millones o miles de millones de cada molécula individual de ADN plasmático en muestras con el uso de la secuenciación masiva en paralelo (26, 27). Por lo tanto, los tamaños del ADN plasmático podrían estudiarse a nivel de todo el genoma y con una resolución de base única. Usando este enfoque, se ha demostrado que el tamaño del ADN circulante generalmente se asemeja al tamaño del ADN mononucleosómico, lo que sugiere que el ADN plasmático podría generarse a través de la apoptosis (26, 27). En mujeres embarazadas, se ha demostrado que el ADN plasmático procedente del feto es más corto que el ADN procedente de la madre (26). La diferencia de tamaño entre el ADN fetal y materno circulante ha proporcionado una nueva base conceptual para cuantificar el ADN fetal en el plasma materno y detectar aneuploidías cromosómicas mediante el análisis del tamaño del ADN plasmático (28). Asimismo, se han observado diferencias en las distribuciones de tamaño del ADN circulante procedente de los órganos trasplantados y de los propios tejidos de los pacientes en los receptores de trasplantes de órganos sólidos o de médula ósea (27).

El plasma de pacientes con cáncer contiene una mezcla de ADN procedente de tumores y ADN no procedente de tumores. Los ejemplos a continuación analizan la distribución del tamaño del ADN plasmático en pacientes con cáncer con carcinoma hepatocelular (CHC). También se analizaron distribuciones de tamaño del ADN plasmático en pacientes con CHC, pacientes con infección crónica por el virus de la hepatitis B (VHB), pacientes con cirrosis hepática y sujetos sanos. Las realizaciones pueden utilizar ciertas regiones anómalas para analizar el perfil de tamaño del ADN procedente del tumor en el plasma de los pacientes con CHC. El uso de dichas regiones anómalas puede superar el desafío de que el ADN plasmático procedente del tumor no se distingue fácilmente del ADN de fondo no procedente del tumor en el plasma.

Algunas realizaciones utilizan brazos cromosómicos que se ven afectados por anomalías en el número de copias (ANC) para inferir la diferencia en las distribuciones de tamaño del ADN plasmático procedente de tumores y no procedentes de tumores. Para los brazos cromosómicos que se amplifican en los tejidos tumorales, la contribución proporcional del ADN procedente de tumores al ADN plasmático aumentaría, mientras que para los brazos cromosómicos que se eliminan en el tumor, la contribución disminuiría. Por consiguiente, la comparación de los perfiles de tamaño de los brazos cromosómicos amplificados y eliminados reflejaría la diferencia de tamaño entre el ADN plasmático procedente de tumores y el no procedente de tumores. Las ANC que implican un brazo cromosómico completo o un tronco grande de un brazo cromosómico son relativamente comunes (33). La eliminación de los cromosomas 1p y 8p y la amplificación de los cromosomas 1q y 8q se observan comúnmente en los tejidos de CHC (34-36). Por tanto, el análisis se centra en los cromosomas 1 y 8 para los análisis de ANC y de perfiles de tamaño del

ADN plasmático.

II. ANÁLISIS DE RECuento PARA IDENTIFICAR REGIONES ANÓMALAS

- 5 Una región anómala incluye una amplificación o una eliminación. Una amplificación significa que una secuencia en la región ocurre con mayor frecuencia que en una secuencia de referencia y, por tanto, la secuencia ha sido amplificada. Normalmente, la amplificación se produciría en una sola copia del cromosoma (haplotipo). Una eliminación significa que se ha eliminado una secuencia en la región en relación con la secuencia de referencia, normalmente sólo una copia del cromosoma tiene la eliminación en el caso de los organismos diploides. Una región puede definirse por al menos dos loci (que están separados entre sí), y los fragmentos de ADN de estos loci pueden utilizarse para obtener un valor colectivo sobre la región.

A. Detección de una región anómala mediante recuento

- 15 La anomalía de una región se puede determinar mediante el recuento de la cantidad de fragmentos de ADN (moléculas) que proceden de la región. Como ejemplos, la cantidad puede ser un número de fragmentos de ADN, un número de bases a las que se superpone un fragmento de ADN u otra medida de fragmentos de ADN en una región. La cantidad de fragmentos de ADN para la región se puede determinar mediante la secuenciación de los fragmentos de ADN para obtener lecturas de secuencia y la alineación de las lecturas de secuencia con un genoma de referencia.
- 20 En una realización, la cantidad de lecturas de secuencia para la región se puede comparar con la cantidad de lecturas de secuencia para otra región a fin de determinar sobrerrepresentación (amplificación) o infrarrepresentación (eliminación). En otra realización, se puede determinar la cantidad de lecturas de secuencia para un haplotipo y compararla con la cantidad de lecturas de secuencia para otro haplotipo.
- 25 Por consiguiente, el número de fragmentos de ADN de una región cromosómica (por ejemplo, determinado mediante el recuento de los marcadores secuenciados alineados con esa región) se puede comparar con un valor de referencia (que se puede determinar a partir de una región cromosómica de referencia, de la región de otro haplotipo o de la misma región en otra muestra que se sabe que es sana). La comparación puede determinar si la cantidad es estadísticamente diferente (por ejemplo, superior o inferior) al valor de referencia. Se puede utilizar un umbral para la diferencia, por ejemplo, correspondiente a 3 desviaciones típicas (DT), como se observa en una distribución de valores observada en una población.

- Como parte de la comparación, se puede normalizar un recuento de marcadores antes de la comparación. Se puede calcular un valor normalizado de las lecturas de secuencia (marcadores) para una región concreta dividiendo el número de lecturas de secuencia que se alinean con esa región entre el número total de lecturas de secuencia alineables con todo el genoma. Este recuento normalizado de marcadores permite comparar los resultados de una muestra con los de otra. Por ejemplo, el valor normalizado puede ser la proporción (por ejemplo, porcentaje o fracción) de lecturas de secuencias que se espera que procedan de la región concreta. Pero, son posibles muchas otras normalizaciones, como resultaría evidente para un experto en la materia. Por ejemplo, se puede normalizar dividiendo el número de recuentos de una región por el número de recuentos de una región de referencia (en el caso anterior, la región de referencia es solo el genoma completo) o utilizando siempre el mismo número de lecturas de secuencia. A continuación, este recuento normalizado de marcadores puede compararse con un valor liminar, que puede determinarse a partir de una o más muestras de referencia que no presenten cáncer.

- 45 En algunas realizaciones, el valor liminar puede ser el valor de referencia. En otras realizaciones, el valor de referencia puede ser el otro valor utilizado para la normalización, y la comparación puede incluir el valor de referencia y el valor liminar. Por ejemplo, la cantidad de la región se puede dividir por el valor de referencia para obtener un parámetro, que se compara con el valor liminar para ver si existe una diferencia estadísticamente significativa. Como otro ejemplo, la cantidad de la región se puede comparar con el valor de referencia más el valor liminar.

- 50 En una realización, la comparación se realiza calculando la puntuación z del caso para la región cromosómica concreta. La puntuación z puede calcularse mediante la siguiente ecuación: puntuación $z = (\text{recuento normalizado de marcadores del caso} - \text{media})/\text{DT}$, donde "media" es el recuento medio normalizado de marcadores alineados con la región cromosómica concreta para las muestras de referencia; y DT es la desviación típica del número de recuentos normalizados de marcadores que se alinean con la región concreta para las muestras de referencia. Por lo tanto, la puntuación z puede corresponder con el número de desviaciones estándar que el recuento normalizado de marcadores de una región cromosómica para el caso analizado se aleja de la media del recuento normalizado de marcadores para la misma región cromosómica de uno o más sujetos de referencia. Esta puntuación z se puede comparar con un umbral, por ejemplo, 3 para amplificación y -3 para eliminación. Las regiones cromosómicas que se amplifican tendrían un valor positivo de la puntuación z por encima del umbral. Las regiones cromosómicas eliminadas tendrían un valor negativo de la puntuación z , que estaría por debajo del umbral.

- La magnitud de la puntuación z puede determinarse mediante varios factores. Uno de los factores es la concentración fraccionaria de ADN procedente del tumor en la muestra biológica (por ejemplo, plasma). Cuanto mayor sea la concentración fraccionaria de ADN procedente del tumor en la muestra (por ejemplo, plasma), mayor será la diferencia entre el recuento normalizado de marcadores del caso probado y los casos de referencia. Por lo tanto, se obtendría

una mayor magnitud de la puntuación z.

Otro factor es la variación del recuento normalizado de marcadores en uno o varios casos de referencia. Con el mismo grado de sobrerepresentación de la región cromosómica en la muestra biológica (por ejemplo, plasma) del caso analizado, una variación menor (es decir, una desviación típica menor) del recuento normalizado de marcadores en el grupo de referencia daría lugar a una puntuación z mayor. De forma similar, con el mismo grado de infrarrepresentación de la región cromosómica en la muestra biológica (por ejemplo, plasma) del caso analizado, una desviación típica menor del recuento normalizado de marcadores en el grupo de referencia daría lugar a una puntuación z más negativa.

Otro factor es la magnitud de la anomalía cromosómica en los tejidos tumorales. La magnitud de la anomalía cromosómica se refiere a los cambios en el número de copias de la región cromosómica concreta (ganancia o pérdida). Cuanto mayores sean los cambios en el número de copias en los tejidos tumorales, mayor será el grado de sobrerepresentación o infrarrepresentación de la región cromosómica concreta en el ADN plasmático. Por ejemplo, la pérdida de ambas copias del cromosoma daría lugar a una mayor infrarrepresentación de la región cromosómica en el ADN plasmático que la pérdida de una de las dos copias del cromosoma y, por lo tanto, daría lugar a una puntuación z más negativa. Normalmente, existen múltiples anomalías cromosómicas en los cánceres. Las anomalías cromosómicas de cada cáncer pueden variar además por su naturaleza (es decir, amplificación o eliminación), su grado (ganancia o pérdida de una o varias copias) y su extensión (tamaño de la anomalía en términos de longitud cromosómica).

La precisión de la medición del recuento normalizado de marcadores se ve afectada por el número de moléculas analizadas. Por ejemplo, puede ser necesario analizar 15.000, 60.000 y 240.000 moléculas para detectar anomalías cromosómicas con un cambio de copia (ya sea ganancia o pérdida) cuando la concentración fraccionaria es aproximadamente un 12,5 %, un 6,3 % y un 3,2 %, respectivamente. Se describen más detalles sobre el recuento de marcadores para la detección de un cáncer en diferentes regiones cromosómicas en la publicación de patente de EE. UU. n.º 2009/0029377 titulada "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing" de Lo *et al.*; y en la patente de EE. UU. n.º 8.741.811 titulada "Detection Of Genetic Or Molecular Aberrations Associated With Cancer" de Lo *et al.*

B. Método

La figura 1 es un diagrama que ilustra un método 100 para identificar una región cromosómica que potencialmente presenta una amplificación de acuerdo con realizaciones de la presente divulgación, pero que no forman parte de la invención reivindicada. El método 100, y otros métodos descritos en el presente documento, puede realizarse total o parcialmente mediante un sistema informático.

En la etapa 110, se pueden identificar una pluralidad de regiones cromosómicas de un organismo. Cada región cromosómica puede incluir una pluralidad de loci. Una región puede tener un tamaño de 1 Mb o cualquier otro tamaño igual. El genoma completo puede incluir entonces aproximadamente 3.000 regiones, cada una de tamaño y ubicación predeterminados. Dichas regiones predeterminadas pueden variar para adaptarse a la longitud de un cromosoma concreto o a un número especificado de regiones que se vayan a utilizar, así como a cualquier otro criterio mencionado en el presente documento. Si las regiones tienen longitudes diferentes, estas longitudes pueden utilizarse para normalizar los resultados, por ejemplo, como se describe en el presente documento.

Las etapas 120 a 140 se pueden realizar para cada una de las regiones cromosómicas. En la etapa 120, para cada región cromosómica, un grupo respectivo de moléculas de ácido nucleico puede identificarse como de la región cromosómica. La identificación puede basarse en la identificación de una ubicación de moléculas de ácido nucleico en un genoma de referencia. Por ejemplo, los fragmentos de ADN extracelular se pueden secuenciar para obtener lecturas de secuencia, y las lecturas de secuencia pueden cartografiarse (alinearse) con el genoma de referencia. Si el organismo fuera un ser humano, el genoma de referencia sería un genoma humano de referencia, potencialmente de una subpoblación particular. Como otro ejemplo, los fragmentos de ADN extracelular pueden analizarse con diferentes sondas (por ejemplo, tras la PCR u otro tipo de amplificación), donde cada sonda corresponde a una ubicación genómica diferente. En algunas realizaciones, el análisis de los fragmentos de ADN extracelular se puede realizar recibiendo lecturas de secuencia u otros datos experimentales correspondientes a los fragmentos de ADN extracelular y analizando después los datos experimentales con un sistema informático.

En la etapa 130, un sistema informático puede calcular una cantidad respectiva del respectivo grupo de moléculas de ácido nucleico. El valor respectivo define una propiedad de las moléculas de ácido nucleico del grupo respectivo. El valor respectivo puede ser cualquiera de los valores mencionados en el presente documento. Por ejemplo, el valor puede ser el número de fragmentos del grupo o un valor estadístico de una distribución de tamaños de los fragmentos del grupo. El valor respectivo también puede ser un valor normalizado, por ejemplo, un recuento de marcadores de la región dividido entre el número total de recuentos de marcadores de la muestra o el número de recuentos de marcadores de una región de referencia. El valor respectivo también puede ser una diferencia o una proporción con respecto a otro valor, proporcionando así la propiedad de una diferencia para la región.

En la etapa 140, la cantidad respectiva puede compararse con un valor de referencia para determinar una clasificación

de si la región cromosómica presenta una anomalía (es decir, una amplificación o una eliminación). En algunas realizaciones, la región cromosómica puede clasificarse como que no presenta anomalía. La comparación puede incluir la determinación de una puntuación z basada en la cantidad respectiva y el valor de referencia. Como ejemplo, el valor de referencia puede ser cualquier umbral o valor de referencia descrito en el presente documento. Por ejemplo, el valor de referencia podría ser un valor liminar determinado para muestras normales. Como otro ejemplo, el valor de referencia podría ser el recuento de marcadores para otra región, y la comparación puede incluir tomar una diferencia o una proporción (o función de los mismos) y después determinar si la diferencia o proporción es mayor que un valor liminar.

El valor de referencia puede variar en función de los resultados de otras regiones. Por ejemplo, si las regiones vecinas también muestran una desviación (aunque sea pequeña en comparación con un umbral, por ejemplo, una puntuación z de 3), puede utilizarse un umbral más bajo. Por ejemplo, si tres regiones consecutivas superan un primer umbral, entonces el cáncer puede ser más probable. Por tanto, este primer umbral puede ser inferior a otro umbral necesario para identificar el cáncer de regiones no consecutivas. Tener tres regiones (o más de tres) que presenten incluso una pequeña desviación puede tener una probabilidad de efecto fortuito lo suficientemente baja como para que se conserven la sensibilidad y la especificidad.

C. Análisis de la puntuación Z a nivel de brazo cromosómico (CAZA)

En algunas realizaciones, un cromosoma se puede dividir en muchas regiones subcromosómicas (por ejemplo, regiones de 1 Mb). Es posible que esta alta resolución no maximice la sensibilidad y especificidad. Otras realizaciones pueden dividir un cromosoma en dos brazos, es decir, p y q . Analizar los dos brazos puede mejorar la especificidad mediante la reducción del ruido provocado por una resolución tan fina. A continuación se proporciona un ejemplo de análisis de la puntuación z a nivel de brazo cromosómico.

Se analizaron un total de 225 muestras de ADN plasmático de 90 pacientes con CHC, 67 pacientes con infección crónica por VHB, 36 pacientes con cirrosis hepática asociada al VHB y 32 sujetos sanos. Se obtuvo una mediana de 31 millones de lecturas (intervalo: de 17 a 79 millones) de cada muestra de plasma. Las cantidades de lecturas de secuencia originadas en brazos cromosómicos que estaban tres DT por debajo (puntuaciones $z < -3$) y tres DT por encima (puntuaciones $z > 3$) de la media de los controles sanos se consideraron que indicaban una infrarrepresentación y sobrerrepresentación significativas del ADN plasmático de esos brazos cromosómicos, respectivamente. Estas anomalías cuantitativas del ADN plasmático generalmente reflejaban la presencia de pérdidas y ganancias del número de copias (ANC) en el tumor (4).

En la figura 2 se muestra un gráfico Circos 200 que identifica regiones que presentan amplificaciones y eliminaciones en muestras de plasma y tejido de un paciente representativo con carcinoma hepatocelular (CHC). De dentro hacia fuera: ANC en el tejido tumoral (en resolución de 1 Mb); ANC a nivel del brazo en el tejido tumoral; ANC de plasma (en resolución de 1 Mb); ANC de plasma a nivel del brazo. Las regiones con ganancias y pérdidas se muestran en verde y rojo, respectivamente. La distancia entre dos líneas horizontales consecutivas representa una puntuación z de 5. Los ideogramas cromosómicos (fuera de los gráficos) están orientados de pter a qter en el sentido de las agujas del reloj.

En la figura 3 se muestran los resultados de las anomalías en el número de copias (ANC) en plasma para todos los sujetos estudiados utilizando una realización de CAZA. Se analizaron los cuatro brazos cromosómicos (1p, 1q, 8p y 8q) que se ven frecuentemente afectados por las ANC en el CHC. Las líneas rojas y verdes representan la infrarrepresentación y la sobrerrepresentación, respectivamente, de los brazos cromosómicos correspondientes en el plasma. Cada línea vertical representa los datos de un caso.

La figura 4 es una tabla 400 que muestra la detectabilidad de las ANC en plasma de pacientes con CHC, portadores del VHB, pacientes con cirrosis hepática y sujetos sanos de acuerdo con realizaciones de la presente divulgación. En la tabla 400 se muestran las categorías de pacientes en la columna más a la izquierda. En las columnas restantes se muestran el número de pacientes y el porcentaje con ANC detectadas en el plasma para diferentes brazos cromosómicos. Setenta y seis (84,4 %) de los 90 pacientes con CHC tenían al menos una ANC a nivel de brazo cromosómico en los cromosomas 1 y 8 en el plasma. Se dispuso de tejidos tumorales de 12 pacientes con CHC para corroborar los hallazgos del ADN plasmático. Las muestras de tejido se secuenciaron y los patrones de ANC se muestran en la figura 5.

En la figura 5 se muestra una tabla 500 de ANC detectadas en el tumor y el plasma correspondiente de 12 pacientes con CHC. En la tabla 500, el número de caso del paciente aparece en la primera columna. Los pacientes están ordenados en orden descendente de fracción de ADN tumoral en plasma, como se muestra en la segunda columna. En la tercera columna se muestra el tamaño del tumor. Las columnas restantes muestran las ANC detectadas en el tumor y el plasma para diferentes brazos cromosómicos. 'Ganancia' indica una ganancia en el número de copias. 'Pérdida' indica una pérdida en el número de copias. 'Nulo' indica que no hay ANC detectables. Se analizaron un total de 48 brazos cromosómicos de los 12 pacientes. Se muestran los números (y porcentajes) de brazos cromosómicos con resultados concordantes y discordantes entre el tumor y el plasma.

De los 48 brazos cromosómicos analizados para los 12 pacientes, se observaron cambios concordantes en el plasma y en los tejidos tumorales en 30 brazos (63 %). Las ANC solo se observaron en el tumor, pero no en el plasma, para 10 (21 %) brazos. Estos casos tendían a tener fracciones de ADN tumoral más bajas en el plasma. Se observaron ANC en el plasma, pero no el tumor, para 7 (15 %) brazos. En un caso (HOT428), se observó una ganancia de 1q en el tumor, pero se observó una pérdida en el plasma. Estos datos podrían sugerir la presencia de heterogeneidad tumoral donde podría haber otros focos o clones de células cancerosas que aportan ADN plasmático.

Entre los portadores del VHB con y sin cirrosis hepática, las proporciones de detección de estas ANC fueron del 22,2 % y del 4,5 %, respectivamente. Un paciente con cirrosis hepática y un portador crónico de VHB sin cirrosis presentaron ANC en plasma, pero no se sabía que tuvieran CHC en el momento de la extracción de sangre, se les diagnosticó CHC a los 3 y 4 meses después, respectivamente. Todos los portadores del VHB y pacientes cirróticos fueron objeto de seguimiento durante al menos 6 meses. Para aquellos sujetos de control sin ninguna ANC en plasma, ninguno de ellos había desarrollado CHC durante el período de seguimiento. Ninguno de los 32 sujetos sanos tenía ANC detectables en el cromosoma 1 u 8 en plasma mediante CAZA. En los pacientes con CHC, el aumento o disminución desproporcionada en las lecturas de secuencia en el plasma debido a la presencia de ANC refleja la concentración fraccionaria de ADN tumoral en la muestra de plasma. La concentración fraccionaria media de ADN procedente del tumor en el plasma de los pacientes con CHC fue del 2,1 % (intervalo: del 0 % al 53,1 %; intervalo intercuartílico: del 1,2 % al 3,8 %).

CAZA proporciona una forma de detectar ANC asociadas a tumores de forma no invasiva. En el CHC, los cromosomas 1 y 8 se ven comúnmente afectados por ANC (34-36). De hecho, los presentes datos mostraron que 76 (84,4 %) de los 90 pacientes con CHC tenían al menos una ANC que afectaba cualquiera de los brazos de los cromosomas 1 y 8 en el plasma, mientras que ninguno de los 32 sujetos sanos presentó ANC para estos dos cromosomas en el plasma. También se detectaron ANC en plasma que implicaban los cromosomas 1 y 8 en el 22,2 % y el 4,5 % de los pacientes cirróticos y portadores del VHB. En un portador del VHB y un paciente con cirrosis hepática, el CHC se diagnosticó poco después de la extracción de sangre. Es probable que el cáncer hubiera estado presente en el momento de la extracción de sangre y estuviera asociado con las ANC en el plasma, demostrando así la capacidad de detección temprana de las realizaciones. La tasa de detección relativamente elevada de ANC en plasma en los pacientes con CHC sugiere que este enfoque podría tener valor futuro en la detección de portadores del VHB. Más aún, las ANC están presentes en casi todos los tipos de cáncer (33). Por consiguiente, este enfoque se puede aplicar como un marcador tumoral genérico con adaptación a los patrones de ANC específicos del cáncer de interés.

III. DETECCIÓN DEL TIPO DE CÁNCER EN FUNCIÓN DEL PATRÓN DE REGIONES ANÓMALAS

Algunas realizaciones pueden utilizar regiones anómalas conocidas (junto con amplificación o eliminación) de un tipo de cáncer para identificar posibles cánceres implicados por anomalías identificadas en la muestra. En el ejemplo anterior, las regiones anómalas conocidas para el CHC se utilizaron para analizar la muestra en busca de CHC. Esta evaluación puede comparar las regiones anómalas identificadas (ya sea amplificación o eliminación) con un conjunto conocido de regiones anómalas. Si se determina una coincidencia suficientemente elevada, entonces ese tipo de cáncer puede marcarse como un posible resultado de prueba.

Un criterio de coincidencia puede ser el porcentaje de regiones del conjunto que también están identificadas en la muestra. Los criterios de coincidencia pueden requerir que regiones específicas sean anómalas. Por ejemplo, la coincidencia se puede identificar para CHC cuando 1p, 1q u 8q es anómalo, o cuando más de uno de estos brazos cromosómicos son anómalos. Por tanto, puede haber subconjuntos específicos para los que se requiera una coincidencia idéntica, pero los subconjuntos pueden ser más pequeños que un conjunto completo de regiones anómalas conocidas para un tipo de cáncer.

Por tanto, un patrón de regiones anómalas para una muestra de prueba se puede comparar con el patrón de regiones anómalas para un tipo particular de cáncer, que puede determinarse a partir de pacientes que se sabe que tienen un tipo particular de cáncer. Las realizaciones se pueden utilizar para detectar el cáncer e identificar el tipo de cáncer en cuestión, en particular cuando el tumor puede ser pequeño (por ejemplo, menos de 2 cm de tamaño). Las técnicas de imagen tienen dificultades para identificar tumores de menos de 2 cm de tamaño. Dichas técnicas también se pueden utilizar para seguir el progreso del paciente después del tratamiento.

A. Método

La figura 6 es un diagrama que ilustra un método 600 de análisis de una muestra biológica de un organismo para determinar si una muestra biológica presenta un primer tipo de cáncer de acuerdo con realizaciones de la presente divulgación, pero que no forman parte de la invención reivindicada. La muestra biológica incluye moléculas de ácido nucleico (también llamadas fragmentos) originarias de células normales y potencialmente de células asociadas con un cáncer. Al menos algunas de estas moléculas pueden ser extracelulares en la muestra.

En una realización de este y cualquier otro método descrito en el presente documento, la muestra biológica incluye fragmentos de ADN extracelulares. Aunque se ha utilizado el análisis del ADN plasmático para ilustrar los diferentes métodos descritos en esta solicitud, estos métodos también pueden aplicarse para detectar anomalías cromosómicas

asociadas a tumores en muestras que contienen una mezcla de ADN normal y procedente de tumores. Los otros tipos de muestras incluyen saliva, lágrimas, líquido pleural, líquido ascítico, bilis, orina, suero, jugo pancreático, heces y muestras de frotis cervical

5 En la etapa 610, se identifica una pluralidad de regiones cromosómicas del organismo. La pluralidad de regiones cromosómicas son subcromosómicas y pueden no superponerse. Las regiones cromosómicas que se cuentan pueden tener restricciones. Por ejemplo, solo pueden contarse las regiones contiguas con al menos otra región (o puede exigirse que las regiones contiguas tengan un tamaño determinado, por ejemplo, cuatro o más regiones). Para las realizaciones en las que las regiones no son iguales, el número también puede representar las longitudes respectivas (por ejemplo, el número podría ser una longitud total de las regiones anómalas). En algunas realizaciones, las regiones corresponden a los brazos de los cromosomas. En otras realizaciones, las regiones pueden ser más pequeñas que los brazos, por ejemplo, regiones de 1 Mb.

15 En algunas realizaciones, una región cromosómica puede ser de un haplotipo particular (es decir, corresponder a una copia particular del cromosoma). En realizaciones que utilizan un análisis de dosis relativa de haplotipos (RHDO, del inglés *relative haplotype dosage*), cada región puede incluir al menos dos loci heterocigotos. Se pueden encontrar más detalles sobre RHDO en la patente de EE. UU. n.º 8.741.811.

20 En la etapa 620, para cada una de una pluralidad de moléculas de ácido nucleico en la muestra biológica del organismo, se puede identificar una ubicación de la molécula de ácido nucleico en un genoma de referencia del organismo. La pluralidad de moléculas de ácido nucleico puede incluir 500.000 o más moléculas (fragmentos). Esta ubicación se puede realizar de diversas maneras, incluida la realización de una secuenciación de una molécula (por ejemplo, mediante una secuenciación aleatoria), para obtener uno o dos marcadores secuenciados (de extremos emparejados) de la molécula y después alinear lo(s) marcador(es) secuenciado(s) con el genoma de referencia. Dicha alineación se puede realizar utilizando herramientas, tales como la herramienta de búsqueda de alineación local básica (BLAST). La ubicación se puede identificar como un número en un brazo de un cromosoma.

25 En la etapa 630, se puede identificar un grupo respectivo de moléculas de ácido nucleico como de la región cromosómica en función de la región identificada, para cada una de la pluralidad de regiones cromosómicas. El grupo respectivo puede incluir al menos una molécula de ácido nucleico ubicada en cada uno de la pluralidad de loci de la región cromosómica.

30 En la etapa 640, un sistema informático puede calcular un valor respectivo del respectivo grupo de moléculas de ácido nucleico para cada una de la pluralidad de regiones cromosómicas. El valor respectivo puede definir una propiedad de las moléculas de ácido nucleico del respectivo grupo. La propiedad puede ser un recuento, un porcentaje o un tamaño de las moléculas de ácido nucleico. El valor respectivo puede incluir una media de una distribución de tamaño, una mediana de la distribución de tamaño, un modo de distribución de tamaño o una proporción de moléculas de ácido nucleico que tienen un tamaño inferior a un umbral de tamaño. El uso del tamaño como propiedad se analiza con mayor detalle en la sección IV.

40 En la etapa 650, el valor respectivo se puede comparar con un valor de referencia respectivo para determinar una clasificación de si la región cromosómica presenta una eliminación o una amplificación. La comparación puede incluir la determinación de una puntuación z basada en el valor respectivo y el valor de referencia respectivo. A continuación, la puntuación z se puede comparar con uno o más valores liminares para determinar si existe una eliminación o una amplificación. Se pueden utilizar diferentes umbrales para una eliminación y una amplificación. En otras realizaciones, el valor de referencia puede incluir el valor liminar, por ejemplo, si los otros valores en la puntuación z se movieran al otro lado de la ecuación. Un valor de referencia puede corresponder a un valor determinado en una muestra sana, otra región cromosómica (por ejemplo, una que no presente una anomalía) o el otro haplotipo cuando la región que se está probando es un primer haplotipo.

50 En la etapa 660, se puede determinar un patrón de prueba de las regiones cromosómicas que presentan una eliminación o amplificación. El patrón de prueba se refiere al patrón de regiones anómalas en la muestra que se está probando. El patrón de prueba puede incluir un conjunto de regiones cromosómicas que presentan una eliminación, una amplificación o son normales. El patrón de prueba también puede incluir un primer subconjunto del conjunto que se identifica como que presenta una amplificación. El patrón de prueba puede incluir además un segundo subconjunto del conjunto que se identifica como que presenta una eliminación. El patrón de prueba puede incluir además un tercer subconjunto del conjunto que se identifica como que no presenta una amplificación o una eliminación.

60 En la etapa 670, el patrón de prueba puede compararse con una pluralidad de patrones de referencia de diferentes tipos de cáncer. Un patrón de referencia para un tipo de cáncer puede incluir un conjunto conocido de regiones anómalas. Los patrones de referencia pueden determinarse a partir de muestras de referencia de tejidos y/o mezclas de moléculas de ácido nucleico extracelulares. El patrón de referencia puede incluir varias regiones, cada una de ellas con un estado definido de amplificación, eliminación o sin anomalía. La comparación puede determinar qué regiones del patrón de prueba tienen la misma anomalía que las regiones de un patrón de referencia. Por ejemplo, se puede determinar si se indica que la misma región tiene una amplificación, una eliminación o es normal tanto en el patrón de prueba como en un patrón de referencia.

En la etapa 680, basándose en la comparación, se puede determinar una cantidad de regiones del patrón de prueba que presentan una misma eliminación o amplificación que un primer patrón de referencia correspondiente a un primer tipo de cáncer. En diversas realizaciones, la cantidad puede ser un número o porcentaje de regiones cromosómicas que coinciden con el conjunto conocido de regiones anómalas.

En la etapa 690, la cantidad de regiones se compara con un primer umbral para determinar una primera clasificación de si la muestra biológica presenta el primer tipo de cáncer. El primer umbral puede ser específico para el primer tipo de cáncer o utilizarse para múltiples tipos de cáncer. Dicho umbral puede ser una cantidad mínima de regiones cromosómicas necesarias para que coincidan con el conjunto conocido de regiones anómalas para el primer tipo de cáncer que se va a identificar. En diversas realizaciones, la cantidad mínima puede ser de 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 regiones cromosómicas. En algunas realizaciones, puede necesitarse que regiones específicas sean anómalas y, por lo tanto, se pueden utilizar otros criterios además de la comparación de la cantidad con el primer umbral. Dichas regiones específicas pueden ser una restricción o tener mayor ponderación que otras regiones. Las regiones anómalas específicas pueden ser un subconjunto del conjunto completo de regiones anómalas conocidas para un tipo de cáncer. El tipo de cáncer puede incluir CHC, cáncer colorrectal, cáncer de mama, cáncer de pulmón o carcinoma nasofaríngeo, entre otros tipos de cáncer.

Un valor liminar utilizado para determinar la clasificación puede variar según las ubicaciones y los tamaños de las regiones que se cuentan. Por ejemplo, la cantidad de regiones de un cromosoma particular o de un brazo de un cromosoma se puede comparar con un umbral para ese cromosoma (o brazo) en particular como criterio para determinar si está implicado un tipo particular de cáncer. Se pueden utilizar varios umbrales. A modo de ejemplo, puede requerirse que la cantidad de regiones coincidentes (es decir, la misma clasificación de anomalía en el patrón de prueba y el patrón de referencia) en un cromosoma particular (o brazo o región subcromosómica más grande) sea mayor que un primer valor liminar, y puede requerirse que la cantidad total de regiones coincidentes en el genoma sea mayor que un segundo valor liminar.

El valor liminar para la cantidad de regiones coincidentes también puede depender de qué tan fuerte sea el desequilibrio para la clasificación de las regiones. Por ejemplo, la cantidad de regiones coincidentes que se utilizan como umbral para determinar una clasificación de un tipo de cáncer puede depender de la especificidad y la sensibilidad (umbral anómalo) utilizadas para detectar una anomalía en cada región. Por ejemplo, si el umbral anómalo es bajo (por ejemplo, una puntuación z de 2), entonces el umbral de cantidad puede seleccionarse como alto (por ejemplo, 15 regiones coincidentes o el 80 %). Pero, si el umbral anómalo es alto (por ejemplo, una puntuación z de 3), entonces el umbral de cantidad puede ser menor (por ejemplo, 5 regiones coincidentes o el 60 %). La cantidad de regiones que muestran una anomalía también puede ser un valor ponderado, por ejemplo, una región que muestre un gran desequilibrio puede tener una ponderación más elevada que una región que solo muestre un pequeño desequilibrio (es decir, hay más clasificaciones que solo positiva y negativa para la anomalía). Dicha ponderación puede actuar de manera similar a ciertas regiones que deben tener una anomalía para que se identifique el tipo de cáncer.

En algunas realizaciones, el umbral se puede determinar dinámicamente en función del número de regiones coincidentes con otros tipos de cáncer. Por ejemplo, el umbral puede ser que el número de regiones coincidentes para el cáncer identificado sea al menos un número específico mayor que las regiones coincidentes para el siguiente tipo de cáncer más probable. Dicho umbral puede ser un criterio adicional además del umbral mínimo. Por tanto, en algunos casos, no se podrá identificar ningún tipo de cáncer si no existe un número suficiente de regiones coincidentes.

B. Resultados

Se probó el método 600 para una pluralidad de tipos de cáncer para determinar la precisión. El método 600 se probó con pacientes con un tipo de cáncer conocido. Además, los umbrales utilizados pueden determinarse con muestras de tipos de cáncer conocidos. Se pueden utilizar diferentes umbrales para distintos tipos de cáncer.

Se secuenció el ADN plasmático de cada uno de los 17 pacientes con cáncer (6 pacientes con CHC, 4 con cáncer colorrectal (CCR), 3 con cáncer de mama (CM), 2 pacientes con cáncer de pulmón (CP) y 2 pacientes con carcinoma nasofaríngeo (CNF)). Se analizaron las anomalías en el número de copias (ANC) para cada brazo cromosómico de cada paciente según el enfoque CAZA.

En la figura 7 se muestran brazos cromosómicos que presentan patrones diferentes para diferentes tipos de cánceres en la tabla 700 de acuerdo con las realizaciones. Las ANC que ocurren en ≥ 50 % de los casos están resaltadas en color. Las pérdidas del número de copias se resaltan en rojo y las ganancias del número de copias se resaltan en verde.

La tabla 700 muestra el brazo cromosómico en la columna más a la izquierda. Cada una de las otras columnas enumera el tipo de cáncer y un número de paciente para el tipo de cáncer. Una eliminación se identifica con '-'. Una amplificación se identifica con '+'. Una región normal se identifica con 'Nulo'.

Los patrones de ANC observados en las muestras de plasma son diferentes para los pacientes que padecen distintos tipos de cáncer. En función de los patrones más comunes de ANC observadas en el plasma, las realizaciones pueden deducir el posible origen tisular de los cánceres en pacientes con ANC observadas en plasma, pero se desconoce la fuente de las ANC. Los patrones de ANC enumerados en la tabla 700 tienen fines ilustrativos, y se puede establecer una lista más completa de ANC mediante el análisis de un número mucho mayor de muestras clínicas.

Los patrones de referencia de las ANC también pueden determinarse a partir del análisis del tejido tumoral. Como ejemplos, ganancias en 1q, 6p, 8q y 17q, y pérdidas en 4q, 8p, 13q, 16q y 17p se detectan normalmente en tejidos tumorales de CHC (Moinzadeh P. *et al.* Br J Cancer 2005;92:935-941). Ganancias en 5p, 8q, 9p, 13q y 20q, y pérdidas en 8p, 9p, 17p y 18q se detectan normalmente en tejidos tumorales de CRC (Farzana *et al.* PLoS One 2012;2:231968 y Lips E. H. *et al.* J Pathol 2007;212:269-77). Ganancias en 5p, 7p, 7q, 8q, 14q, 17q y 20q, y pérdidas en 3p, 8p, 9p, 13q y 18q se detectan normalmente en tejidos de cáncer de pulmón no microcítico, mientras que las ganancias en 3q, 5p, 14q y 19q, y pérdidas en 3p, 4p, 4q, 5q, 10p, 10q, 13q, 15q, 17p y 22q se detectan normalmente en tejidos de cáncer de pulmón microcítico (Zhao X. *et al.* Cancer Res 2005;65:5561-70). Ganancias en 1q, 8q, 17q y 20q, y pérdidas en 4p, 5q, 8p, 11q y 13q son comunes en los tejidos de cáncer de mama (Andre F. *et al.* Clin Cancer Res 2009;15:441-51). Los patrones de ANC descritos en el presente documento sirven como ejemplos ilustrativos y no pretenden ser los únicos patrones que se pueden utilizar en los métodos descritos en el presente documento.

En función de los patrones ANC de este ejemplo, se supone que se realizó la secuenciación del ADN plasmático del paciente CM2 con el fin de detectar el cáncer. Se observaron ANC, incluidas ganancias en el número de copias para 1q, 3q, 8q y 14q y pérdidas en el número de copias para 2p, 2q, 3p, 4p, 7q, 8p, 9p, 11p, 12p, 12q, 16q y 17p. Las ANC en su plasma coincidieron con 13 ANC típicas para cánceres de mama. En cambio, sus ANC solo coincidieron con 3, 6, 4 y 1 ANC típicas de CHC, CCR, CP y CNF, respectivamente. Por consiguiente, en función del patrón ANC de su ADN plasmático, se deduce que el cáncer más probable que tiene es el cáncer de mama. El umbral seleccionado se puede utilizar para determinar si el número de ANC observado es compatible con las ANC típicas de determinados tipos de cáncer. En este ejemplo, se puede utilizar un umbral de 7, 8, 9, 10, 11, 12 o 13 para clasificar las ANC como compatibles con el cáncer de mama. También se puede utilizar un porcentaje de regiones coincidentes. Por ejemplo, se puede utilizar un porcentaje de regiones que coincidan con las regiones comúnmente anómalas. Las regiones comúnmente anómalas se pueden definir como regiones que tienen una anomalía particular en más de un 50 % de las muestras de referencia.

En otras realizaciones, se pueden realizar otros enfoques estadísticos, por ejemplo, pero sin limitación, la agrupación jerárquica, para deducir el tipo de cáncer más probable que tiene un paciente. Por ejemplo, a cada muestra de referencia se le puede asignar una función de datos multidimensional, donde cada dimensión corresponde a una región diferente. En una implementación, a cada dimensión se le puede asignar un -1 (para una eliminación), 0 (normal) o un 1 (para una amplificación). Podrían ser posibles números más altos para diferentes niveles de amplificación. Las muestras de un tipo particular de cáncer se agruparán y se podrá asignar una nueva muestra a un grupo. El umbral podría corresponder a la métrica utilizada para determinar a qué grupo (si hay alguno) se debe asignar la nueva muestra, donde el grupo asignado corresponde al tipo de cáncer identificado para la muestra. Por ejemplo, un grupo puede tener un centroide correspondiente a regiones de los patrones de referencia del grupo compartidos por al menos un número predeterminado de patrones de referencia del grupo. El grupo puede incluir un límite que define qué patrones de prueba se encuentran dentro del grupo. El límite puede tener varias formas más allá de la simplemente esférica. El límite se puede determinar como parte del análisis de agrupamiento cuando se determina qué patrones de referencia pertenecen a qué grupo, donde los patrones de referencia más alejados del centroide, pero dentro del grupo, pueden definir el límite. El umbral para determinar si un patrón de prueba es parte de un grupo puede considerarse la distancia desde el centroide hasta el límite en la dirección del centroide hasta el patrón de prueba.

En otra realización más, se puede determinar la probabilidad relativa de tener diferentes tipos de cáncer. El patrón de ANC de un paciente se puede comparar con la probabilidad de una ANC para cada tipo de cáncer. Por ejemplo, un paciente que tiene una ganancia en 1q se compararía con la probabilidad de la ganancia de 1q para diferentes tipos de cáncer. Para fines ilustrativos, se supone que se puede producir una ganancia en 1q en el 70 % de los pacientes con CHC, en el 20 % de los pacientes con CP y en el 1 % de los pacientes con CCR. Con estas probabilidades, se puede determinar una razón de probabilidades en función del porcentaje relativo de pacientes con diferentes tipos de cáncer que tienen la ANC. A modo de ejemplo, en función de la ganancia en 1q, se puede considerar que el paciente tiene 3,5 veces más probabilidades de tener CHC que CP y 70 veces más probabilidades de tener CHC que CCR. La razón de probabilidades de CHC a CP a CCR puede ser de 70:20:1. Una experto entendería que esta razón de probabilidades podría expresarse de varias maneras diferentes, aunque equivalentes. También se pueden determinar los razones de probabilidades para diferentes ANC en brazos cromosómicos distintos de 1q. A continuación, se puede calcular una razón de probabilidades general con las probabilidades o razones de probabilidades de las ANC individuales. En otras palabras, dado un patrón ANC de un paciente y las probabilidades de que diferentes tipos de cáncer tengan el patrón de ANC dado, las probabilidades de los diferentes tipos de cáncer se pueden comparar entre sí en una razón de probabilidades general. Aunque este ejemplo utilizó probabilidades de ANC en diferentes brazos cromosómicos, se pueden utilizar las probabilidades de ANC en diferentes regiones subcromosómicas distintas de los brazos cromosómicos. En algunas realizaciones, si no se encuentra ANC en un brazo cromosómico u otra región subcromosómica en un paciente, el patrón de ausencia de ANC se puede comparar con la probabilidad de no encontrar

una ANC en el brazo cromosómico o en la región subcromosómica para diferentes tipos de cáncer. El patrón de regiones sin ANC de un paciente se puede utilizar después para determinar la probabilidad de diferentes tipos de cáncer. Asimismo, la combinación del análisis de regiones con ANC y regiones sin ANC se puede utilizar para determinar la probabilidad o probabilidad relativa de un tipo de cáncer con una precisión potencialmente mayor que si solo se utiliza un tipo de región.

En otro ejemplo, se supone que el paciente CNF1 tiene el ADN plasmático secuenciado. Se observaron ANC, incluidas ganancias en el número de copias para 2q, 12q y 22q, y pérdidas en el número de copias para 6q y 18q. El patrón de ANC de este paciente coincidió con cuatro de las ANC típicas para CNF. En comparación, este patrón de ANC coincidió con 0, 2, 0 y 0 ANC típicas para los patrones de CHC, CCR, CM y CP. En otra realización, también se puede contabilizar la falta ANC típicas para un tipo de cáncer. Por ejemplo, en este paciente no se encontraron ninguna de las ANC típicas para CNF. En cambio, en este paciente no se encontraron 7, 16, 13 y 8 ANC típicas para CHC, CCR, CM y CP. Por consiguiente, el patrón ANC de este paciente no es indicativo de CHC, CCR, CM y CP.

En las figuras 8A, 8B y 8C se muestra cómo se puede mejorar aún más la precisión de este enfoque mediante el uso de un análisis de ANC de mayor resolución en la tabla 800. Se identificaron las ANC que afectan a las regiones de 1 Mb en esta cohorte de pacientes con cáncer. La tabla 800 tiene las coordenadas genómicas de las regiones de 1 MB enumeradas en la columna más a la izquierda. Cada una de las otras columnas enumera el tipo de cáncer y un número de paciente para el tipo de cáncer. Una eliminación se identifica con '-'. Una amplificación se identifica con '+'. Una región normal se identifica con 'Nulo'.

En este ejemplo, se identificaron las ANC que abarcaban 1 Mb y estaban presentes en todos los pacientes que tenían el mismo tipo de cáncer. Con la mayor resolución, se pueden identificar ANC subcromosómicas que están presentes en una alta proporción de pacientes con el mismo tipo de cáncer. Estas ANC específicas del tipo de cáncer no se identifican en el análisis basado en el brazo. Por ejemplo, se identificaron ganancias en el número de copias en el cromosoma 18 que abarca las coordenadas 30-31 Mb y 44-45 Mb en los tres pacientes con cáncer de pulmón, pero fueron poco comunes en pacientes con otros tipos de cáncer. Como se ha indicado anteriormente, se pueden utilizar diferentes pruebas estadísticas para determinar qué patrón de ANC específico del cáncer es más similar al caso analizado. Diferentes pruebas estadísticas pueden incluir, por ejemplo, el recuento del número de ANC típicas en diferentes patrones de ANC asociados al cáncer y agrupamiento jerárquico.

IV. ANÁLISIS DEL TAMAÑO DE FRAGMENTOS DE ADN PROCEDENTES DE TUMORES EN PLASMA

Una diferencia estadísticamente significativa en la distribución del tamaño de los fragmentos de ADN se puede utilizar para identificar una anomalía, de manera similar que el número de recuentos puede. Se ha informado que la distribución del tamaño del ADN plasmático total (es decir, tumoral más no tumoral) aumenta en pacientes con cáncer (Wang B. G., *et al.* Cancer Res. 2003; 63: 3966-8). Sin embargo, si se estudia específicamente el ADN procedente de tumores (en lugar de la cantidad total de ADN (es decir, el tumoral más el no tumoral)), entonces se ha observado que la distribución de tamaños de las moléculas de ADN procedentes de tumores es más corta que la de las moléculas procedentes de células no tumorales (Diehl *et al.* Proc Natl Acad Sci USA. 2005;102:16368-73). Por consiguiente, la distribución del tamaño del ADN circulante se puede utilizar para determinar si existen anomalías cromosómicas asociadas al cáncer.

El análisis del tamaño puede utilizar varios parámetros, como se menciona en el presente documento y en la patente de EE. UU. n.º 8.620.593. Por ejemplo, pueden utilizarse los valores Q o F anteriores. Estos valores de tamaño no necesitan una normalización por recuentos de otras regiones, ya que estos valores no aumentan con el número de lecturas. Se pueden utilizar técnicas que implican la profundidad y el perfeccionamiento de una región. En algunas realizaciones, al comparar dos regiones, puede tenerse en cuenta el sesgo de GC de una región concreta. En algunas implementaciones, el análisis de tamaño utiliza sólo moléculas de ADN.

A. Método

La figura 9 es un diagrama que ilustra un método 900 para analizar una muestra biológica de un organismo de acuerdo con realizaciones de la presente divulgación, pero que no forman parte de la invención reivindicada. La muestra biológica puede incluir moléculas de ácido nucleico originarias de células normales y potencialmente de células asociadas con el cáncer. Al menos parte de las moléculas de ácido nucleico puede ser extracelular en la muestra biológica. En un aspecto, el método 900 puede estar dirigido a determinar una clasificación de un desequilibrio de secuencia en función de un valor de separación (por ejemplo, una diferencia o relación) para el tamaño de los fragmentos de un primer cromosoma y el tamaño de los fragmentos de uno o más cromosomas de referencia.

En la etapa 910, para cada una de una pluralidad de moléculas de ácido nucleico en la muestra biológica, se puede medir el tamaño de la molécula de ácido nucleico. La obtención del tamaño de una molécula de ácido nucleico se describe en la publicación de patente de EE. UU. n.º 2013/0237431 titulada "Size-Based Analysis of Fetal DNA Fraction in Maternal Plasma" de Lo *et al.* presentada el 7 de marzo de 2013.

En la etapa 920, se puede identificar una ubicación de la molécula de ácido nucleico en un genoma de referencia del

organismo. La ubicación puede ser cualquier parte de un genoma, como se describe para la etapa 120 y en otros lugares. Por ejemplo, se identifica de qué cromosoma procede cada una de la pluralidad de moléculas de ácido nucleico. Esta determinación se puede realizar mediante la cartografía de un genoma de referencia.

- 5 En la etapa 930, para cada una de la pluralidad de regiones cromosómicas, se puede identificar un grupo respectivo de moléculas de ácido nucleico como perteneciente a una primera región cromosómica basándose en las ubicaciones identificadas. La primera región cromosómica puede incluir una pluralidad de primeros loci.

- 10 En la etapa 940, un sistema informático puede calcular un primer valor estadístico de una distribución de tamaño del primer grupo de moléculas de ácido nucleico. En unas realizaciones, el primer valor estadístico puede determinarse mediante el cálculo de un área bajo una primera curva de un tamaño específico. La primera curva puede ser un gráfico de una frecuencia acumulada de moléculas de ácido nucleico para la primera región cromosómica en un intervalo de tamaños. En una realización, el primer valor estadístico puede ser un promedio, media, mediana o moda de la distribución del tamaño de los fragmentos correspondientes al primer cromosoma. En otra realización, el primer valor estadístico puede incluir una suma de la longitud de los fragmentos por debajo de un primer tamaño, que puede ser un tipo de valor de corte. Por ejemplo, cada uno de los fragmentos que sean menores de 200 pb pueden tener sus longitudes sumadas. La suma se puede dividir por otro número, tal como una suma de las longitudes de todos los fragmentos correspondientes al primer cromosoma o una suma de las longitudes de los fragmentos mayores que un segundo valor de corte de tamaño (que puede ser el mismo que el primer tamaño). Por ejemplo, el primer valor estadístico puede ser una relación de la longitud total de los fragmentos por debajo de un primer valor de corte de tamaño en relación con la longitud total de los fragmentos, o una relación de la longitud total de los fragmentos pequeños en relación con la longitud total de los fragmentos grandes.

- 25 En la etapa 950, el primer valor estadístico puede compararse con un primer valor de referencia para determinar una clasificación de si la primera región cromosómica presenta una anomalía. En unas realizaciones, el primer valor de referencia puede ser un valor estadístico de una distribución de tamaño de un segundo grupo de moléculas de ácido nucleico de una segunda región cromosómica. La segunda región cromosómica puede considerarse una región cromosómica de referencia. El primer valor de referencia puede determinarse mediante el cálculo de un área bajo una segunda curva del tamaño específico. La segunda curva puede ser un gráfico de la frecuencia acumulada de moléculas de ácido nucleico para la segunda región cromosómica en el intervalo de tamaños. En una realización, el primer valor de referencia puede ser un valor estadístico para una pluralidad de cromosomas de referencia. En una implementación, los valores estadísticos se pueden combinar de tal manera que el valor estadístico pueda ser de uno o más segundos cromosomas. En otra realización, los valores estadísticos de la pluralidad de cromosomas de referencia pueden compararse individualmente. La comparación puede determinar una clasificación de si la primera región cromosómica presenta una eliminación o una amplificación.

- 40 El primer valor estadístico y el primer valor de referencia pueden compararse para obtener un valor de separación. En una realización, el valor de separación puede ser una diferencia entre el primer valor estadístico y el primer valor de referencia determinado. En otra realización, el valor de separación puede ser una relación entre el primer valor estadístico y el primer valor de referencia. En otra realización más, se pueden determinar varios valores de separación, por ejemplo, uno para cada valor de referencia, que se puede calcular para cada cromosoma de referencia.

- 45 El valor de separación puede ser una diferencia en la proporción de fragmentos cortos de ADN entre la primera región cromosómica y la región cromosómica de referencia utilizando la siguiente ecuación:

$$\Delta F = P(\leq 150 \text{ pb})_{\text{prueba}} - P(\leq 150 \text{ pb})_{\text{ref}}$$

- 50 donde $P(\leq 150 \text{ pb})_{\text{prueba}}$ indica la proporción de fragmentos secuenciados originados en la primera región cromosómica con tamaños $\leq 150 \text{ pb}$ y $P(\leq 150 \text{ pb})_{\text{ref}}$ indica la proporción de fragmentos secuenciados originados en la región cromosómica de referencia con tamaños $\leq 150 \text{ pb}$. En otras realizaciones, se pueden utilizar otros umbrales de tamaño, por ejemplo, pero sin limitación, 100 pb, 110 pb, 120 pb, 130 pb, 140 pb, 160 pb y 166 pb. En otras realizaciones, los umbrales de tamaño pueden expresarse en bases, nucleótidos u otras unidades. En algunas implementaciones, la región cromosómica de referencia se puede definir como todas las regiones subcromosómicas excluida la primera región cromosómica. En otras implementaciones, la región de referencia puede ser solo una parte de las regiones subcromosómicas excluida la primera región cromosómica.

- 60 Los mismos grupos de controles utilizados en el análisis basado en recuento se pueden utilizar en el análisis basado en tamaño. Se puede calcular una puntuación z basada en el tamaño de la región analizada utilizando los valores de media y desviación típica de ΔF de los controles:

$$\text{Puntuación z basada en tamaño} = \frac{\Delta F_{\text{muestra}} - \text{media } \Delta F_{\text{control}}}{\text{DT } \Delta F_{\text{control}}}$$

- 65 El valor de separación puede compararse con uno o más valores de corte. En una realización, la comparación se puede realizar para cada uno de una pluralidad de valores de separación. Por ejemplo, se puede determinar un valor de separación diferente entre el primer valor estadístico y cada valor de referencia. En diversas implementaciones,

cada valor de separación se puede comparar con el mismo valor de corte o con valores de corte diferentes. En otra realización, se compara un valor de separación con dos valores de corte para determinar si el valor de separación está dentro de un intervalo particular. El intervalo puede incluir un valor de corte para determinar si ocurre una función de datos no normal (por ejemplo, una anomalía) y se podría usar un segundo valor de corte para determinar si es probable que la función de datos esté provocada por un error en la medición o el análisis (por ejemplo, si el valor de separación es mayor de lo que se esperaría, incluso para una muestra enferma).

Se determina una clasificación para determinar si existe un desequilibrio de secuencia (por ejemplo, una anomalía) para la primera ubicación genómica en función de la comparación. En una realización, se pueden utilizar varios valores de corte (por ejemplo, N valores de corte) para un único valor de separación. En dicha realización, se pueden determinar N+1 clasificaciones. Por ejemplo, se pueden utilizar dos valores de corte para determinar la clasificación de si la región cromosómica es normal o sana, indeterminada o aberrante (por ejemplo, amplificación o eliminación). En otra realización en la que se realizan varias comparaciones (por ejemplo, una para cada valor de separación), la clasificación puede basarse en cada una de las comparaciones. Por ejemplo, un método basado en reglas puede analizar las clasificaciones resultantes de cada una de las comparaciones. En una implementación, una clasificación definitiva sólo se proporciona cuando todas las clasificaciones son coherentes. En otra implementación, se utiliza la clasificación mayoritaria. En aún otra implementación, se puede utilizar una fórmula más complicada en función de qué tan cerca esté cada uno de los valores de separación de un valor de corte respectivo, y estos valores de cercanía se pueden analizar para determinar una clasificación. Por ejemplo, los valores de cercanía podrían sumarse (junto con otros factores, tal como una normalización) y el resultado podría compararse con otro valor de corte. En otras realizaciones, también se pueden aplicar variaciones del método 900 a una comparación directa de un valor estadístico para el primer cromosoma con un valor de corte, que puede proceder de una muestra de referencia.

B. Correlación del tamaño con el cáncer

Para análisis adicionales, se exploró por separado moléculas de ADN plasmático de tres grupos de tamaños diferentes, en concreto, aquellos menores de 150 pb, aquellos entre 150 y 180 pb, y aquellos por encima de 180 pb. Existe una correlación positiva (r de Pearson = 0,6; (valor de $p < 0,001$) entre la proporción de fragmentos de ADN menores de 150 pb y la fracción de ADN tumoral en plasma (Figura 10A). La fracción de ADN tumoral en las figuras 10A, 10B y 10C se muestran en escala logarítmica. No se observó correlación ($r = -0,07$; valor de $p = 0,95$) entre la proporción de fragmentos de ADN con tamaños entre 150 y 180 pb y la fracción de ADN tumoral en plasma (Figura 10B). Se observó una correlación negativa ($r = -0,41$; valor de $p < 0,001$) entre la proporción de ADN de más de 180 pb y la fracción de ADN tumoral en plasma (Figura 10C).

Es más probable que se produzca una fracción menor de ADN tumoral en las primeras etapas del cáncer, y es más probable que se produzca una fracción mayor de ADN tumoral en las etapas posteriores del cáncer. Por tanto, la existencia de un tamaño promedio mayor (u otro valor estadístico) que el normal para los fragmentos de ADN puede indicar un cáncer en etapa temprana, y la existencia de un tamaño promedio menor que el normal para los fragmentos de ADN indica un cáncer en etapa posterior.

En otras realizaciones, se puede medir la fracción de ADN tumoral. Cuando la fracción de ADN tumoral está por debajo de un cierto umbral, se puede realizar un análisis de tamaño para determinar si un valor estadístico de una distribución de tamaño es mayor que un umbral (es decir, probar si los fragmentos de ADN son largos). Cuando la fracción de ADN del tumor está por encima de un cierto umbral, se puede realizar un análisis de tamaño para determinar si un valor estadístico de una distribución de tamaño es menor que un umbral (es decir, probar si los fragmentos de ADN son cortos).

Los métodos de análisis de tamaño y los datos relativos a la relación del tamaño con el cáncer se analizan en la publicación de patente de EE. UU. n.º 2013/0040824 titulada "Detection of Genetic or Molecular Aberrations Associated with Cancer" de Lo *et al.* presentada el 30 de noviembre de 2011.

V. CONFIRMACIÓN DE LA ANOMALÍA ANC CON ANÁLISIS DE TAMAÑO

Se utilizó secuenciación masiva en paralelo para estudiar los perfiles de tamaño de muestras de ADN plasmático con una resolución de base única y a nivel de todo el genoma. Se utilizó CAZA para identificar ADN plasmático procedente de tumores para estudiar sus perfiles de tamaño específicos.

En este estudio, se utilizó el enfoque CAZA para identificar brazos cromosómicos que mostraban anomalías cuantitativas del ADN plasmático que sugieren la presencia de ANC asociadas a tumores. Después de identificar los brazos cromosómicos con amplificaciones o eliminaciones, se pone el foco en estas regiones como estrategia para comparar el ADN plasmático procedente de tumores (enriquecido en las regiones amplificadas) y el no procedente de tumores (enriquecido en las regiones eliminadas). Se cree que este enfoque puede proporcionar un medio más sólido para identificar ADN tumoral por el análisis de perfiles de tamaño que el basado en la detección de mutaciones asociadas al cáncer. Para esta última, en promedio, se ha informado que existen del orden de miles de mutaciones puntuales en los genomas del cáncer (29-32, 39). Para CAZA, por otra parte, sería útil cualquiera de la innumerables moléculas de ADN plasmático procedentes de las regiones genómicas que presentan ANC, cuyo total asciende a

decenas de megabases.

A. Análisis combinado

En la figura 11 se muestra una ilustración esquemática del principio del análisis del tamaño del ADN plasmático en pacientes con cáncer. En la figura 11 se muestran las etapas 1110 a 2150. La etapa 1110 muestra las células de los tejidos en plasma. Las células tumorales pueden incluir amplificaciones y/o eliminaciones en varias regiones, como se ha descrito anteriormente. El ejemplo muestra una región amplificada en un cromosoma particular y otra región eliminada.

En la fase 1120, el plasma se muestra con contribuciones de varias regiones. En la muestra de plasma se muestran fragmentos de ADN. En pacientes con cáncer, el ADN plasmático procede tanto de células tumorales (moléculas rojas) como de células no tumorales (moléculas azules). Las regiones genómicas que se amplifican en el tejido tumoral aportarían más ADN tumoral al plasma. Las regiones genómicas eliminadas en el tejido tumoral aportarían menos ADN al plasma.

En la fase 1130, se realiza la secuenciación de extremos emparejados. La secuenciación de extremos emparejados se puede utilizar para determinar el tamaño de los fragmentos de ADN en la muestra de plasma.

En la fase 1140, se utiliza un análisis basado en recuentos para identificar regiones anómalas. En el ejemplo mostrado, se utilizó un análisis CAZA para determinar si un brazo cromosómico está sobrerrepresentado o infrarrepresentado en el ADN plasmático, sugestivo de la presencia de amplificación o eliminación del brazo cromosómico en el tumor. Una puntuación z positiva grande puede indicar la presencia de una amplificación del brazo cromosómico, mientras que una puntuación z negativa grande puede indicar la presencia de una eliminación del brazo cromosómico. Se pueden utilizar otros tamaños de regiones además de los brazos.

En la fase 1150, se puede analizar la distribución del tamaño de una región de prueba. Como se ha explicado anteriormente, los fragmentos de ADN tumoral son más cortos que los fragmentos de ADN de las células sanas. Se pueden analizar los fragmentos de ADN de una región anómala para confirmar que el análisis de tamaño también muestra la misma anomalía. En el ejemplo mostrado, se compara una distribución de tamaño de una región que presenta una amplificación con una distribución de tamaño de una región que presenta una eliminación. Por tanto, en algunas realizaciones, se pueden comparar los perfiles de tamaño de las moléculas de ADN plasmático que se originan en brazos cromosómicos que están infrarrepresentados (enriquecidos con ADN no tumoral) y sobrerrepresentados (enriquecidos con ADN procedente de tumores), como se describe con mayor detalle a continuación.

B. Diferencia de tamaño entre dos regiones

Para comparar los perfiles de tamaño del ADN plasmático procedente de tejidos tumorales y no tumorales, se analizaron los fragmentos de ADN plasmático de los brazos cromosómicos con ANC. En función de estudios previos (34-36) así como los hallazgos en este estudio, las ANC típicas asociadas con el CHC incluyen eliminaciones de 1p y 8p y amplificaciones de 1q y 8q. Para ilustrar el principio se utiliza un caso de CHC (H291) con el 53 % de ADN procedente del tumor en el plasma. Este caso mostró eliminación de 8p y amplificación de 8q en plasma. Por tanto, el tumor liberaría más ADN plasmático de la región amplificada de 8q que de la región eliminada de 8p. Como resultado, 8q estaría relativamente enriquecido con ADN procedente de tumores y 8p estaría relativamente desprovisto de ADN tumoral (o en otras palabras, relativamente enriquecido con ADN no tumoral) en comparación con regiones sin ANC. Los perfiles de tamaño del ADN plasmático para 8p y 8q se muestran en la figura 12A. El perfil de tamaño para 8q estaba en el lado izquierdo del de 8p, lo que indica que la distribución del tamaño del ADN plasmático para 8q era más corta que para 8p. Debido a que 8q está enriquecido con ADN tumoral, los datos sugieren que el ADN liberado por el tumor tiende a ser más corto que el ADN que no se origina en el tumor.

Para cuantificar el grado de acortamiento, se construyeron gráficos de frecuencia acumulativa (Figura 12B) para los perfiles de tamaño de 8p y 8q para cada muestra de plasma. Estos gráficos muestran la acumulación progresiva de moléculas de ADN, de tamaños cortos a largos, como proporción de todas las moléculas de ADN plasmático en la muestra. La diferencia entre las dos curvas ΔS (Figura 12C) se calculó entonces como

$$\Delta S = S_{8q} - S_{8p}$$

donde ΔS representa la diferencia en las frecuencias acumuladas entre 8p y 8q en un tamaño particular, y S_{8p} y S_{8q} representan las proporciones de fragmentos de ADN plasmático menores de un tamaño particular en 8p y 8q, respectivamente. Un valor positivo de ΔS para un tamaño particular indica una mayor abundancia de ADN más corto que ese tamaño particular en 8q en comparación con 8p. Con este método, se analiza los valores de ΔS de 50 pb a 250 pb para todos los casos de CHC que mostraron ANC en 8p y 8q en plasma. La diferencia en las frecuencias acumuladas, ΔS , entre 8q y 8p para el caso de CHC H291 se representa como una línea roja en la figura 12C. En comparación con los controles sanos (líneas grises), todos estos casos de CHC mostraron una mayor abundancia de ADN plasmático de menos de 200 pb procedente de 8q (enriquecido con ADN tumoral) que de 8p (enriquecido con

ADN no tumoral) (Figura 13A). En la figura 13A se muestra un gráfico de ΔS frente al tamaño de todos los casos de CHC con diferentes ANC en 8p y 8q en plasma. Los casos con diferentes intervalos de concentraciones fraccionarias de ADN tumoral en plasma se muestran en diferentes colores. A medida que aumenta la concentración fraccionaria de ADN tumoral, la ΔS aumenta, lo que indica una mayor abundancia de fragmentos de ADN más cortos. Estos datos respaldan además que el ADN procedente de tumores era más corto que el ADN no procedente de tumores.

El valor de ΔS alcanzó un máximo a 166 pb, lo que sugiere que la diferencia clave entre el ADN plasmático procedente de tejidos tumorales y no tumorales es la abundancia relativa de ADN <166 pb y ≥ 166 pb. Se indica este valor como ΔS_{166} . La ΔS_{166} se mostró en un gráfico para todos los sujetos de este estudio, incluidos los portadores del VHB y los pacientes con cirrosis hepática (Figura 13B). Para el grupo de CHC, los pacientes con y sin diferentes ANC en 8p y 8q según lo determinado por el análisis CAZA plasmático están representados por puntos rojos y negros, respectivamente. Para casi todos los sujetos sin CHC, los valores de ΔS_{166} fueron cercanos a 0, lo que indica que las distribuciones de tamaño del ADN de 8p y 8q eran similares. La ΔS_{166} (o el valor en algún otro tamaño especificado) se puede comparar con un umbral, y si la diferencia excede el umbral, entonces al menos una de las regiones puede identificarse como que presenta una anomalía. Si se sabe que una región no tiene una anomalía (por ejemplo, a partir del análisis de ANC), entonces la otra región se identificaría como que presenta una anomalía cuando la diferencia excede un umbral. En dicha realización, el signo de la diferencia puede indicar el tipo de anomalía. Por ejemplo, cuando la primera región tiene una amplificación y la segunda región no, entonces la diferencia sería un número positivo. Cuando la primera región tiene una eliminación y la segunda región no, entonces la diferencia sería un número negativo. Si se determina una anomalía, entonces ambas regiones se pueden identificar como potencialmente portadoras de una anomalía, con el signo indicando el tipo de anomalía que puede tener cada región. Si la diferencia es suficientemente grande, puede indicar que una región tiene una amplificación y la otra región tiene una eliminación (o la cantidad de amplificación es diferente), ya que entonces la diferencia sería mayor en una región amplificada en comparación con una región normal. El análisis del número de copias puede proporcionar una clasificación inicial para las regiones, para que se pueda elegir un umbral adecuado.

También se realizó un análisis de tamaño basado en los perfiles de tamaño del ADN plasmático de 1p y 1q (Figuras 14 y 15) y mostró la misma tendencia. En la figura 15, para el grupo de CHC, los pacientes con y sin diferentes ANC en 1p y 1q según lo determinado por el análisis CAZA plasmático están representados por puntos rojos y negros, respectivamente. Este análisis de tamaño se puede realizar utilizando una región amplificada en una región normal, o una región normal y una región eliminada.

En otra realización, se puede comparar una distribución de tamaño para una región amplificada o eliminada con una distribución de tamaño de uno o más sujetos de referencia que se sabe que tienen cáncer o que están sanos. La distribución de tamaño se puede representar mediante un valor, por ejemplo, un valor estadístico, tal como un tamaño medio o mediano.

Por consiguiente, la anomalía de una región cromosómica se puede utilizar para seleccionar regiones particulares para un análisis de tamaño. El análisis del tamaño de las regiones seleccionadas se puede utilizar entonces para determinar una clasificación de un nivel de cáncer. La combinación del uso de ANC y el análisis del tamaño puede proporcionar una mayor precisión. El análisis de ANC puede ocasionalmente producir falsos positivos, es decir, pacientes que no tienen cáncer, pero que tienen regiones con anomalía en el número de copias. Por tanto, un paciente que se identifica como portador de cáncer debido a una cantidad suficiente de regiones que presentan anomalía puede luego confirmarse mediante un análisis de tamaño. En una realización, las regiones seleccionadas son las que tienen amplificación.

Este estudio fue diseñado con la intención de explorar el perfil de tamaño del ADN plasmático de pacientes con CHC de una manera integral y de alta resolución que pueda arrojar luz sobre los mecanismos relacionados con la generación o liberación de ADN plasmático por los tejidos tumorales. Otro objetivo del estudio fue resolver algunas de las aparentes incoherencias que existían en la bibliografía respecto a los perfiles de tamaño del ADN plasmático asociados al cáncer. Unos estudios han informado la presencia de ADN más largo en el plasma de pacientes con cáncer (20-23) mientras que otros informaron una mayor prevalencia de mutaciones de ADN asociadas al cáncer entre las moléculas de ADN plasmático más cortas (12, 25). Para lograr estos objetivos de estudio, se adoptó un enfoque de dos etapas. En primer lugar, se midieron las longitudes de todas las moléculas de ADN en muestras de plasma de los sujetos reclutados mediante el uso de secuenciación masiva en paralelo de extremos pareados. Este enfoque permite determinar la longitud de moléculas individuales de ADN plasmático hasta una resolución de una sola base. Además, se podrían analizar las moléculas de ADN plasmático en todo el genoma y se podrían determinar las cantidades relativas entre ADN de diferentes tamaños con alta precisión. Por lo tanto, se podría obtener un estudio amplio y profundo del perfil de tamaño del ADN plasmático. En segundo lugar, se aprovechó la diferencia relativa en el contenido de ADN tumoral en el ADN plasmático originado en ubicaciones genómicas que estaban asociadas con amplificaciones o eliminaciones, la estrategia CAZA, como un medio para identificar ADN plasmático procedente de tumores para un análisis detallado.

Este estudio proporciona una serie de conocimientos sobre los mecanismos biológicos que podrían estar implicados en la liberación de ADN plasmático. El ADN plasmático de todos los sujetos reclutados, incluidos los portadores del VHB, los pacientes con cirrosis hepática o CHC, presentó un pico prominente a 166 pb (Figuras 14 y 16). Este patrón

es análogo a las observaciones en el plasma de mujeres embarazadas y receptores de trasplantes de órganos (26, 27). La presencia del pico característico a 166 pb en el perfil de tamaño del ADN plasmático de todos los grupos de pacientes estudiados sugiere que la mayoría de las moléculas de ADN circulantes en el plasma humano, incluida las de las mujeres embarazadas, los receptores de trasplantes, los pacientes con CHC, cirrosis hepática o VHB crónico, se asemejan a unidades mononucleosomales y es probable que se originen a partir del proceso de apoptosis.

El estudio del perfil de tamaño de las moléculas de ADN plasmático que contienen ANC asociadas a tumores indica que dichas moléculas son más cortas que aquellas que no llevan dichas características (Figura 13). Esto es coherente con la presente observación de que con el aumento de las concentraciones fraccionarias de ADN tumoral en el plasma, el perfil de tamaño del ADN plasmático se desplazaría hacia la izquierda. Sin embargo, el hecho de que los pacientes con CHC con bajas concentraciones fraccionarias de ADN tumoral en plasma tuvieran una distribución de tamaño aparentemente más larga que los controles sanos sugiere que había un componente adicional del ADN plasmático que no portaba las características genómicas asociadas al tumor. Es posible que este componente provenga de los tejidos hepáticos no neoplásicos que rodean el tumor. Estas moléculas de ADN largas podrían proceder de la necrosis en lugar de la apoptosis. Se ha informado que la muerte celular asociada con la necrosis tisular puede generar fragmentos de ADN más largos además de los fragmentos de ADN oligonucleosómicos típicos (37, 38). Para futuros estudios, sería interesante estudiar el perfil de metilación del ADN de estas moléculas de ADN más largas para ver si tienen similitudes con el esperado para el hígado.

Se demostró que poblaciones de moléculas de ADN anómalo cortas y largas coexistían en el plasma de pacientes con carcinoma hepatocelular. Las cortas portaban preferentemente las anomalías en el número de copias asociadas al tumor.

Resumiendo, se hizo el perfil de la distribución del tamaño del ADN plasmático en pacientes con CHC con resolución de un solo nucleótido. Se demostró una diferencia en el tamaño del ADN plasmático procedente de tejidos tumorales y no tumorales.

La relación entre ΔS y el tamaño del tumor también se analizó. Las muestras de ADN plasmático de 10 pacientes con CHC con eliminación de 8p y amplificación de 8q en plasma se analizaron utilizando el análisis de ΔS . La ΔS se determinó para la diferencia de tamaño entre los fragmentos de ADN plasmático asignados a 8p y 8q. Un valor positivo para ΔS indica la mayor abundancia de fragmentos de ADN cortos por debajo de 150 pb para 8q en comparación con 8p. En la figura 30, los valores de ΔS se representaron gráficamente frente a la dimensión más larga del tumor de los pacientes con CHC.

Se observó una correlación positiva entre ΔS y el tamaño del tumor ($r = 0,876$, correlación de Pearson). Esta observación sugiere que la distribución del tamaño de los fragmentos de ADN plasmático de regiones que presentan diferentes tipos de ANC se puede utilizar para reflejar el tamaño del tumor en pacientes con CHC.

También se analizó la distribución general del tamaño del ADN plasmático total de estos 10 pacientes con CHC. Se determinó el porcentaje de fragmentos de ADN plasmático de menos de 150 pb ($P(<150)$) para cada caso y se representó gráficamente frente al tamaño del tumor en la figura 31. La proporción de fragmentos cortos fue significativamente mayor en pacientes con cáncer de mayor tamaño, de más de 3 cm en la dimensión más grande. En una realización, la proporción de fragmentos cortos se puede utilizar para reflejar el tamaño y la gravedad del cáncer. En otras implementaciones, se pueden utilizar otros valores de corte para el tamaño, por ejemplo, pero sin limitación, 100 pb, 110 pb, 120 pb, 130 pb, 140 pb, 160 pb y 166 pb.

Se puede utilizar una función de calibración para proporcionar una relación entre el tamaño del tumor y un valor estadístico. La función de calibración puede determinarse a partir de funciones de datos de calibración de muestras de referencia de organismos con tumores de tamaño conocido. La función de datos de calibración puede incluir una medición del tamaño del tumor y una medición estadística correspondiente de los tamaños de las moléculas de ácido nucleico de una región cromosómica. Cuando se obtiene una nueva muestra de un nuevo sujeto, se puede determinar el valor estadístico y se puede utilizar la función de calibración para convertir el valor estadístico en un tamaño de tumor. Un ejemplo de una función de calibración es un ajuste lineal, similar al ajuste lineal que se muestra en la figura 30. Otros tipos de análisis de regresión, tal como un ajuste de mínimos cuadrados, se pueden utilizar para generar la función de calibración.

La función de calibración se puede definir de diversas maneras, por ejemplo, como una pluralidad de coeficientes de una función especificada, tal como una función lineal o no lineal. Otras realizaciones pueden almacenar una pluralidad de funciones de datos de calibración (por ejemplo, funciones de datos de la función de calibración) de modo que se pueda generar la función de calibración. Además, se puede realizar una centrifugación entre dichas funciones de datos de calibración para obtener la función de calibración. La función de calibración puede almacenarse y recuperarse desde la memoria del ordenador.

C. Método

La figura 16 es un diagrama que ilustra un método 1600 para realizar CAZA y análisis de tamaño para analizar una

muestra biológica de un organismo de acuerdo con realizaciones de la presente invención.

En la etapa 1605, se pueden identificar una pluralidad de regiones cromosómicas de un organismo. Cada región cromosómica puede incluir una pluralidad de loci. Se puede seleccionar una de la pluralidad de regiones cromosómicas como primera región cromosómica. La identificación de la pluralidad de regiones cromosómicas puede ser similar a la etapa 610 de la figura 6.

En la etapa 1610, se puede identificar una localización de una molécula de ácido nucleico en un genoma de referencia del organismo para cada una de una pluralidad de moléculas de ácido nucleico. La identificación de la ubicación de la molécula de ácido nucleico se puede realizar de manera similar a la etapa 620 de la figura 6.

En la etapa 1615, se puede medir el tamaño de una molécula de ácido nucleico para cada una de la pluralidad de moléculas de ácido nucleico en la muestra biológica. El tamaño de la molécula de ácido nucleico se puede medir de manera similar a la etapa 910 de la figura 9.

En la etapa 1620, se puede identificar un grupo respectivo de moléculas de ácido nucleico, en función de las ubicaciones identificadas, como que es de una región cromosómica para cada región cromosómica de la pluralidad de regiones cromosómicas. El grupo respectivo puede incluir al menos una molécula de ácido nucleico ubicada en cada uno de la pluralidad de loci de la región cromosómica. La identificación del respectivo grupo de moléculas de ácido nucleico puede ser similar a la etapa 120 de la figura 1.

En la etapa 1625, un sistema informático puede calcular una cantidad respectiva del respectivo grupo de moléculas de ácido nucleico. El cálculo de la cantidad respectiva puede ser similar al cálculo de la etapa 130 de la figura 1.

En la etapa 1630, la cantidad respectiva puede compararse con un valor de referencia de recuento para determinar una clasificación de recuento de si la región cromosómica exhibe una amplificación. Basándose en la comparación, la primera región cromosómica puede identificarse como potencialmente que presenta una anomalía. Las etapas 1620 a 1630 se pueden realizar de manera similar a las etapas 120 a 140 de la figura 1 o las etapas 630 a 650 de la figura 6.

En la etapa 1640, un primer grupo de moléculas de ácido nucleico puede identificarse como de la primera región cromosómica.

En la etapa 1645, un sistema informático puede calcular un primer valor estadístico de una primera distribución de tamaño del primer grupo de moléculas de ácido nucleico. El primer valor estadístico puede determinarse mediante el cálculo de un área bajo una primera curva de un tamaño específico. La primera curva puede ser un gráfico de frecuencia acumulada de moléculas de ácido nucleico para la primera región cromosómica en un intervalo de tamaños. El cálculo del primer valor estadístico en la etapa 1645 puede ser similar al cálculo del primer valor estadístico en la etapa 940 de la figura 9.

En la etapa 1650, el primer valor estadístico puede compararse con un valor de referencia de tamaño para determinar una clasificación de tamaño de si la primera región cromosómica presenta una anomalía. El valor de referencia de tamaño se puede determinar mediante el cálculo de un área bajo una segunda curva en el tamaño especificado. La segunda curva puede ser un gráfico de la frecuencia acumulada de moléculas de ácido nucleico para la segunda región cromosómica en el intervalo de tamaños. La comparación puede basarse en una diferencia entre las dos curvas. En algunas realizaciones, comparar el primer valor estadístico con el valor de referencia de tamaño puede ser similar a la etapa 950 de la figura 9.

En la etapa 1655, se puede determinar una clasificación final de si la primera región cromosómica presenta una anomalía. Por ejemplo, se puede utilizar al menos una de las clasificaciones de tamaño y de recuento para determinar si existe anomalía para la primera región cromosómica. En algunas realizaciones, la clasificación final puede ser que la primera anomalía exista sólo cuando la clasificación de recuento y la clasificación de tamaño indiquen la misma anomalía. Por tanto, la comparación del primer valor estadístico con el valor de referencia de tamaño puede confirmar si la primera región cromosómica presenta una anomalía. En algunas realizaciones, se puede determinar un conjunto de clasificaciones de tamaño para un conjunto de regiones cromosómicas identificadas como anómalas según las clasificaciones de recuento correspondientes. En función del conjunto de clasificaciones de tamaño, cada una de las regiones cromosómicas puede confirmarse como anómala o no anómala.

En algunas realizaciones, la clasificación final de si la primera región cromosómica presenta una anomalía puede basarse en múltiples valores de referencia de recuento y múltiples valores de referencia de tamaño. Cada uno de los valores de referencia de recuento puede corresponder a una clasificación de recuento diferente (por ejemplo, una discriminación entre un par único de clasificación de recuento, tal como entre el nivel 1 y el nivel 2, o entre el nivel 2 y el nivel 3). De forma similar, cada uno de los valores de referencia de tamaño puede corresponder a una clasificación de tamaño diferente. La clasificación final se puede determinar a partir de la combinación particular de clasificación de tamaño y clasificación de recuento.

La clasificación de tamaño puede incluir múltiples clasificaciones dependiendo de un valor estadístico de la distribución de tamaño. Por ejemplo, una gran diferencia entre el valor estadístico y un valor de referencia de tamaño puede dar lugar a una clasificación de tamaño correspondiente a una probabilidad elevada de anomalía, mientras una diferencia pequeña entre el valor estadístico y el valor de referencia de tamaño puede dar lugar a una clasificación de tamaño correspondiente a una probabilidad baja de anomalía. De forma similar, la clasificación de recuento puede incluir múltiples clasificaciones dependiendo de la cantidad de un grupo de moléculas de ácido nucleico. Por ejemplo, una gran diferencia entre la cantidad de un grupo de moléculas de ácido nucleico en comparación con un valor de referencia de recuento puede dar como resultado una clasificación de recuento correspondiente a una probabilidad elevada de anomalía, mientras que una diferencia pequeña puede dar lugar a una clasificación de recuento correspondiente a una probabilidad baja de anomalía.

Por consiguiente, la clasificación final puede basarse en diferentes umbrales para diferentes clasificaciones de tamaño y clasificaciones de recuento. A modo de ejemplo, una clasificación de tamaño que indica una probabilidad elevada de una anomalía puede dar como resultado una clasificación final que indica una anomalía dada una clasificación de recuento que indica una cierta probabilidad, posiblemente baja, de una anomalía. A medida que aumenta la probabilidad de una anomalía, como lo indica una de la clasificación de tamaño o la clasificación de recuento, entonces se reduce el umbral de probabilidad indicado por la otra clasificación. En algunos casos, una clasificación puede mostrar una probabilidad elevada de un primer tipo de anomalía, la otra clasificación puede mostrar una probabilidad baja de un segundo tipo de anomalía y la clasificación final puede indicar que está presente el primer tipo de anomalía. En algunos casos, la clasificación final puede corresponder a una probabilidad o verosimilitud de anomalía.

D. Casos de ejemplo

La especificidad de la detección de ANC asociadas al cáncer se puede mejorar mediante el análisis del tamaño del ADN plasmático, como se muestra en los dos casos siguientes. El caso 1 fue un paciente con cirrosis asociada a hepatitis B y el caso 2 fue un portador crónico de infección de hepatitis B. Ninguno de los dos tenía conocimiento de tener algún cáncer en el momento del reclutamiento. Se les había hecho un seguimiento clínico durante dos años desde el reclutamiento y no se detectó cáncer. Se recogió sangre venosa de cada uno de los dos sujetos en el momento del reclutamiento. Se secuenció el ADN plasmático. En cada uno de estos dos pacientes se detectaron ANC que afectaba el cromosoma 1q. Para el caso 1, la puntuación z para 1p y 1q fueron -2,3 y 15,5, respectivamente. Estos resultados son coherentes con la interpretación de la amplificación de 1q. En el análisis del tamaño de los fragmentos de ADN plasmático, la ΔS fue -0,019. El valor negativo de ΔS indica que los fragmentos de ADN cortos fueron menos abundantes en 1q en comparación con 1p. Como el análisis basado en el recuento sugiere que 1q se amplificó, el resultado del análisis basado en el tamaño es opuesto a lo que esperábamos para las ANC asociadas al cáncer. En pacientes con cáncer, se espera que las regiones con ganancia en el número de copias muestren una distribución de tamaño general más corta debido a la presencia de más fragmentos cortos procedentes del cáncer en comparación con las regiones con amplificación o regiones sin ninguna ANC. Por consiguiente, el análisis de tamaño en este caso no sugiere la presencia de ANC asociadas al cáncer en el ADN plasmático.

Para el caso 2, las puntuaciones z para 1p y 1q fueron 0,4 y -4,4, respectivamente. Estos resultados son compatibles con la interpretación de la eliminación de 1q. En el análisis del tamaño de los fragmentos de ADN plasmático, la ΔS fue 0,044. El valor positivo de ΔS indica que los fragmentos de ADN cortos fueron más abundantes en 1q en comparación con 1p. Como el análisis basado en el recuento sugiere que se eliminó 1q, el resultado del análisis basado en el tamaño es opuesto a lo que esperábamos para las ANC asociadas al cáncer. En pacientes con cáncer, se espera que las regiones con pérdida en el número de copias muestren una distribución de tamaño general más largo debido a la presencia de menos fragmentos cortos procedentes del cáncer en comparación con las regiones con amplificación o regiones sin ninguna ANC. Por consiguiente, el análisis de tamaño en este caso no sugiere la presencia de ANC asociadas al cáncer en el ADN plasmático.

VI. DETERMINACIÓN DE ESTADIOS DEL CÁNCER

Como se ha mencionado anteriormente, el tamaño de los fragmentos de ADN puede indicar el estadio del cáncer. Una etapa posterior del cáncer presenta fragmentos más pequeños para las regiones que presentan amplificación.

Aparte del interés biológico intrínseco, el perfil del tamaño del ADN plasmático también puede ser útil para el desarrollo de enfoques de diagnóstico para detectar cambios asociados al cáncer en el plasma. Por ejemplo, el enriquecimiento del ADN tumoral del plasma puede lograrse centrándose en el análisis de fragmentos cortos de ADN. Asimismo, se observó que la proporción de moléculas de ADN cortas tenía una relación positiva con la concentración fraccionaria de ADN procedente del tumor en el plasma. Los cambios en los perfiles de tamaño se pueden utilizar para el seguimiento de los pacientes durante el curso del tratamiento. Además, la presencia de la población de moléculas de ADN largas en el plasma de los pacientes con y sin CHC justifica una mayor investigación. Cuando se comprenda mejor la fuente tisular o el proceso patológico que regula la liberación de estas moléculas de ADN, la medición de la proporción de ADN largo en el plasma podría ser útil para la evaluación de dichas enfermedades.

A. Distribución del tamaño del ADN plasmático de pacientes con CHC

Las distribuciones del tamaño del ADN plasmático de los pacientes con CHC, portadores del VHB, pacientes con cirrosis y controles sanos se muestran en las figura 18 y 19. En la figura 19, cada individuo está representado por un color diferente. En general, el pico más destacado se observó a 166 pb en el gráfico de distribución de tamaño de cada sujeto. Esta observación es coherente con informes previos sobre mujeres embarazadas y receptores de trasplantes (26-28), lo que sugiere que la mayoría de las moléculas de ADN circulantes proceden de la apoptosis. De manera interesante, en comparación con el perfil de distribución del tamaño medio de 32 controles sanos (línea negra gruesa en la figura 18), los tamaños del ADN plasmático en pacientes con CHC con concentraciones fraccionarias bajas de ADN tumoral fueron más largos. Sin embargo, con concentraciones fraccionarias crecientes de ADN tumoral en plasma, la distribución del tamaño del ADN plasmático se desplazó progresivamente hacia la izquierda (Figura 18).

Como se ha descrito anteriormente, la figura 13A es un gráfico de ΔS frente al tamaño de todos los casos de CHC con diferentes ANC en 8p y 8q en plasma. A medida que la concentración fraccionaria de ADN tumoral en el plasma aumenta de menos de un 2 % a más de un 8 %, la ΔS aumenta, lo que indica una mayor abundancia de fragmentos de ADN más cortos. La concentración fraccionaria de ADN tumoral en el plasma puede aumentar a medida que progresa la etapa del cáncer. Como resultado, la cantidad de fragmentos de ADN más cortos puede indicar una etapa posterior del cáncer. En la figura 13B se muestra que ΔS_{166} es más elevada para los pacientes con CHC, en comparación con sujetos sin CHC, lo que indica que la abundancia relativa de ADN <166 pb y ≥ 166 pb puede utilizarse para indicar la presencia de cáncer. Por consiguiente, la ΔS_{166} también puede indicar la etapa del cáncer.

En la figura 20 se muestra un ejemplo de cuándo se puede utilizar la proporción de fragmentos cortos para diferenciar pacientes con CHC de sujetos de control sanos. Se representó gráficamente la proporción de fragmentos de ADN plasmático de menos de 150 pb para 32 sujetos sanos, pacientes con CHC con una fracción de ADN tumoral inferior al 2 % en plasma y pacientes con CHC con una fracción de ADN tumoral superior al 6 % en plasma. En comparación con los sujetos de control sanos (marcados como 'CTR'), los pacientes con CHC con una fracción de ADN tumoral inferior al 2 % tuvieron una proporción significativamente menor de fragmentos de ADN cortos de menos de 150 pb ($p = 0,0002$, prueba de la t), y aquellos con una fracción de ADN tumoral mayor al 6 % tuvieron una proporción significativamente mayor de fragmentos cortos ($p = 0,003$, prueba de la t). Los pacientes con CHC con una fracción de ADN tumoral del 2 % al 6 % tienen una proporción de fragmentos de ADN entre los pacientes con CHC con una fracción tumoral inferior al 2 % y los pacientes con CHC con una fracción tumoral superior al 6 %. De esta manera, los pacientes con CHC con una fracción tumoral del 2 % al 6 % pueden tener una distribución similar a la de los sujetos de control sanos.

En la figura 21 se muestra una curva de eficacia diagnóstica (ROC) para aplicar $P(<150)$ para diferenciar pacientes con CHC con una fracción de ADN tumoral inferior al 2 % de sujetos de control sanos. La fracción tumoral se determinó en función de la magnitud de la irrepresiblemente de las regiones cromosómicas que presentaban infrarrepresentación en el plasma y que eran compatibles con una pérdida en el número de copias en el tumor. Para los casos sin infrarrepresentación significativa de ningún brazo cromosómico, la magnitud de la sobrerrepresentación de las regiones que eran compatibles con la ganancia en el número de copias se utilizó para determinar la fracción tumoral con el supuesto de una ganancia de copia única. La fracción tumoral se puede determinar con la siguiente ecuación:

$$\text{Fracción tumoral} = \frac{|P_{\text{prueba}} - P_{\text{normal}}|}{P_{\text{normal}} \times \Delta N / 2}$$

donde P_{prueba} representa la proporción de fragmentos asignados al brazo cromosómico de interés para el caso de prueba, P_{normal} representa la proporción media de fragmentos asignados al brazo cromosómico para los controles sanos y ΔN representa la magnitud del cambio en el número de copias (por ejemplo, 1 para una duplicación o una eliminación, y números más elevados para amplificaciones de orden superior). El área bajo la curva (ABC) fue de 0,776 con límites de confianza del 95 % de 0,670 y 0,882. Este resultado indica que el análisis de tamaño se puede utilizar para identificar pacientes con CHC con una fracción tumoral inferior al 2 % en plasma. El análisis de la curva ROC indica que se pueden seleccionar diferentes umbrales para lograr diferentes sensibilidades y especificidades.

En la figura 22, igual que en la figura 21, se muestra que el análisis de tamaño con $P(<150)$ también puede detectar pacientes con CHC con una fracción tumoral superior al 6 % en el plasma. El ABC para diferenciar a estos pacientes de sujetos sanos fue de 0,893 con límites de confianza del 95 % de 0,761 y 1,000.

En la figura 23 se muestra que la proporción de fragmentos largos de ADN plasmático se puede utilizar para detectar CHC, como se muestra en la figura 20 con la proporción de fragmentos cortos de ADN plasmático. En este ejemplo, la proporción de fragmentos mayores de 180 pb, indicados como $P(>180)$, se representó gráficamente para pacientes con CHC con menos del 2 % y más del 6 % de fracción de ADN tumoral en plasma y sujetos de control sanos. Esta proporción fue significativamente mayor en pacientes con CHC con menos del 2 % de fracción de ADN tumoral ($p < 0,00001$, prueba de la t).

En la figura 24 se muestra una curva ROC para utilizar $P(>180)$ para diferenciar pacientes con CHC con una fracción de ADN tumoral inferior al 2 % de sujetos de control sanos. El ABC fue de 0,883 con límites de confianza del 95 % de 0,805 y 0,961.

La figura 25 proporciona otro ejemplo de las diferentes distribuciones de tamaño de fragmentos de ADN con diferentes fracciones de ADN tumoral. En la figura 25 se muestran diagramas de caja del tamaño medio de los fragmentos de sujetos de control sanos, pacientes con CHC con una fracción de ADN tumoral inferior al 2 % y pacientes con CHC con una fracción de ADN tumoral superior al 6 %. El tamaño medio de los fragmentos de ADN de los pacientes con CHC con menos del 2 % de fracción de ADN tumoral fue significativamente mayor ($P < 0,00001$, prueba de la t) que el de los sujetos de control sanos. En cambio, el tamaño medio de los fragmentos de ADN de los pacientes con CHC con una fracción de ADN tumoral superior al 6 % fue significativamente más corto ($p = 0,03$, prueba de la t). La figura 25 apoya el uso del tamaño del fragmento de ADN como una forma de determinar el estadio del cáncer. Un tamaño medio más largo se asocia con una fracción de ADN tumoral más pequeña, mientras que un tamaño medio más corto se asocia con una fracción de ADN tumoral más grande. Si un individuo tiene una fracción de ADN tumoral más pequeña por debajo de un primer valor de corte y un tamaño medio por encima de un umbral de tamaño largo, entonces se puede confirmar que el cáncer está en etapa temprana. Por otra parte, si un individuo tiene una fracción de ADN tumoral más grande que un segundo valor de corte y un tamaño medio por debajo de un umbral de tamaño corto, entonces se puede confirmar que el cáncer está en etapa avanzada.

Los pacientes con CHC con una fracción de ADN tumoral del 2 % al 6 % tienen un tamaño medio de fragmentos de ADN entre los pacientes con CHC con una fracción tumoral inferior al 2 % y los pacientes con CHC con una fracción tumoral superior al 6 %. De esta manera, los pacientes con CHC con una fracción tumoral del 2 % al 6 % pueden tener una distribución similar a la de los sujetos de control sanos en la figura 25. Por lo tanto, si un individuo tiene una fracción de ADN tumoral desde el valor de corte inferior hasta el valor de corte superior y un tamaño medio desde un umbral de tamaño corto hasta un umbral de tamaño largo, entonces se puede confirmar que el cáncer está en etapa intermedia.

Las figuras 26 y 27 son curvas ROC que muestran que se pueden utilizar umbrales de tamaño diferentes para diferenciar a los pacientes con CHC de los sujetos de control sanos. La figura 26 es una curva ROC para utilizar el tamaño medio de los fragmentos para diferenciar entre pacientes con CHC con una fracción de ADN tumoral inferior al 2 % y sujetos de control sanos. El ABC fue de 0,812 con límites de confianza del 95 % de 0,718 y 0,907.

La figura 27 es una curva ROC para utilizar el tamaño medio de los fragmentos para diferenciar entre pacientes con CHC con una fracción de ADN tumoral superior al 2 % y sujetos de control sanos. El ABC fue de 0,795 con límites de confianza del 95 % de 0,627 y 0,963.

Otras características estadísticas de la distribución de tamaño (por ejemplo, mediana, media y percentil) se pueden utilizar como parámetro para diferenciar entre pacientes con CHC y sujetos sanos.

Además de analizar la distribución del tamaño de los fragmentos de ADN plasmático que surgen de todas las regiones genómicas, el análisis de tamaño también puede centrarse en fragmentos de ADN que surgen de regiones genómicas específicas. Una región genómica específica puede ser un brazo cromosómico.

En la figura 28 se muestra un diagrama de caja de la proporción de fragmentos cortos de ADN plasmático de menos de 150 pb que se alinearon con el cromosoma 1q para pacientes con CHC con una fracción de ADN tumoral superior al 6 % y para sujetos de control sanos. La proporción de fragmentos cortos fue significativamente mayor ($p < 0,00001$, prueba de la t) en los pacientes con CHC.

La figura 29 es una curva ROC para utilizar la proporción de fragmentos cortos de ADN plasmático de menos de 150 pb para diferenciar entre pacientes con CHC con una fracción de ADN tumoral superior al 6 % y sujetos de control sanos. El ABC fue de 0,915 con un intervalo de confianza del 95 % de 0,808 a 1,000.

B. Método

La figura 17 es un diagrama que ilustra un método 1700 para analizar una muestra biológica de un organismo de acuerdo con realizaciones de la presente divulgación, pero que no forman parte de la invención reivindicada. La muestra biológica puede incluir moléculas de ácido nucleico originarias de células normales y de células asociadas con el cáncer. Al menos algunas de las moléculas de ácido nucleico son extracelulares en la muestra biológica.

En la etapa 1710, para cada una de una pluralidad de las moléculas de ácido nucleico en la muestra biológica, se mide el tamaño de la molécula de ácido nucleico. El tamaño de la molécula de ácido nucleico se puede medir de manera similar a la etapa 910 de la figura 9.

En la etapa 1720, se identifica la ubicación de la molécula de ácido nucleico en un genoma de referencia del organismo. La identificación de la ubicación de la molécula de ácido nucleico se puede realizar de manera similar a la etapa 620 de la figura 6.

En la etapa 1730, se identifica un primer grupo de moléculas de ácido nucleico como perteneciente a una primera región cromosómica basándose en las ubicaciones identificadas. La primera región cromosómica puede incluir una

pluralidad de primeros loci. La identificación del respectivo grupo de moléculas de ácido nucleico puede ser similar a la etapa 120 de la figura 1.

En la etapa 1740, un sistema informático puede calcular un primer valor estadístico de una distribución de tamaño del primer grupo de moléculas de ácido nucleico. El cálculo de la cantidad respectiva puede ser similar al cálculo de la etapa 130 de la figura 1.

En la etapa 1750, se puede medir una fracción de moléculas de ácido nucleico originarias de células asociadas con el cáncer. La fracción se puede calcular de acuerdo con los métodos descritos en la publicación de patente de EE. UU. n.º 2013/0040824 titulada "Detection of Genetic or Molecular Aberrations Associated with Cancer" de Lo *et al.* presentada el 30 de noviembre de 2011. La fracción de moléculas de ácido nucleico tumoral corresponde a una proporción de las moléculas de ácido nucleico en la muestra que provienen del tumor o tumores. La fracción/proporción puede expresarse como cualquier porcentaje o valor decimal.

Los siguientes ejemplos son métodos para la medición de la fracción de ácidos nucleicos tumorales, pero se pueden utilizar otros métodos. La fracción de ácidos nucleicos tumorales se puede determinar en función de la magnitud de la infrarrepresentación (o sobrerepresentación) en el plasma para las regiones que presentan una infrarrepresentación significativa que es compatible con la pérdida en el número de copias (o ganancia en el número de copias) en los tejidos tumorales. Otro ejemplo es determinar el grado de desequilibrio alélico en dos cromosomas homólogos para regiones afectadas por anomalías en el número de copias, por ejemplo, regiones con pérdida de una copia de los dos cromosomas homólogos. Otro ejemplo es determinar la concentración fraccionaria de una mutación asociada al cáncer, incluida la mutación de un solo nucleótido, eliminación de nucleótido(s) y translocación. La fracción tumoral se puede determinar mediante los métodos descritos en la figura 21 anterior.

En la etapa 1760, se puede seleccionar un primer valor de referencia basado en la fracción medida. En un ejemplo, la selección del primer valor de referencia puede incluir la selección de un umbral de tamaño cuando la fracción medida está por debajo de un valor de corte. En otro ejemplo, la selección del primer valor de referencia puede incluir la selección de un umbral de tamaño cuando la fracción medida está por encima de un valor de corte. En estos ejemplos, los valores de corte y los umbrales de tamaño pueden diferir y pueden depender del valor de la fracción medida.

En la etapa 1770, el primer valor estadístico puede compararse con un primer valor de referencia para determinar un estadio de cáncer de la muestra biológica. El primer valor estadístico puede ser cualquier valor estadístico descrito en el presente documento.

La existencia de cáncer se puede confirmar mediante el análisis de tamaño junto con la fracción medida de moléculas de ácido nucleico que se originan de células asociadas con el cáncer. Por ejemplo, cuando la fracción medida está por debajo de un valor de corte inferior, se puede confirmar si la distribución de tamaño es más larga que la de los controles sanos (por ejemplo, si el primer valor estadístico está por encima del umbral de tamaño). Si la distribución de tamaño es más larga que para los controles sanos, esto puede confirmar una etapa temprana del cáncer. Algunos ejemplos de valor de corte bajo son 0,01, 0,015, 0,02 o 0,025. Como otro ejemplo, cuando la fracción medida está por encima de un valor de corte elevado, se puede confirmar si la distribución de tamaño es más corta que la de los controles sanos (por ejemplo, si el primer valor estadístico está por debajo del umbral de tamaño). Si la distribución de tamaño es más corta para los controles sanos, esto puede confirmar una etapa avanzada del cáncer. Los ejemplos del valor de corte elevado pueden ser una fracción de 0,03, 0,035, 0,04, 0,045, 0,05, 0,055, 0,06, 0,065 o 0,07.

Se demostró que había poblaciones adicionales de moléculas de ADN más cortas y más largas en el plasma de pacientes con CHC. Estos datos podrían haber resuelto las aparentes incoherencias que existían en la bibliografía donde los grupos informaban de la presencia de un aumento en las moléculas de ADN más largas o más cortas en el plasma de pacientes con cáncer.

VII. MATERIALES Y MÉTODOS

A continuación, se indican las técnicas utilizadas en la obtención de los resultados de las figuras 2 a 5. Estas técnicas se pueden utilizar en otros ejemplos anteriores.

Los sujetos reclutados para el estudio incluyeron 90 pacientes con CHC ingresados en el Departamento de Cirugía del Hospital Prince of Gales, Hong Kong, para la resección del tumor. Todas las muestras de sangre se recogieron antes de la operación. Se reclutaron sesenta y siete portadores del VHB y 36 pacientes con cirrosis relacionada con el VHB del Departamento de Medicina y Terapéutica del Hospital Prince of Gales, Hong Kong. Todos los pacientes dieron su consentimiento informado por escrito y el estudio fue aprobado por el comité de revisión institucional.

Para extraer ADN y preparar bibliotecas de secuencias, se recogieron muestras de sangre periférica en tubos que contenían EDTA. Las muestras de sangre periférica se centrifugaron a 1.600g durante 10 min a 4 °C. La parte de plasma se volvió a centrifugar a 16.000g durante 10 min a 4 °C para obtener plasma libre de células. Se extrajo ADN de 3 a 4,8 ml de plasma utilizando el kit QIAamp DSP DNA Blood Mini (Qiagen). El ADN plasmático se concentró con un concentrador SpeedVac (Savant DNA120; Thermo Scientific) en un volumen final de 75 µl por muestra. Las

bibliotecas de ADN catalogadas se prepararon utilizando el kit de preparación de bibliotecas Kapa (Kapa Biosystems) siguiendo las instrucciones del fabricante. El ADN ligado al adaptador se enriqueció mediante una PCR de 14 ciclos utilizando el kit de PCR KAPA HiFi HotStart ReadyMix (Kapa Biosystems). A continuación, las bibliotecas se analizaron con un bioanalizador 2100 (Agilent) y se cuantificaron con el kit de cuantificación de bibliotecas Kapa (Kapa Biosystems) antes de la secuenciación.

Para secuenciar y alinear el ADN, cada biblioteca de ADN se diluyó y se hibridó en una cubeta de flujo de secuenciación de extremos emparejados (Illumina). Los grupos de ADN se generaron en un sistema de generación de grupos cBot (Illumina) con el kit de generación de grupos TruSeq PE v3 (Illumina), seguido de 76×2 ciclos de secuenciación en un sistema HiSeq 2000 (Illumina) con el TruSeq SBS Kit v3 (Illumina). La secuenciación se realizó utilizando un protocolo 4-plex. Se realizaron 7 ciclos adicionales de secuenciación para decodificar la secuencia índice en cada molécula de ADN secuenciada. El análisis de imágenes en tiempo real y la lectura de bases se realizaron utilizando el programa informático HiSeq Control (HCS) v1.4 y el programa informático Real Time Analysis (RTA) v1.13 (Illumina), mediante los que los cálculos automatizados de la matriz y el desfase se basaron en el control PhiX v3 secuenciado con las bibliotecas. Después de la lectura de bases, se eliminaron las secuencias adaptadoras y las bases de baja calidad (es decir, puntuación de calidad <5).

Para el análisis de datos de secuenciación, las secuencias de cada carril se asignaron a las muestras correspondientes según las secuencias de índice de seis bases. A continuación, las lecturas secuenciadas se alinearon con el genoma de referencia humano sin ocultación de repetición (NCBI build 37/hg19) utilizando el Programa de Alineación de Oligonucleótidos Cortos 2 (SOAP2) (40). Se permitieron hasta dos desapareamientos de nucleótidos para cada miembro de las lecturas de extremos emparejados, pero no se permitieron inserciones ni eliminaciones. Las lecturas asignadas a una ubicación genómica única se utilizaron para análisis posteriores. Las lecturas de extremos emparejados alineadas al mismo cromosoma con una orientación correcta y que abarcan un tamaño de inserto de ≤ 600 pb se conservaron para los análisis de tamaño posteriores. Después de la alineación con el genoma humano de referencia, el tamaño de cada fragmento de ADN plasmático podría deducirse de las coordenadas de los nucleótidos en los extremos más externos de cada par de lecturas de secuencia. Las primeras lecturas de un solo extremo se utilizaron para el análisis de ANC. Se aceptaron lecturas con una calidad de cartografiado mayor a 30 (es decir, 1 alineación errónea por cada 1000 alineaciones) utilizando el programa informático Bowtie 2 (41).

Para realizar el análisis CAZA para ANC, todo el genoma humano se dividió en intervalos de 100 kb. El recuento de lectura corregido por GC se determinó para cada intervalo de 100 kb como se informó anteriormente (42). El número de lecturas corregidas por GC para cada brazo cromosómico de interés se determinó mediante la suma de todos los valores de cada intervalo de 100 kb en el brazo cromosómico. Se utilizó una estadística de puntuación z para determinar si la representación del ADN plasmático en un brazo cromosómico aumentaría o disminuiría significativamente en comparación con el grupo de referencia. Se calculó el porcentaje de lecturas de secuenciación asignadas a cada brazo cromosómico y se comparó con el valor medio de los 32 sujetos de control sanos para el brazo cromosómico respectivo. Se calculó una puntuación z a nivel del brazo como

$$Puntuación\ z = \frac{P_{prueba} - P_{normal}}{DT_{normal}}$$

donde P_{prueba} representa la proporción de fragmentos asignados al brazo cromosómico de interés para el caso de prueba; P_{normal} y DT_{normal} representan la media y la desviación típica de la proporción de fragmentos asignados al brazo cromosómico para los controles sanos, respectivamente. Los brazos cromosómicos con puntuaciones z de <-3 y >3 se consideraron como portadores de ANC en plasma correspondientes a eliminaciones y amplificaciones, respectivamente.

La concentración fraccionaria de ADN procedente del tumor en el plasma (F) se puede calcular como

$$F = \frac{|P_{prueba} - P_{normal}|}{\Delta N / 2 \times P_{normal}}$$

donde P_{prueba} representa la proporción de fragmentos asignados al brazo cromosómico de interés para el caso de prueba; P_{normal} representa la proporción media de fragmentos asignados al brazo cromosómico para los controles sanos y ΔN representa el cambio del número de copias. Para los casos que muestran una eliminación en al menos un brazo cromosómico, se calculó F en función de los brazos cromosómicos eliminados. Como la mayoría de las eliminaciones del brazo cromosómico afectan solo a uno de los dos cromosomas homólogos (33), se asumió una pérdida de copia única para nuestro análisis. Para los 24 casos con solo amplificación del brazo cromosómico, pero sin eliminación, F se calculó en función del brazo amplificado con el supuesto de ganancia de copia única.

El análisis de datos de secuenciación se realizó mediante programas bioinformáticos escritos en lenguajes Perl y R. Un valor de $p < 0,05$ se consideró estadísticamente significativo y todas las probabilidades fueron bilaterales.

VIII. SISTEMA INFORMÁTICO

Cualquiera de los sistemas informáticos mencionados en el presente documento puede utilizar cualquier número adecuado de subsistemas. En la figura 32, en el aparato informático 10 se muestran ejemplos de dichos subsistemas. En algunas realizaciones, un sistema informático incluye un único aparato informático, donde los subsistemas pueden ser los componentes del aparato informático. En otras realizaciones, un sistema informático puede incluir varios aparatos informáticos, siendo cada uno de ellos un subsistema, con componentes internos. Un sistema informático puede incluir ordenadores de escritorio y portátiles, tabletas, teléfonos móviles y otros dispositivos móviles.

Los subsistemas mostrados en la figura 32 están interconectados a través de un bus de sistema 75. Se muestran subsistemas adicionales, tales como una impresora 74, un teclado 78, un dispositivo(s) de almacenamiento 79, un monitor 76, que está acoplado al adaptador de pantalla 82, y otros. Los dispositivos periféricos y de entrada/salida (I/O, del inglés *input/output*), que se acoplan al controlador de I/O 71, pueden conectarse al sistema informático mediante cualquier tipo de medio conocido en la materia, tal como el puerto de entrada/salida (I/O) 77 (por ejemplo, USB, FireWire®). Por ejemplo, el puerto de I/O 77 o la interfaz externa 81 (por ejemplo, Ethernet, Wi-Fi, etc.) se pueden utilizar para conectar el sistema informático 10 a una red de área amplia, tal como Internet, un dispositivo de entrada de ratón o un escáner. La interconexión mediante el bus de sistema 75 permite que el procesador central 73 se comunique con cada subsistema y controle la ejecución de instrucciones procedentes de la memoria del sistema 72 o de los dispositivos de almacenamiento 79 (por ejemplo, un disco fijo, tal como un disco duro o un disco óptico), así como el intercambio de información entre subsistemas. La memoria del sistema 72 y/o los dispositivos de almacenamiento 79 pueden ser realizaciones de un medio legible por ordenador. Otro subsistema es un dispositivo de recogida de datos 85, tal como una cámara, micrófono, acelerómetro y similares. Cualquiera de los datos mencionados en el presente documento puede ser la salida de un componente a otro componente y puede ser la salida al usuario.

Un sistema informático puede incluir una pluralidad de los mismos componentes o subsistemas, por ejemplo, conectados entre sí por la interfaz externa 81 o por una interfaz interna. En algunas realizaciones, los sistemas informáticos, subsistema o aparatos pueden comunicarse a través de una red. En dichos casos, un ordenador puede considerarse como cliente y otro ordenador como servidor, donde cada uno puede formar parte de un mismo sistema informático. Un cliente y un servidor pueden incluir cada uno múltiples sistemas, subsistemas o componentes.

El método de la presente invención está implementado por ordenador. Debe entenderse que cualquiera de las realizaciones de la presente invención puede implementarse en forma de lógica de control utilizando un equipo físico (por ejemplo, un circuito integrado específico de la aplicación o una matriz de selección programable por campo) y/o utilizando un programa de ordenador con un procesador generalmente programable de manera modular o integrada. Como se utiliza en el presente documento, un procesador incluye un procesador de un solo núcleo, un procesador de varios núcleos en un mismo chip integrado o múltiples unidades de procesamiento en una sola placa de circuito o en red. Basándose en la divulgación y las enseñanzas proporcionadas en el presente documento, un experto habitual en la materia conocerá y percibirá otras formas y/o métodos para implementar las realizaciones de la presente invención utilizando equipos físicos y una combinación de equipos físicos y programas informáticos.

Puede implementarse cualquiera de los componentes o las funciones de programas informáticos descritos en la presente solicitud como código de programación para su ejecución mediante un procesador utilizando cualquier lenguaje informático adecuado, tal como, por ejemplo, Java, C, C++, C#, Objective-C, Swift o lenguaje de guiones, tal como Perl o Python, utilizando, por ejemplo, técnicas convencionales u orientadas por objetos. El código del programa informático puede almacenarse como una serie de instrucciones o comandos en un medio legible por ordenador para su almacenamiento y/o transmisión, incluyendo los medios adecuados la memoria de acceso aleatorio (RAM), una memoria de solo lectura (ROM), un medio magnético, tal como un disco duro o disquete, o un medio óptico, tal como un disco compacto (CD) o un DVD (disco digital versátil), memoria *flash* y similares. El medio legible por ordenador puede ser cualquier combinación de tales dispositivos de almacenamiento o transmisión.

Dichos programas también pueden codificarse y transmitirse utilizando señales portadoras adaptadas para la transmisión a través de redes cableadas, ópticas y/o inalámbricas que se ajusten a una variedad de protocolos, incluida Internet. Así pues, se puede crear un medio legible por ordenador utilizando una señal de datos codificada con dichos programas. Los medios legibles por ordenador codificados con el código del programa pueden empaquetarse con un dispositivo compatible o proporcionarse por separado de otros dispositivos (por ejemplo, a través de descarga de Internet). Cualquier medio legible por ordenador puede residir en o dentro de un solo producto informático (por ejemplo, un disco duro, un CD o un sistema informático completo) y puede estar presente en o dentro de diferentes productos informáticos dentro de un sistema o red. Un sistema informático puede incluir un monitor, una impresora u otro dispositivo de presentación adecuado para proporcionar a un usuario cualquiera de los resultados mencionados en el presente documento.

Cualquiera de los métodos descritos en el presente documento puede realizarse total o parcialmente con un sistema informático que incluya uno o más procesadores, que pueden configurarse para realizar las etapas. Por tanto, las realizaciones pueden dirigirse a sistemas informáticos configurados para realizar las etapas de cualquiera de los métodos descritos en el presente documento, potencialmente con diferentes componentes que realizan una etapa o un grupo de etapas respectivas. Aunque se presentan como etapas numeradas, las etapas de los métodos descritos

en el presente documento pueden realizarse al mismo tiempo o en un orden diferente.

La descripción anterior de las realizaciones ilustrativas de la invención se ha presentado con fines ilustrativos y descriptivos. No pretende ser exhaustivo ni limitar la invención a la forma precisa descrita, y son posibles muchas modificaciones y variaciones dentro del alcance de las reivindicaciones adjuntas que definen la invención.

Una mención de "un", "uno/una" o "el/la" se entiende como "uno o más" a menos que se indique específicamente lo contrario. Se entiende que el uso de "o" hace referencia a un "o inclusivo", y no a un "o exclusivo", a menos que se indique específicamente lo contrario.

IX. Bibliografía

1. Chan K. C. A. (2013) Scanning for cancer genomic changes in plasma: toward an era of personalized blood-based tumor markers. *Clin Chem* 59(11):1553-1555.
2. Dawson S. J., Rosenfeld N., y Caldas C. (2013) Circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 369(1):93-94.
3. Bidard F. C., Weigelt B., y Reis-Filho J. S. (2013) Going with the flow: from circulating tumor cells to DNA. *Sci Transl Med* 5(207):207ps214.
4. Chan K. C. A., *et al.* (2013) Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* 59(1):211-224.
5. Heitzer E., *et al.* (2013) Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int J Cancer* 133(2):346-356.
6. Heitzer E., *et al.* (2013) Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* 5(4):30.
7. Leary R. J., *et al.* (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* 4(162):162ra154.
8. Chan K. C. A., *et al.* (2013) Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci USA* 110(47):18761-18768.
9. Chan K. C. A., *et al.* (2008) Quantitative analysis of circulating methylated DNA as a biomarker for hepatocellular carcinoma. *Clin Chem* 54(9):1528-1536.
10. Wong I. H., *et al.* (1999) Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Res* 59(1):71-73.
11. Balgouranidou I., *et al.* (2014) Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer. *Br J Cancer* 110(8):2054-2062.
12. Diehl F., *et al.* (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci USA* 102(45):16368-16373.
13. Yung T. K. F., *et al.* (2009) Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res* 15(6):2076-2084.
14. Murtaza M., *et al.* (2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 497(7447):108-112.
15. Forshew T., *et al.* (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 4(136):136ra168.
16. Lo Y. M. D., *et al.* (1999) Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. *Cancer Res* 59(6):1188-1191.
17. Chan K. C. A., *et al.* (2013) Early detection of nasopharyngeal carcinoma by plasma Epstein-Barr virus DNA analysis in a surveillance program. *Cancer* 119(10):1838-1844.
18. McBride D. J., *et al.* (2010) Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes, Chromosomes & Cancer* 49(11):1062-1069.

19. Leary R. J., *et al.* (2010) Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* 2(20):20ra14.
- 5 20. Chan K. C. A., Leung S. F., Yeung S. W., Chan A. T. C., y Lo Y. M. D. (2008) Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin Cancer Res* 14(13):4141-4145.
- 10 21. Gao Y. J., *et al.* (2010) Increased integrity of circulating cell-free DNA in plasma of patients with acute leukemia. *Clin Chem Lab Med* 48(11):1651-1656.
22. Umetani N., *et al.* (2006) Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clin Chem* 52(6):1062-1069.
- 15 23. Wang B. G., *et al.* (2003) Increased plasma DNA integrity in cancer patients. *Cancer Res* 63(14):3966-3968.
24. Umetani N., *et al.* (2006) Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 24(26):4270-4276.
- 20 25. Schwarzenbach H., *et al.* (2012) Loss of heterozygosity at tumor suppressor genes detectable on fractionated circulating cell-free tumor DNA as indicator of breast cancer progression. *Clin Cancer Res* 18(20):5719-5730.
26. Lo Y. M. D., *et al.* (2010) Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2(61):61ra91.
- 25 27. Zheng Y. W. L., *et al.* (2012) Nonhematopoietically derived DNA is shorter than hematopoietically derived DNA in plasma: a transplantation model. *Clin Chem* 58(3):549-558.
28. Yu S. C. Y., *et al.* (2014) Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A* 111(23):8583-8588.
- 30 29. Pleasance E. D., *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191-196.
- 35 30. Fujimoto A., *et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 44(7):760-764.
31. Tao Y., *et al.* (2011) Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data. *Proc Natl AcadSci USA* 108(29):12042-12047.
- 40 32. Totoki Y., *et al.* (2011) High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 43(5):464-469.
33. Beroukhim R., *et al.* (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463(7283):899-905.
- 45 34. Chiang D. Y., *et al.* (2008) Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res* 68(16):6779-6788.
35. Kan Z., *et al.* (2013) Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* 23(9):1422-1433.
- 50 36. Kim T. M., *et al.* (2008) Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q. *Int J Cancer* 123(12):2808-2815.
- 55 37. Nakano H. y Shinohara K. (1994) X-ray-induced cell death: apoptosis and necrosis. *Radiation Research* 140(1):1-9.
38. Walker N. I., Harmon B. V., Gobe G. C., y Kerr J. F. (1988) Patterns of cell death. *Methods and Achievements in Experimental Pathology* 13:18-54.
- 60 39. Alexandrov L. B., *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
- 65 40. Li R., *et al.* (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25(15):1966-1967.

41. Langmead B. y Salzberg S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357-359.
42. Chen E. Z., *et al.* (2011) Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One* 6(7):e21791.

REIVINDICACIONES

1. Un método implementado por ordenador para analizar una muestra biológica de un organismo, incluyendo la muestra biológica moléculas de ácido nucleico procedentes de células normales y de células procedentes de tejido tumoral, en donde al menos algunas de las moléculas de ácido nucleico son extracelulares en la muestra biológica, comprendiendo el método:
 - identificar una pluralidad de regiones cromosómicas del organismo, incluyendo cada región cromosómica una pluralidad de loci y dando lugar a una división de un cromosoma en una pluralidad de regiones, en donde la pluralidad de regiones cromosómicas no se superpone, y en donde una primera región cromosómica es una de la pluralidad de regiones cromosómicas;
 - para cada una de la pluralidad de moléculas de ácido nucleico en la muestra biológica:
 - medir el tamaño de la molécula de ácido nucleico;
 - identificar la ubicación de la molécula de ácido nucleico en un genoma de referencia del organismo;
 - para cada una de la pluralidad de regiones cromosómicas:
 - identificar un grupo respectivo de una pluralidad de moléculas de ácido nucleico en la muestra biológica como de la región cromosómica basándose en las ubicaciones identificadas, incluyendo el grupo respectivo al menos una molécula de ácido nucleico ubicada en cada uno de la pluralidad de loci de la región cromosómica;
 - calcular, con un sistema informático, una cantidad respectiva del respectivo grupo de moléculas de ácido nucleico; y
 - comparar la cantidad respectiva con un valor de referencia de recuento para determinar una clasificación de recuento de si la región cromosómica presenta una anomalía, siendo la anomalía una anomalía en el número de copias, indicando la clasificación de recuento una primera anomalía para la primera región cromosómica de la pluralidad de regiones cromosómicas;
 - identificar un primer grupo de moléculas de ácido nucleico de la muestra como de la primera región cromosómica basándose en las ubicaciones identificadas;
 - calcular, con un sistema informático, un primer valor estadístico de una primera distribución de tamaño del primer grupo de moléculas de ácido nucleico;
 - comparar el primer valor estadístico con un valor de referencia de tamaño para determinar una clasificación de tamaño de si la primera región cromosómica presenta la primera anomalía; y
 - determinar una clasificación final de si la primera región cromosómica presenta la primera anomalía utilizando la clasificación de recuento y la clasificación de tamaño de la primera región cromosómica.
 2. El método de una cualquiera de las reivindicaciones anteriores, en donde el primer valor estadístico comprende una media de la primera distribución de tamaño, o una mediana de la primera distribución de tamaño.
 3. El método de una cualquiera de las reivindicaciones anteriores, en donde el valor de referencia de tamaño corresponde a un segundo valor estadístico de una segunda distribución de tamaño de una segunda región cromosómica.
 4. El método de una cualquiera de las reivindicaciones 1 a 2, en donde el valor de referencia de tamaño es de uno o más sujetos de referencia que se sabe que tienen cáncer o que se sabe que están sanos.
 5. El método de una cualquiera de las reivindicaciones 1 a 2, en donde el valor de referencia de tamaño se determina mediante el cálculo de un área en un tamaño especificado bajo un gráfico de frecuencia acumulada de las moléculas de ácido nucleico para una segunda región cromosómica en un intervalo de tamaños.
 6. El método de una cualquiera de las reivindicaciones anteriores, que comprende además:
 - determinar un conjunto de clasificaciones de tamaño para un conjunto de regiones cromosómicas identificadas como anómalas en función de las clasificaciones de recuento correspondientes;
 - confirmar si el conjunto de regiones cromosómicas es anómalo en función del conjunto de clasificaciones de tamaño; y
 - confirmar una clasificación de un nivel de cáncer basado en las regiones cromosómicas determinadas como anómalas.
 7. El método de una cualquiera de las reivindicaciones anteriores, que comprende además:
 - comparar la clasificación de tamaño con un primer umbral, y
 - comparar la clasificación de recuento con un segundo umbral, en donde:
 - a medida que aumenta la probabilidad de una anomalía, como lo indica una de la clasificación de tamaño o la

clasificación de recuento, entonces se reduce el umbral de la probabilidad indicado por la otra clasificación, y la clasificación final se determina mediante la comparación de la clasificación de tamaño con el primer umbral y mediante la comparación de la clasificación de recuento con el segundo umbral.

- 5 8. El método de una cualquiera de las reivindicaciones anteriores, que comprende además determinar que la primera anomalía es una amplificación asociada con un cáncer cuando la primera distribución de tamaño del primer grupo de moléculas de ácido nucleico indica un enriquecimiento para moléculas de ácido nucleico más cortas que para la primera anomalía que es una eliminación.
- 10 9. El método de una cualquiera de las reivindicaciones anteriores, en donde la medición del tamaño de la molécula de ácido nucleico incluye alinear los extremos más externos de las lecturas de secuencia de extremos emparejados de la molécula de ácido nucleico con un genoma humano de referencia.
- 15 10. El método de una cualquiera de las reivindicaciones anteriores, en donde la clasificación de tamaño indica la primera anomalía, confirmando así la clasificación de recuento.
- 20 11. El método de una cualquiera de las reivindicaciones anteriores, en donde la clasificación de recuento para la primera región cromosómica que indica la primera anomalía asociada con un cáncer se utiliza para seleccionar la primera región cromosómica para el análisis de tamaño.
12. Un programa informático que comprende una pluralidad de instrucciones que pueden ejecutarse en un sistema informático, que cuando se ejecutan de este modo controlan el sistema informático para realizar el método de una cualquiera de las reivindicaciones anteriores.

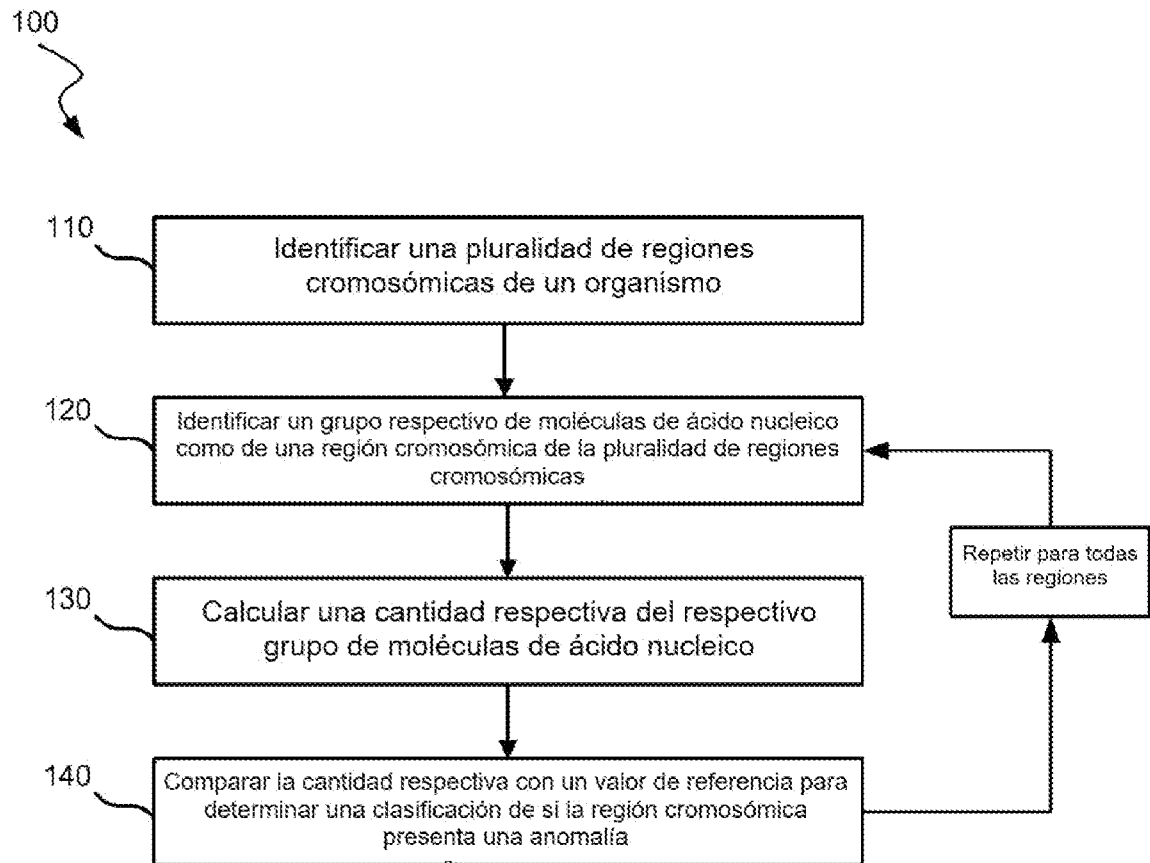
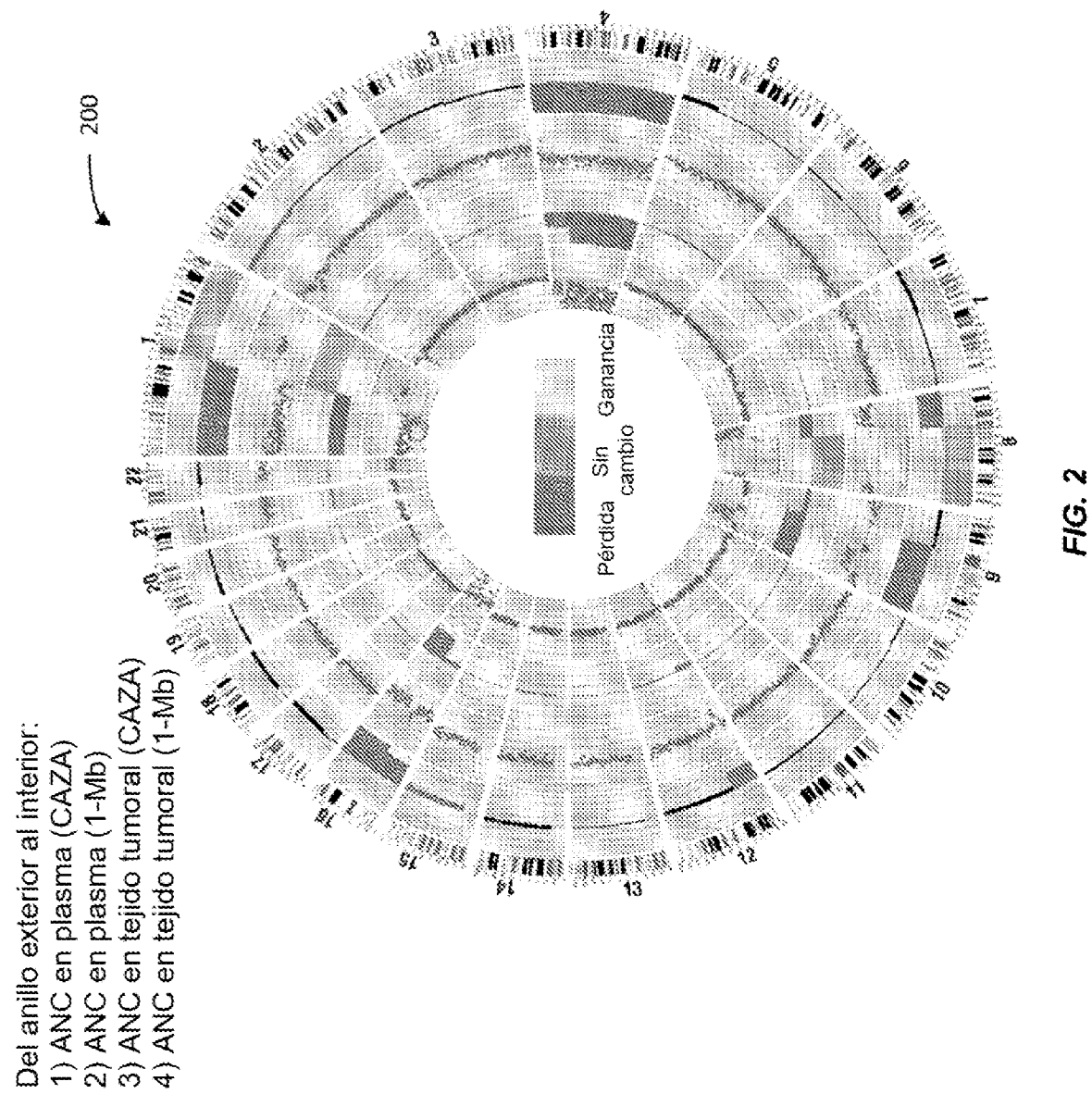


FIG. 1



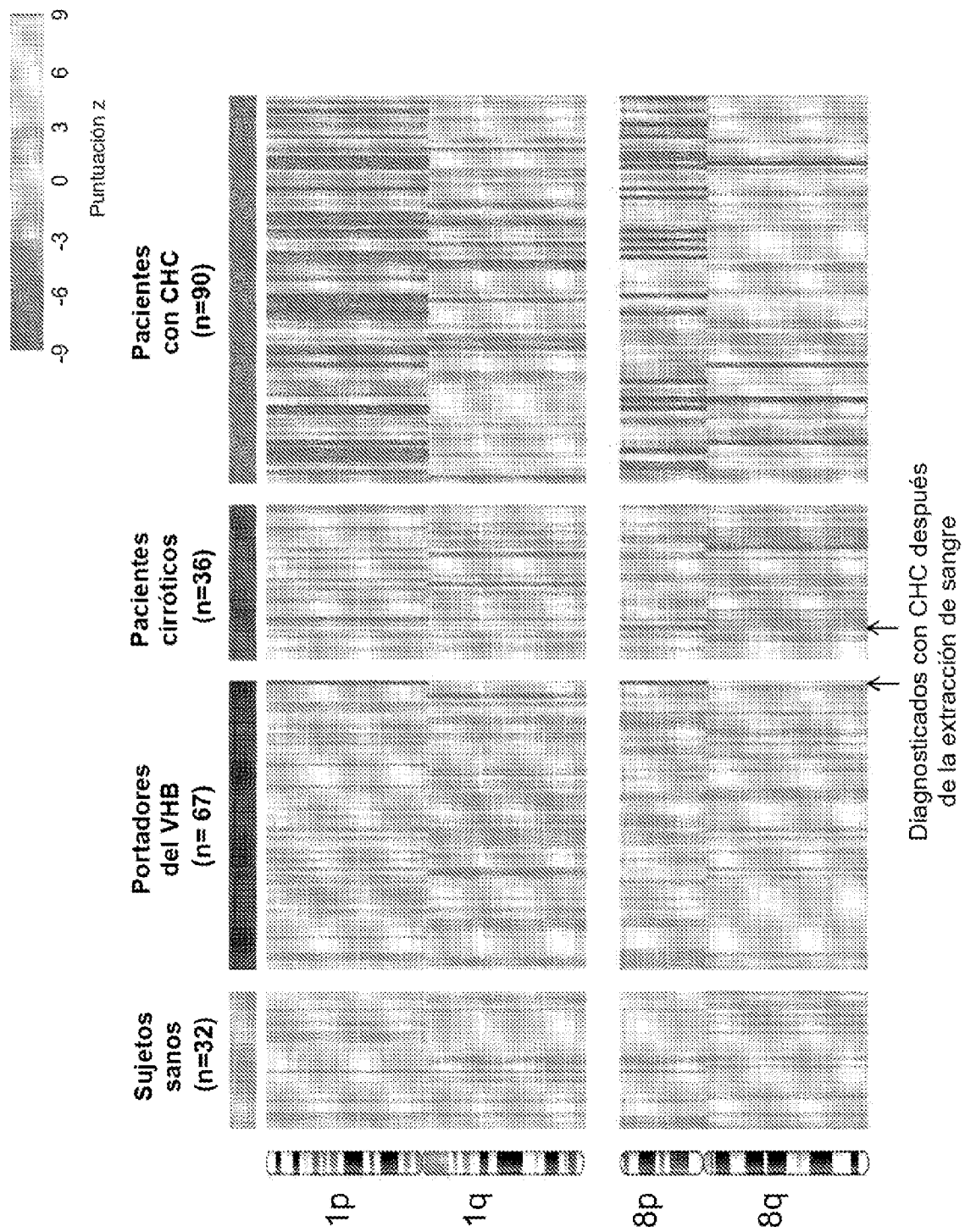


FIG. 3

400

| | N.º de pacientes (porcentaje) con ANC detectadas en plasma para | | | | | |
|-----------------------------------|---|------------------|------------------|------------------|---------------------------------|--|
| | Cr 1p | Cr 1q | Cr 8p | Cr 8q | Cualquiera de
cr 1p/1q/8p/8q | |
| Pacientes con
CHC (n=90) | 39/90
(43,3%) | 46/90
(51,1%) | 33/90
(36,7%) | 63/90
(70,0%) | 76/90
(84,4%) | |
| Pacientes
cirróticos
(n=36) | 3/36
(8,3%) | 5/36
(13,9%) | 1/36
(2,8%) | 1/36
(2,8%) | 8/36
(22,2%) | |
| Portadores
del VHB
(n=67) | 1/67
(1,5%) | 3/67
(4,48%) | 1/67
(1,5%) | 1/67
(1,5%) | 3/67
(4,5%) | |
| Sujetos
sanos
(n=32) | 0/32
(0%) | 0/32
(0%) | 0/32
(0%) | 0/32
(0%) | 0/32
(0%) | |

FIG. 4

| N.º de caso | Fracción de ADN tumoral en plasma (%) | Tamaño del tumor (cm) | 1p | | 1q | | 8p | | 8q | |
|-------------|---------------------------------------|-----------------------|----------|----------|----------|----------|---------|----------|----------|----------|
| | | | Tumor | Plasma | Tumor | Plasma | Tumor | Plasma | Tumor | Plasma |
| HOT412 | 27,5 | 6,5 | Ganancia | Ganancia | Ganancia | Ganancia | Pérdida | Pérdida | Ganancia | Ganancia |
| HOT407 | 19,4 | 8 | Nulo | Nulo | Nulo | Ganancia | Pérdida | Pérdida | Ganancia | Ganancia |
| HOT414 | 18,7 | 15 | Pérdida | Pérdida | Ganancia | Ganancia | Pérdida | Pérdida | Ganancia | Ganancia |
| HOT426 | 4,1 | 6,5 | Pérdida | Pérdida | Ganancia | Ganancia | Pérdida | Pérdida | Nulo | Ganancia |
| HOT394 | 2,5 | 4 | Pérdida | Pérdida | Ganancia | Ganancia | Pérdida | Pérdida | Ganancia | Ganancia |
| HOT397 | 2,4 | 5,5 | Pérdida | Pérdida | Ganancia | Nulo | Pérdida | Nulo | Ganancia | Ganancia |
| HOT393 | 1,8 | 3 | Pérdida | Pérdida | Nulo | Loss | Nulo | Ganancia | Nulo | Ganancia |
| HOT427 | 1,7 | 1,5 | Pérdida | Nulo | Ganancia | Ganancia | Nulo | Nulo | Nulo | Nulo |
| HOT413 | 1,7 | 2 | Nulo | Pérdida | Ganancia | Nulo | Pérdida | Nulo | Ganancia | Ganancia |
| HOT425 | 1,5 | 8,5 | Pérdida | Pérdida | Nulo | Nulo | Pérdida | Nulo | Ganancia | Ganancia |
| HOT428 | 1,5 | 2 | Pérdida | Nulo | Ganancia | Pérdida | Pérdida | Nulo | Ganancia | Ganancia |
| HOT432 | 1,0 | 2 | Pérdida | Nulo | Nulo | Nulo | Nulo | Nulo | Nulo | Ganancia |

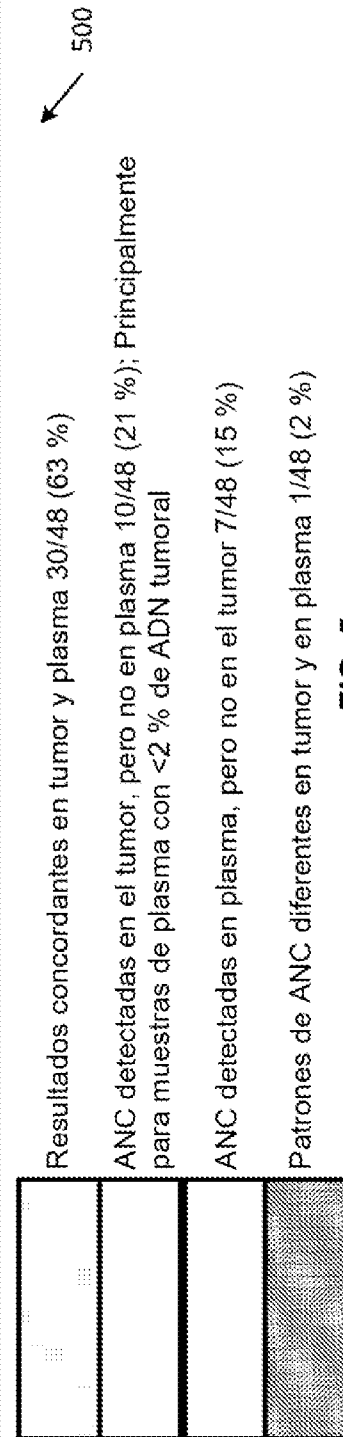


FIG. 5

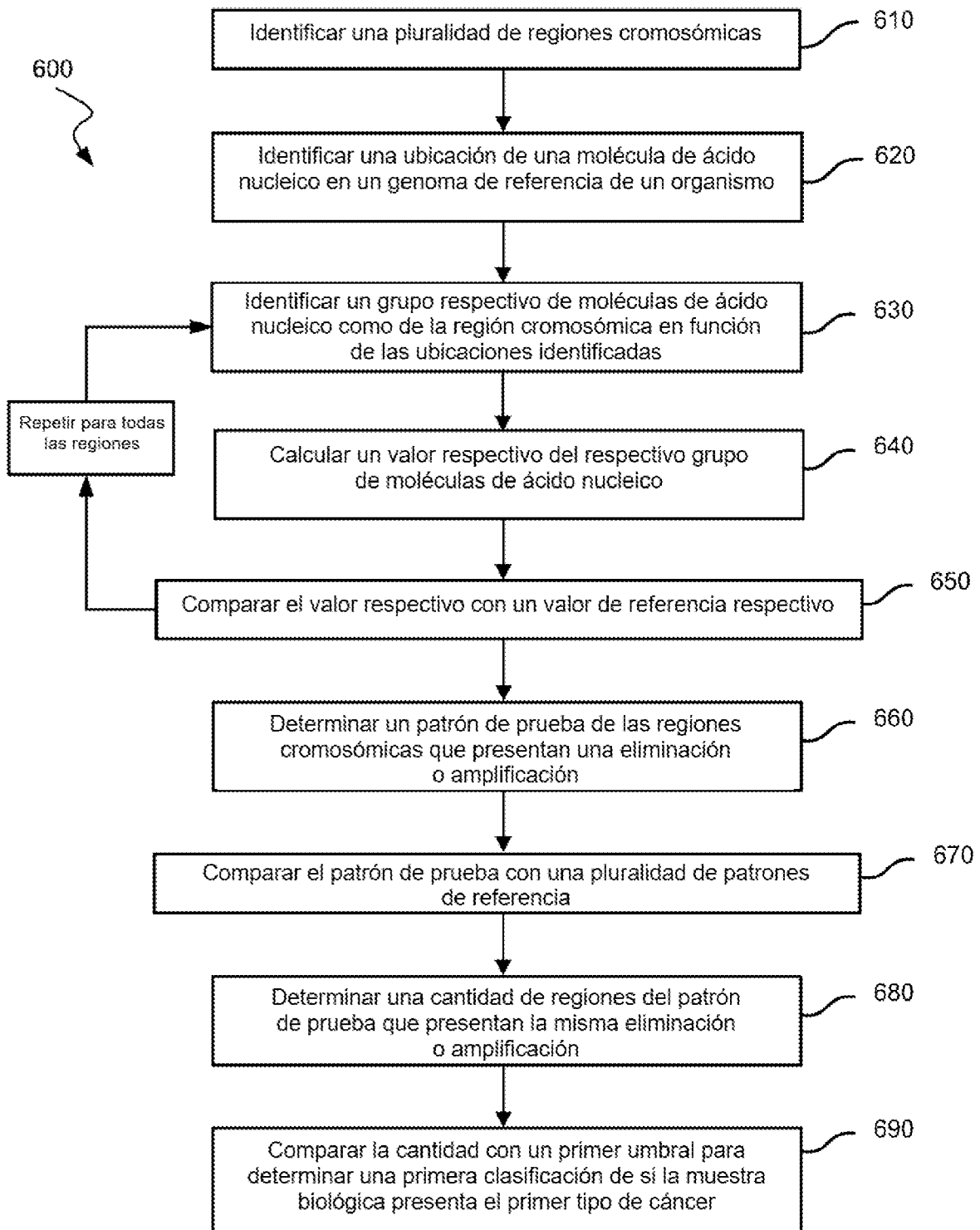


FIG. 6

| | CHC1 | CHC2 | CHC3 | CHC4 | CHC5 | CHC6 | CCR1 | CCR2 | CCR3 | CCR4 | CM1 | CM2 | CM3 | CP1 | CP2 | CNF1 | CNF2 |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1p | | Nufo | + | | | | | | | | | | | Nufo | + | Nufo | - |
| 1q | | | + | Nufo | + | | | | | | | | | | | Nufo | + |
| 2p | + | + | - | Nufo | + | Nufo | | | | | Nufo | | Nufo | + | - | Nufo | + |
| 2q | Nufo | + | - | Nufo | Nufo | Nufo | | | | | | | | | | | |
| 3p | Nufo | + | + | Nufo | Nufo | Nufo | Nufo | Nufo | + | - | | | | | | Nufo | - |
| 3q | Nufo | + | + | Nufo | Nufo | Nufo | + | + | + | - | | | | + | | Nufo | - |
| 4p | + | + | + | + | + | Nufo | | | | | | | | | | Nufo | - |
| 4q | Nufo | - | - | Nufo | Nufo | Nufo | | | | | | | Nufo | Nufo | - | Nufo | - |
| 5p | Nufo | Nufo | + | + | Nufo | Nufo | + | + | + | + | | | Nufo | + | Nufo | Nufo | Nufo |
| 5q | Nufo | Nufo | Nufo | Nufo | Nufo | Nufo | Nufo | + | - | - | | | | | Nufo | Nufo | Nufo |
| 6p | Nufo | + | + | Nufo | Nufo | Nufo | + | + | - | + | + | | Nufo | + | + | Nufo | - |
| 6q | Nufo | + | + | Nufo | Nufo | Nufo | + | + | + | + | + | | Nufo | + | - | - | Nufo |
| 7p | Nufo | Nufo | - | Nufo | + | Nufo | + | + | + | + | + | | Nufo | + | + | Nufo | + |
| 7q | Nufo | Nufo | + | Nufo | Nufo | Nufo | + | + | + | - | + | | | + | Nufo | Nufo | + |
| 8p | - | - | + | + | + | Nufo | | | | | | | | | | Nufo | Nufo |
| 8q | + | + | + | + | + | | + | + | + | + | + | | | + | - | Nufo | + |
| 9p | + | - | + | Nufo | Nufo | Nufo | + | - | + | - | + | | | | | Nufo | - |
| 9q | Nufo | - | Nufo | Nufo | Nufo | Nufo | + | - | + | - | - | | | | | Nufo | - |
| 11p | Nufo | Nufo | Nufo | Nufo | Nufo | Nufo | + | Nufo | - | + | - | | | - | + | Nufo | Nufo |
| 12p | - | - | + | Nufo | Nufo | - | + | + | Nufo | + | | | | + | Nufo | Nufo | + |
| 12q | + | Nufo | + | Nufo | - | Nufo | - | + | Nufo | + | | | | + | Nufo | + | |
| 13q | - | Nufo | - | Nufo | Nufo | Nufo | + | + | + | + | | | | Nufo | Nufo | | |
| 14q | Nufo | - | + | Nufo | Nufo | Nufo | | | | | - | | Nufo | - | Nufo | Nufo | - |
| 15p | Nufo | - | Nufo | Nufo | - | Nufo | - | Nufo | - | + | - | | Nufo | | | Nufo | Nufo |
| 16q | Nufo | - | Nufo | Nufo | Nufo | Nufo | Nufo | Nufo | + | Nufo | | | | Nufo | + | Nufo | Nufo |
| 17p | | | | Nufo | | | | | | | | | | | | Nufo | Nufo |
| 17q | Nufo | Nufo | - | | | | | | | | Nufo | Nufo | + | Nufo | Nufo | Nufo | Nufo |
| 18p | - | Nufo | - | Nufo | Nufo | Nufo | | + | | | + | | Nufo | Nufo | Nufo | Nufo | + |
| 18q | Nufo | Nufo | - | Nufo | + | Nufo | | | | | + | | Nufo | Nufo | Nufo | Nufo | + |
| 22q | - | Nufo | Nufo | - | - | Nufo | | | + | | Nufo | Nufo | + | + | Nufo | + | + |

FIG. 7

FIG. 8A

FIG. 8B

FIG. 8C

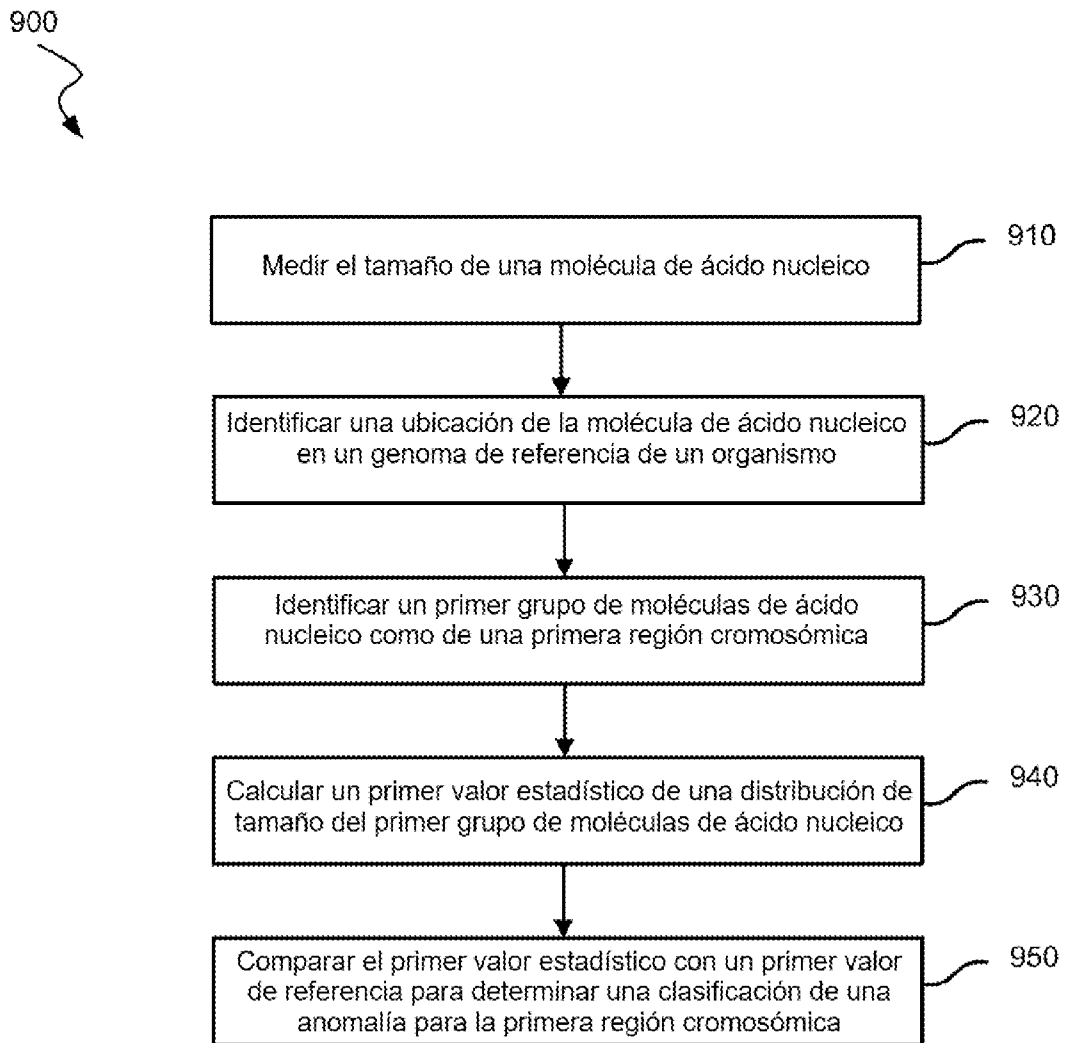


FIG. 9

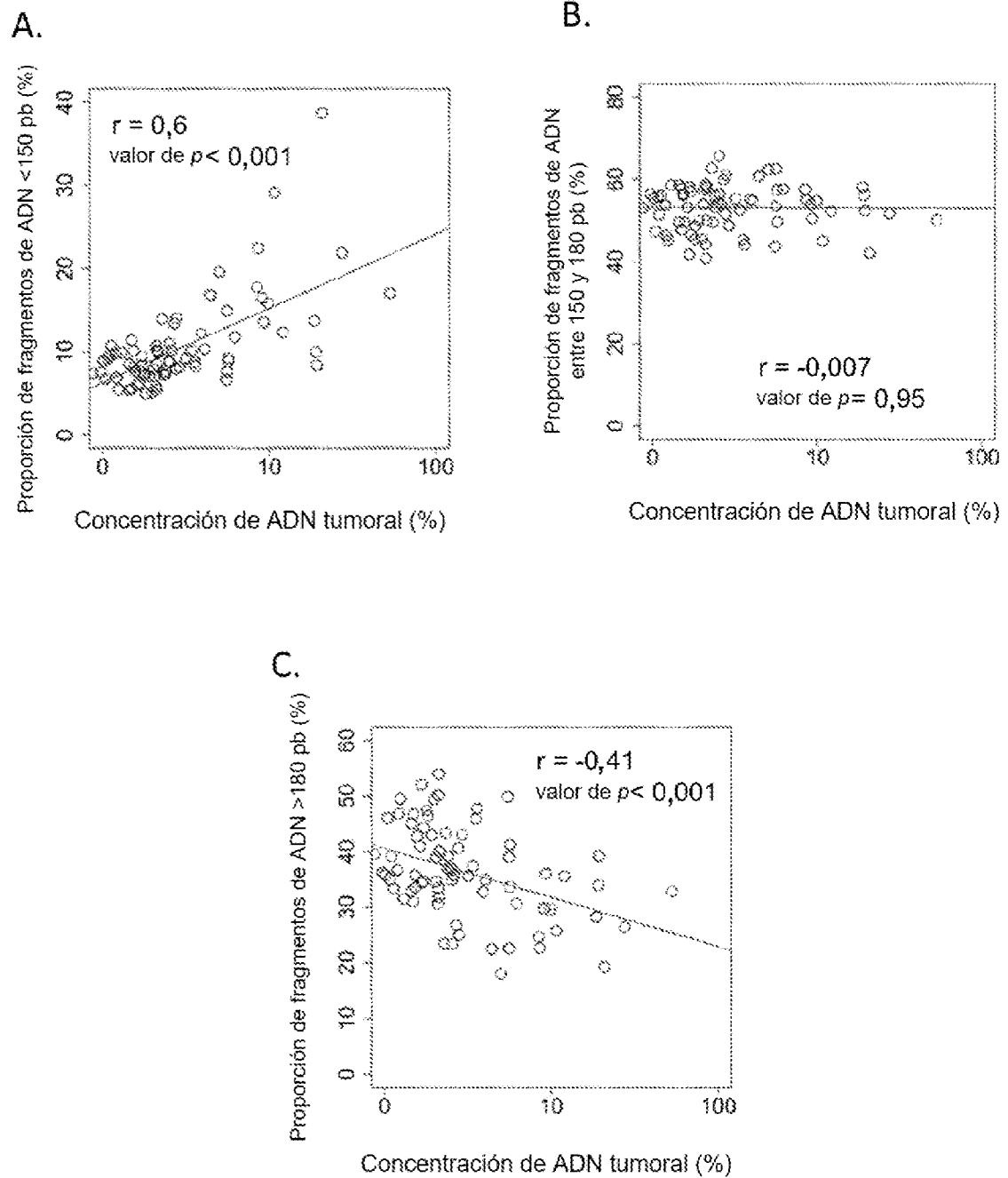
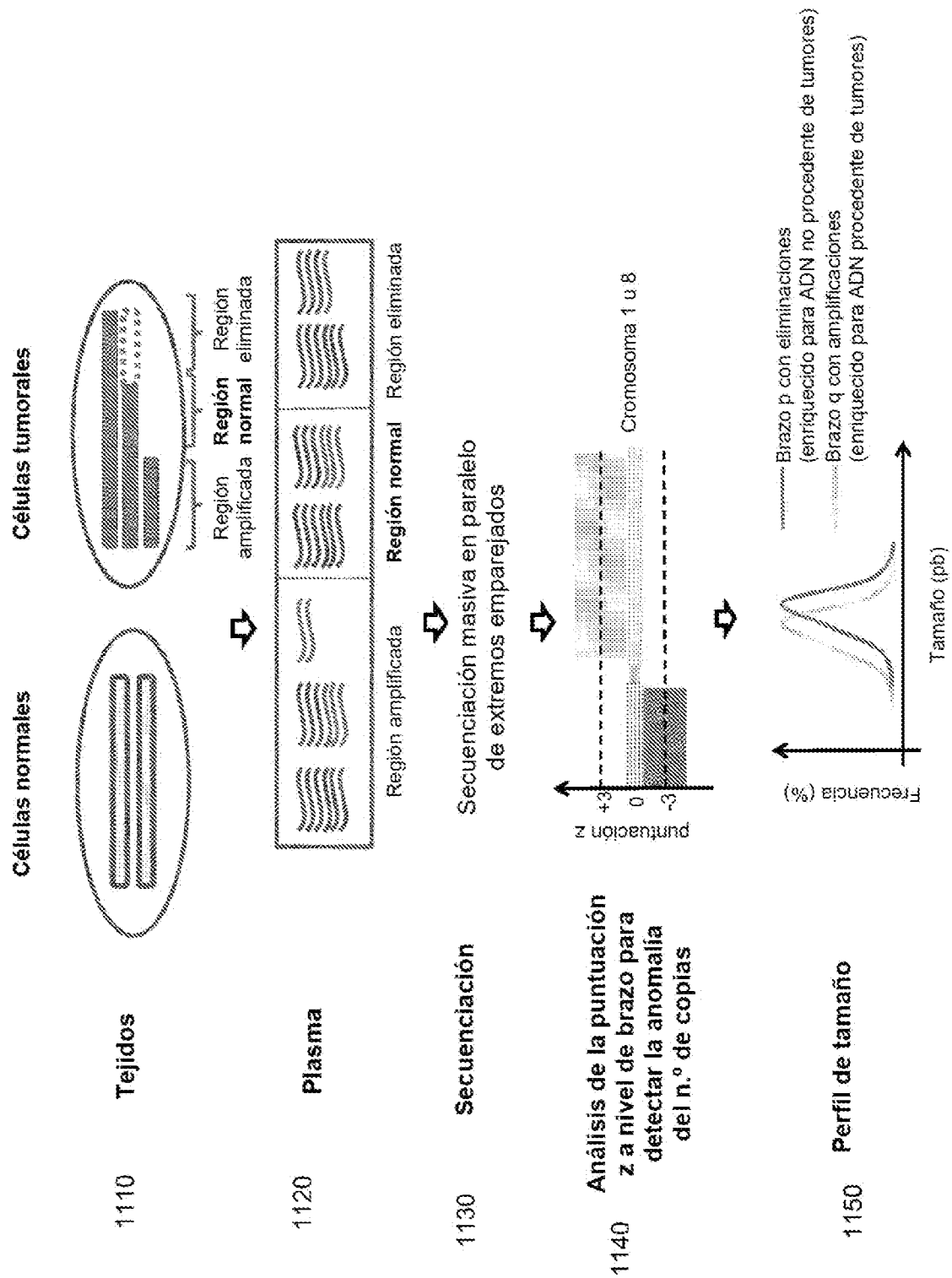


FIG. 10



11-19-61

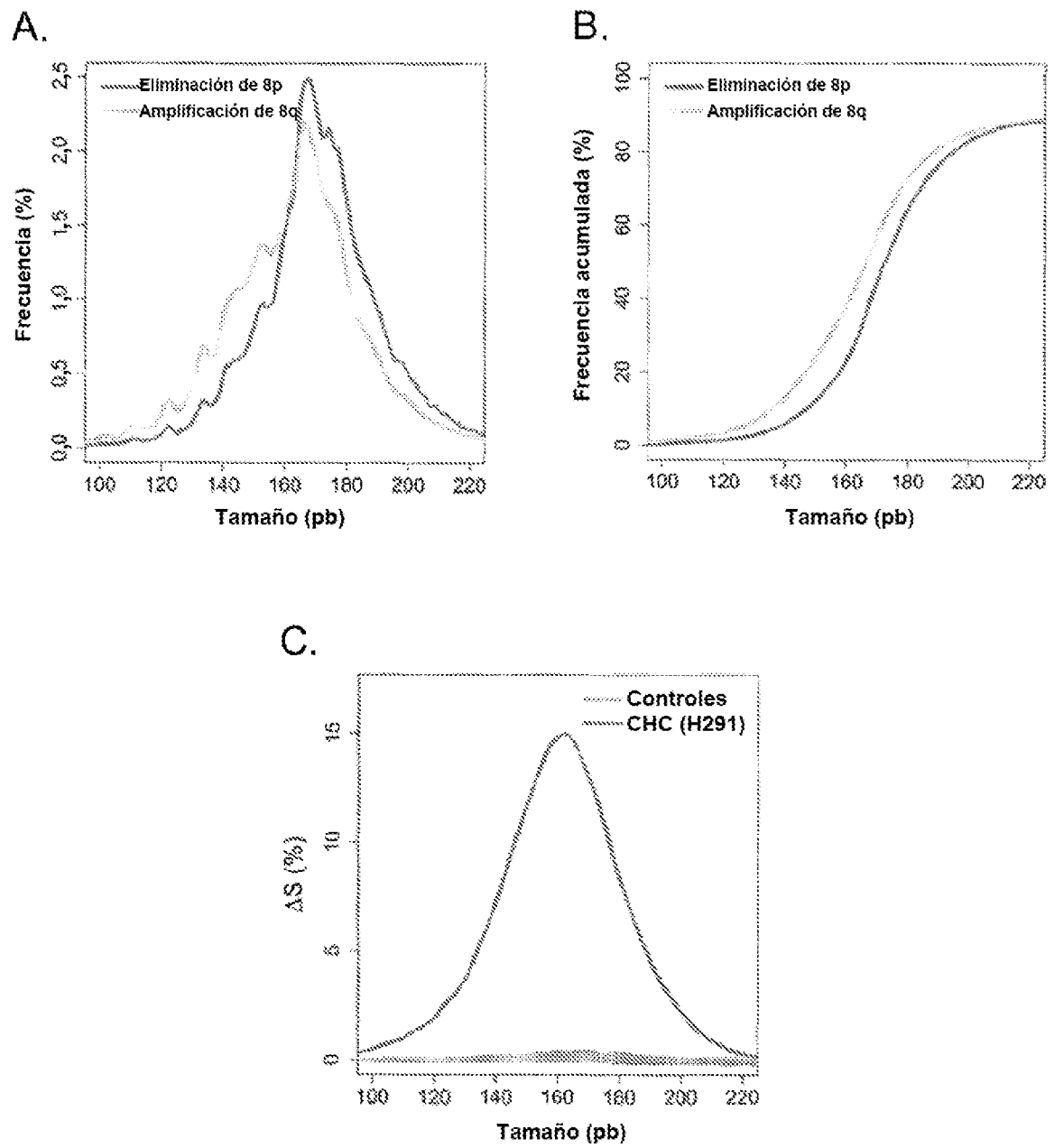


FIG. 12

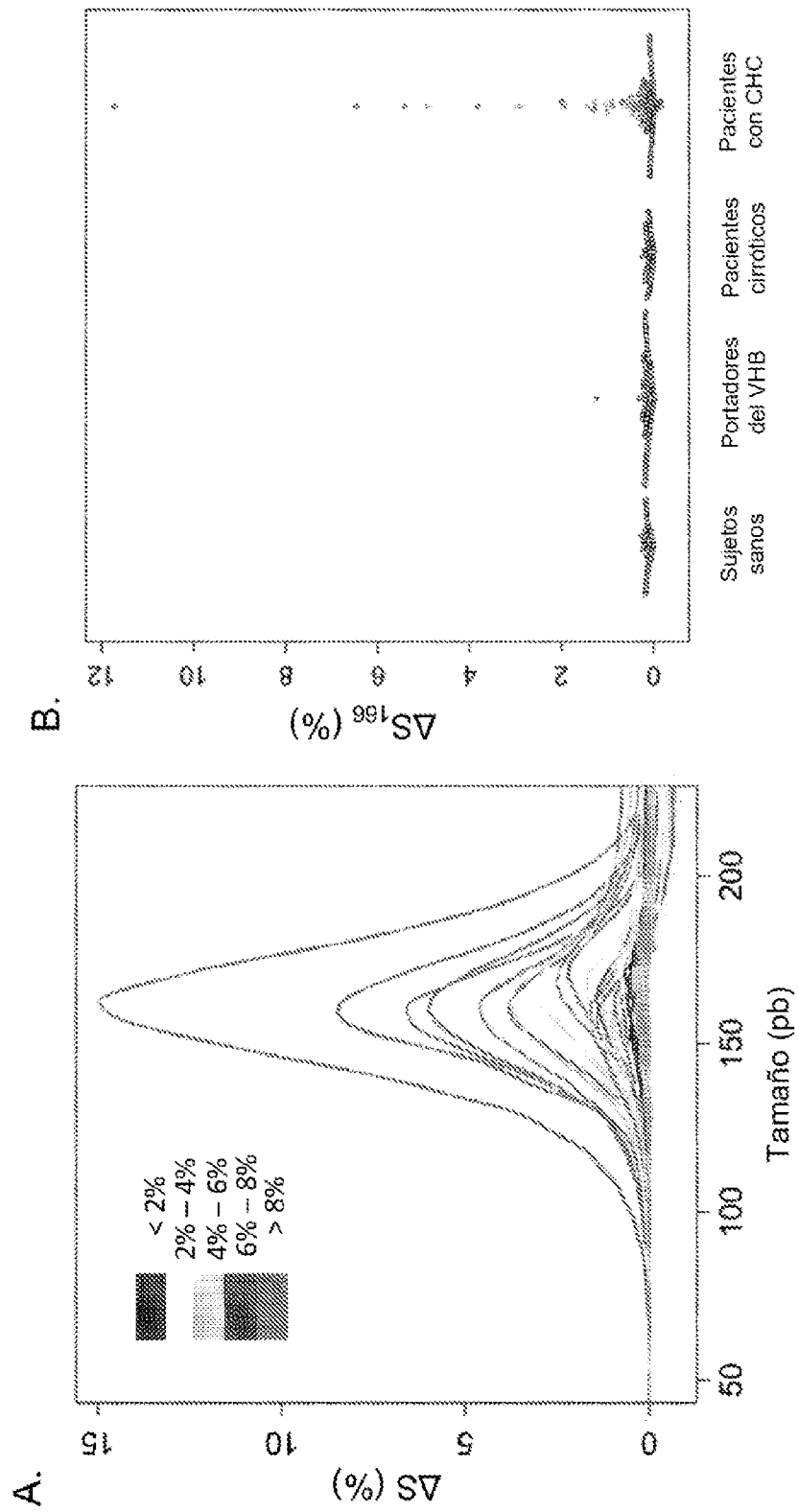


FIG. 13

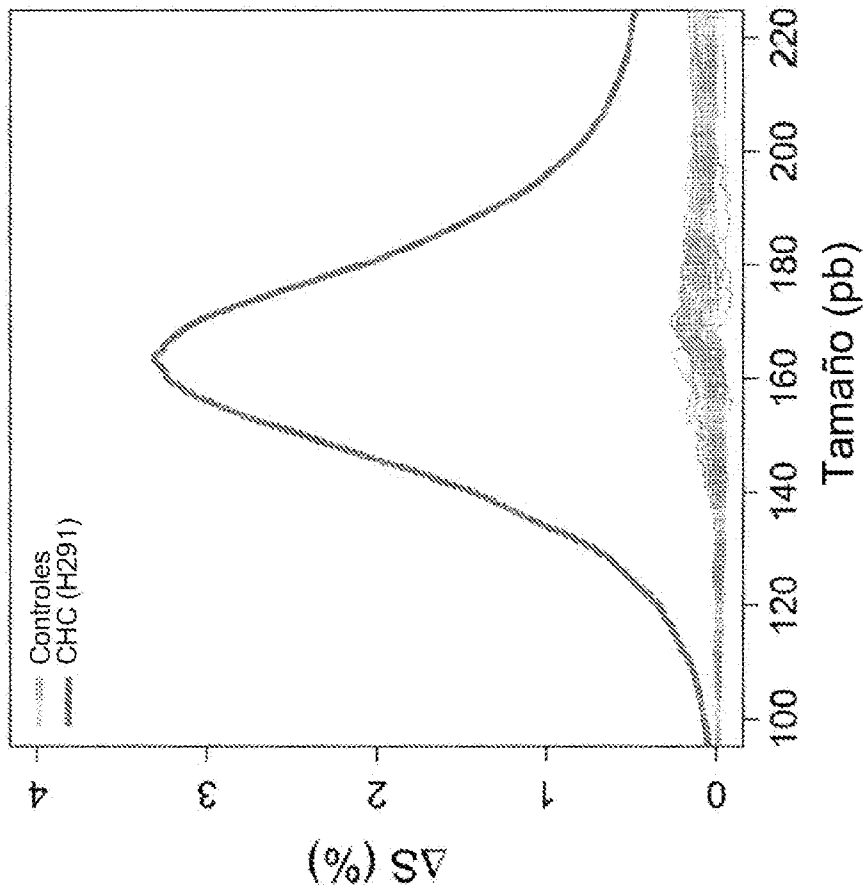


FIG. 14

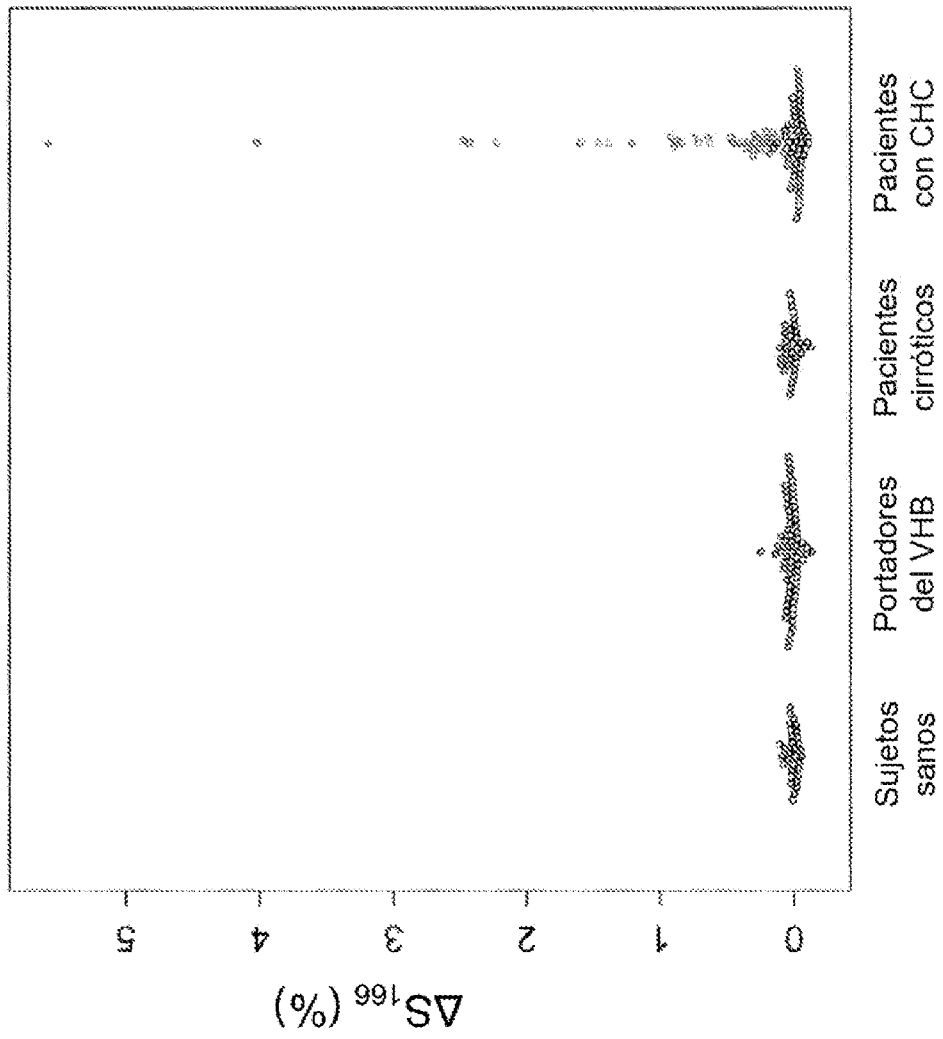


FIG. 15

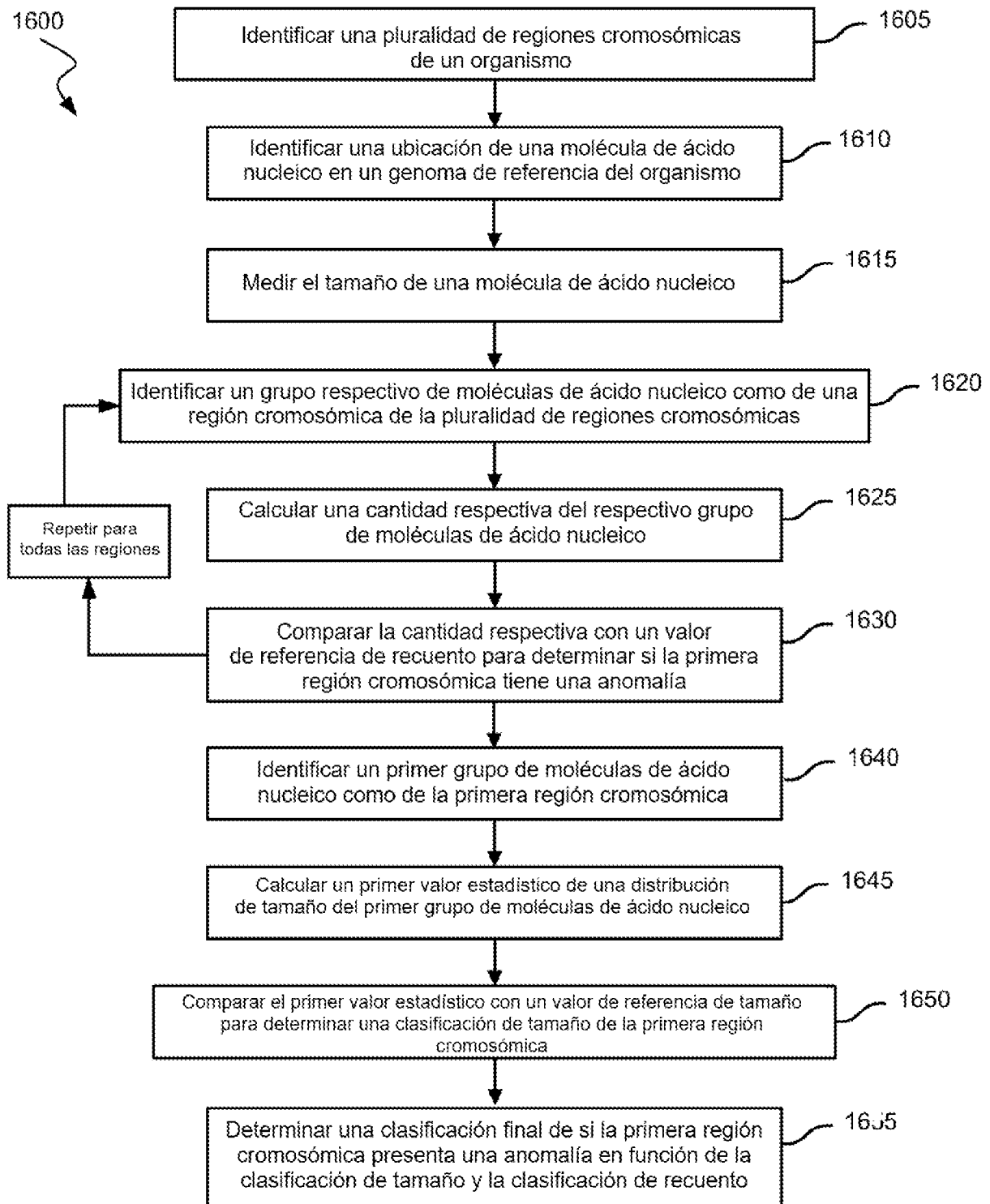


FIG. 16

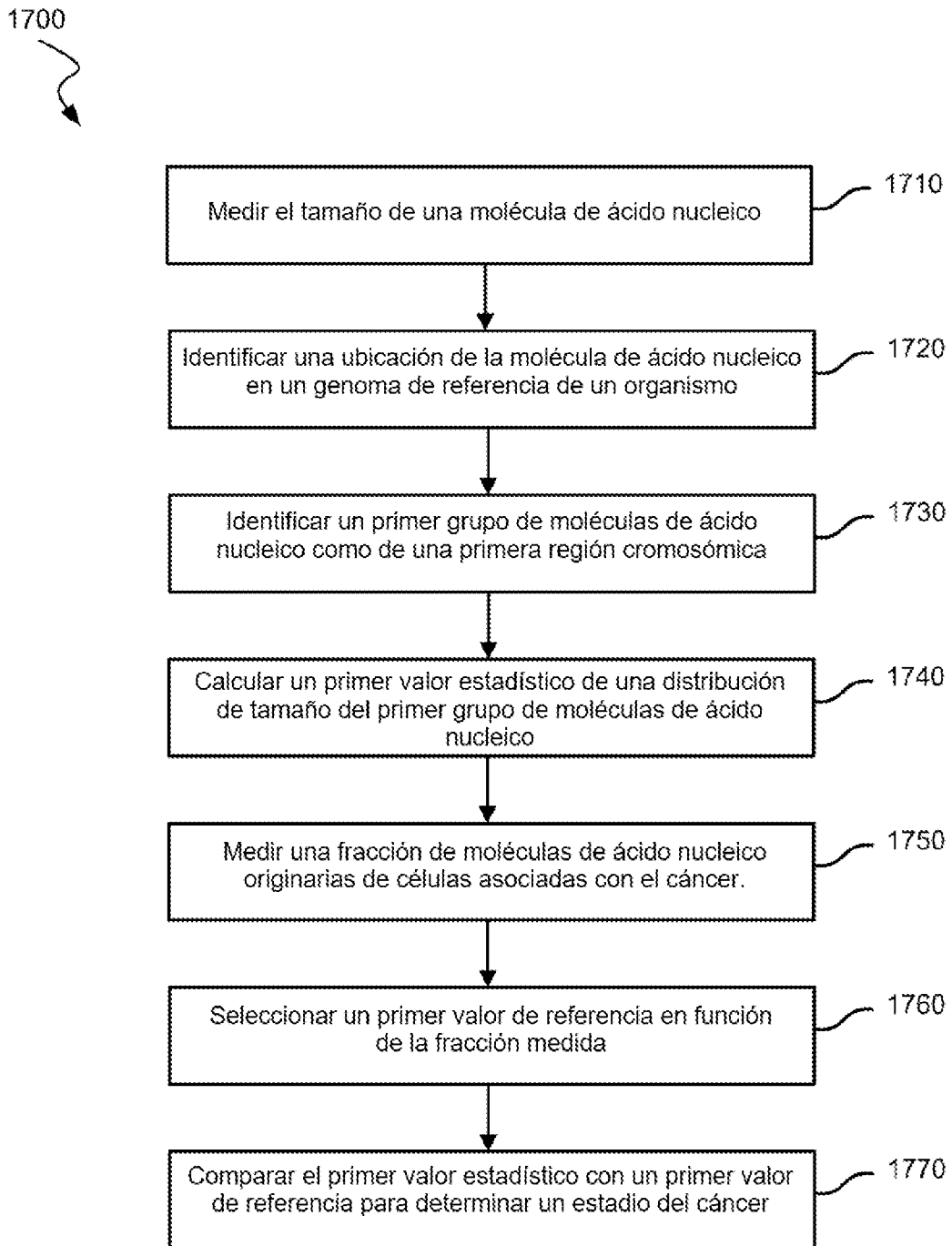


FIG. 17

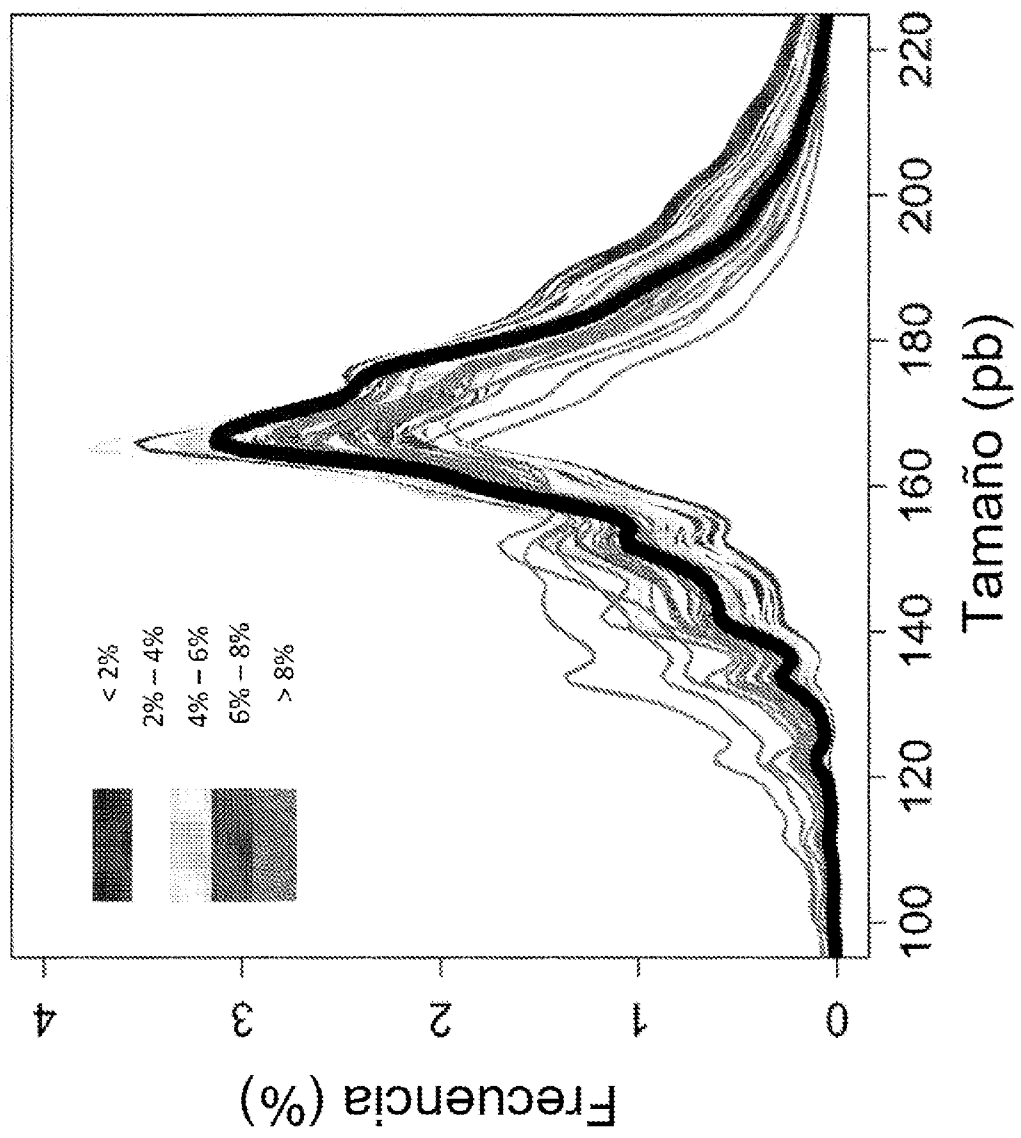


FIG. 18

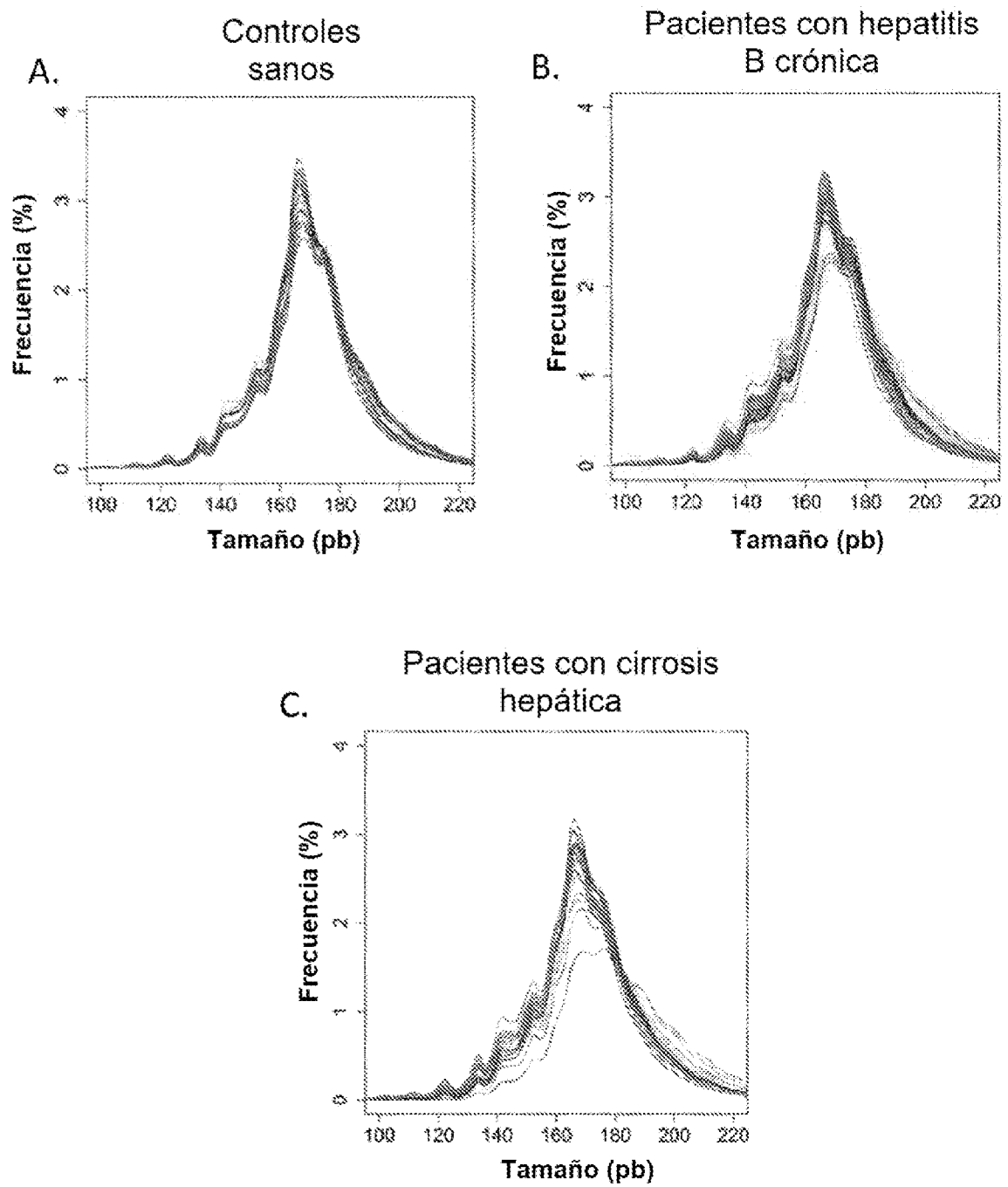


FIG. 19

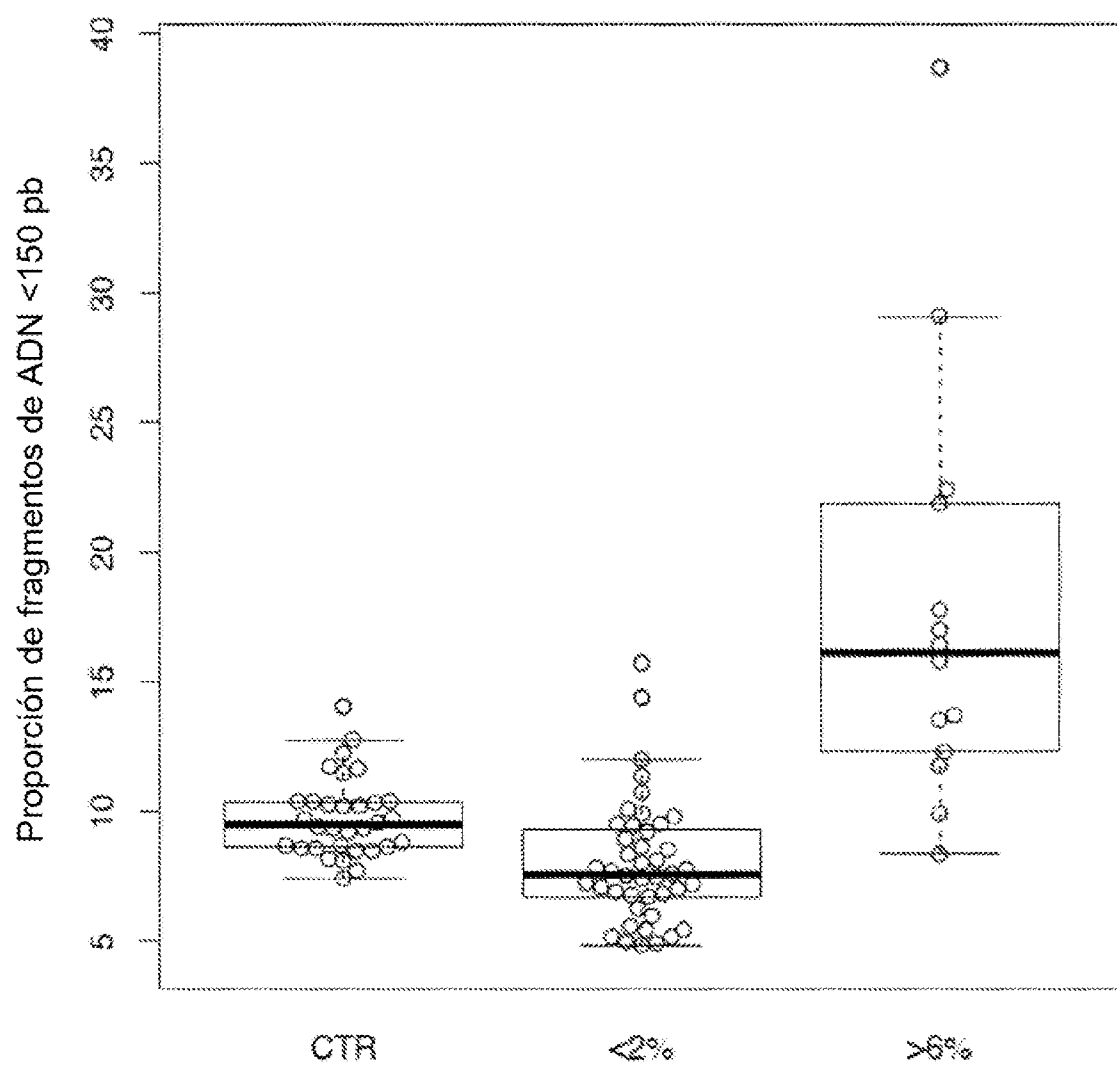


FIG. 20

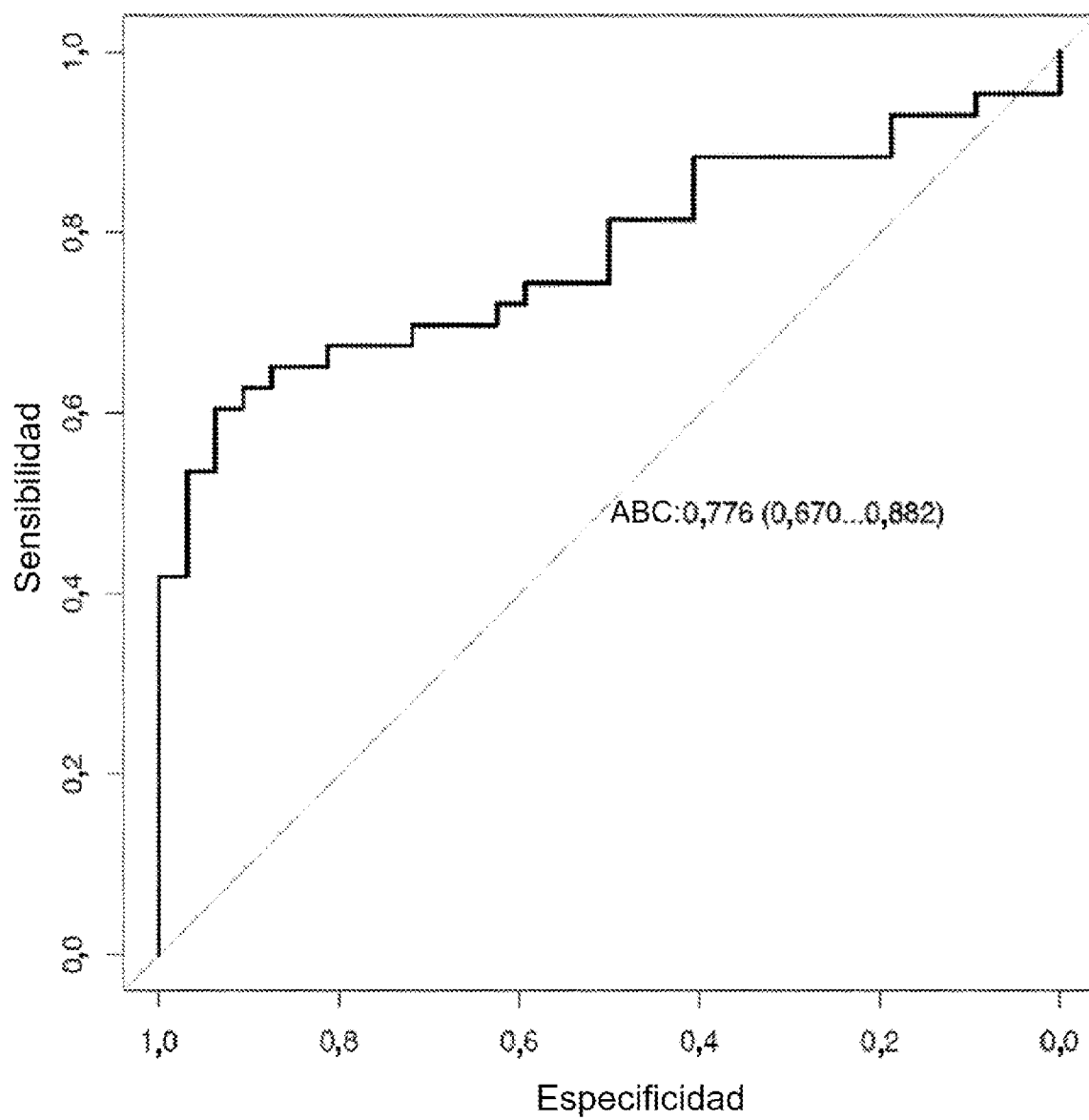


FIG. 21

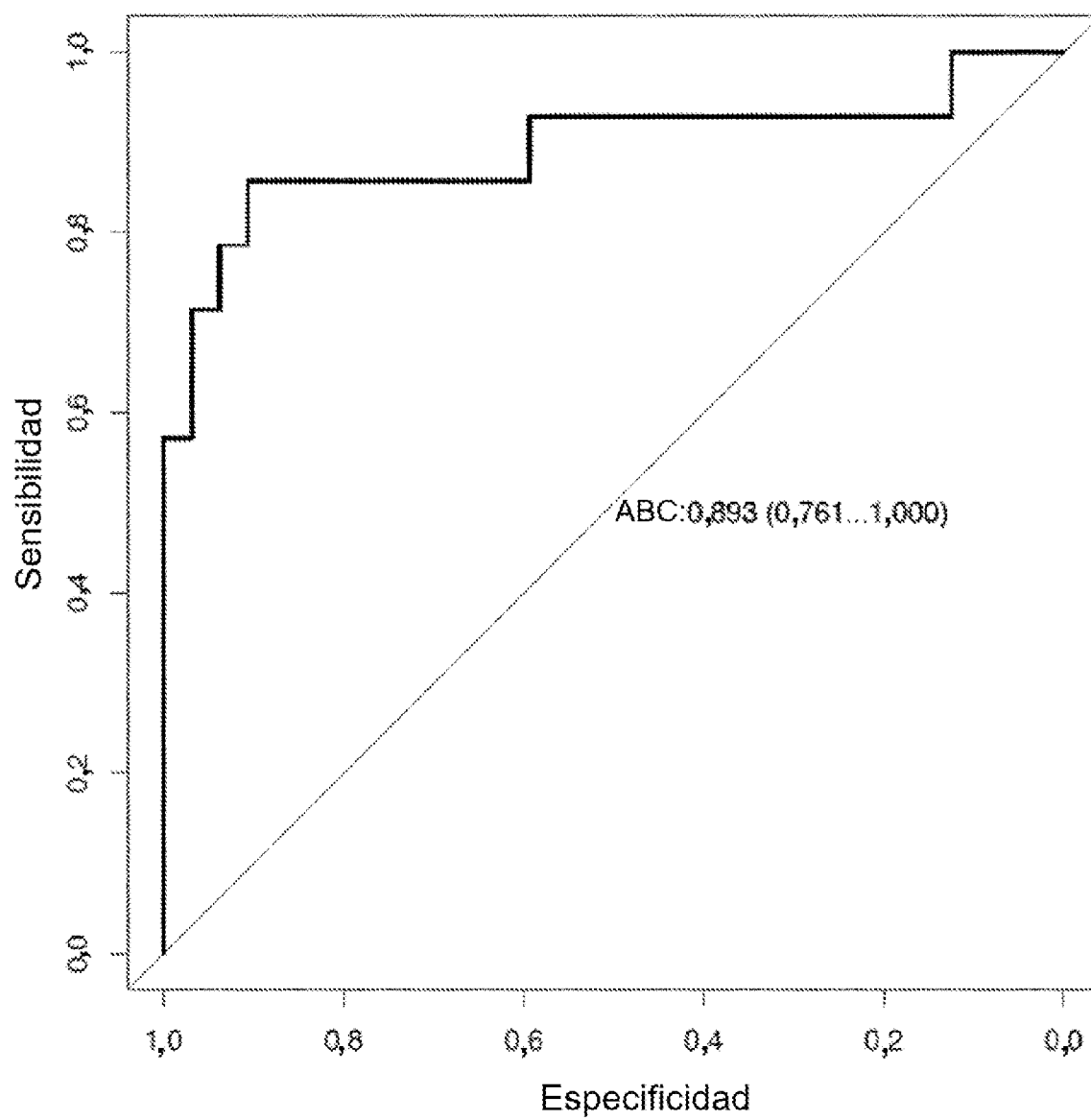


FIG. 22

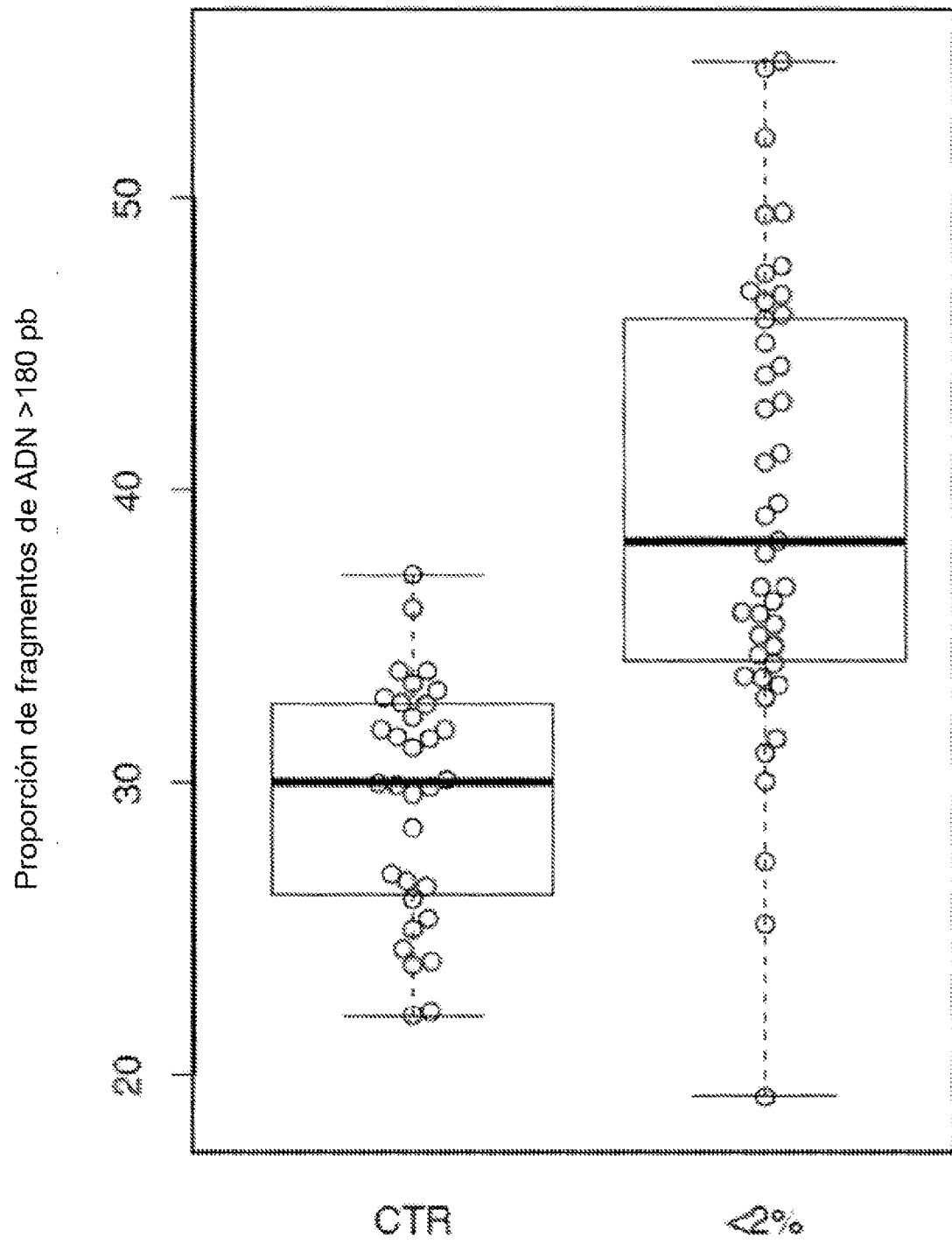


FIG. 23

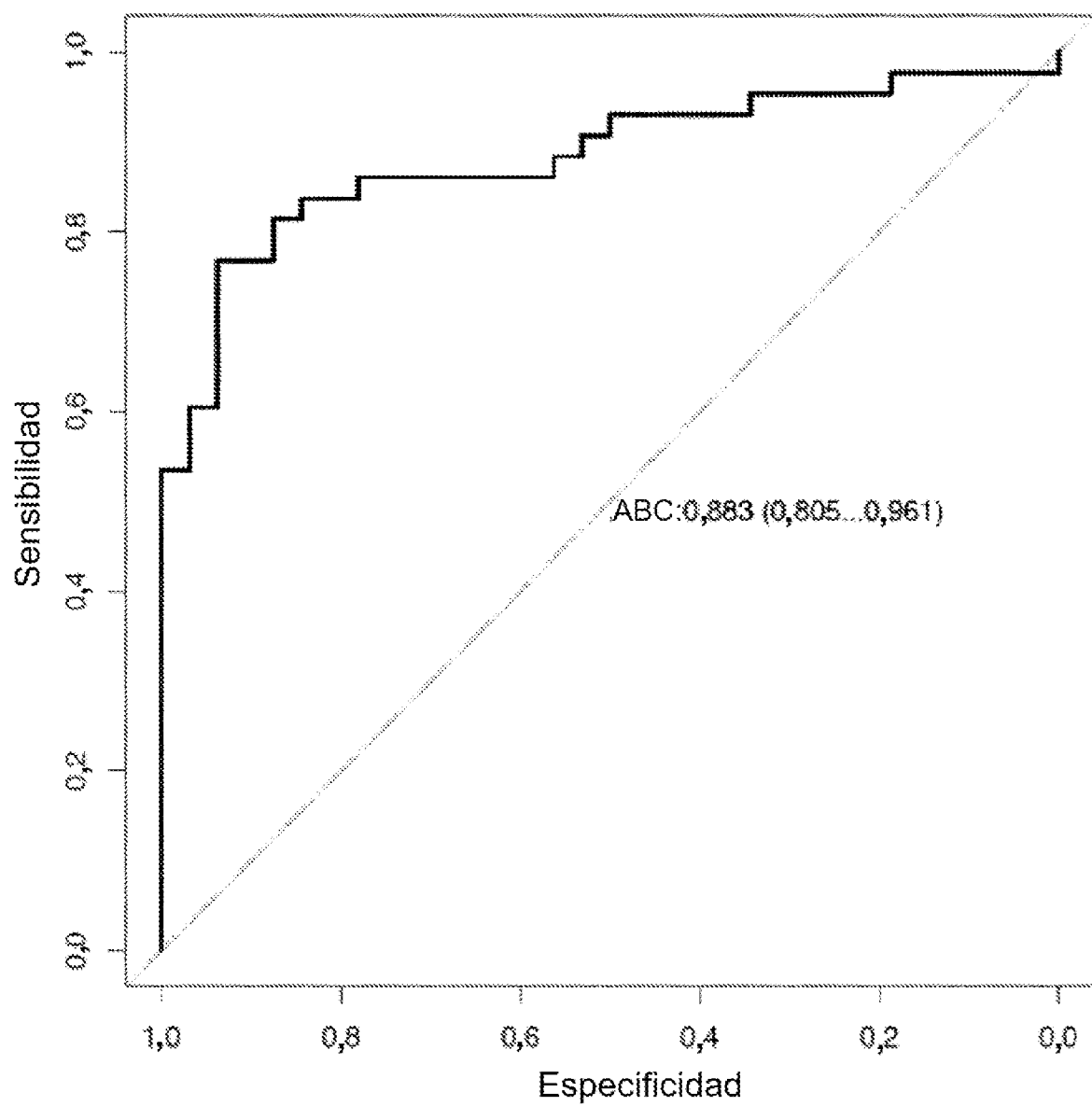


FIG. 24

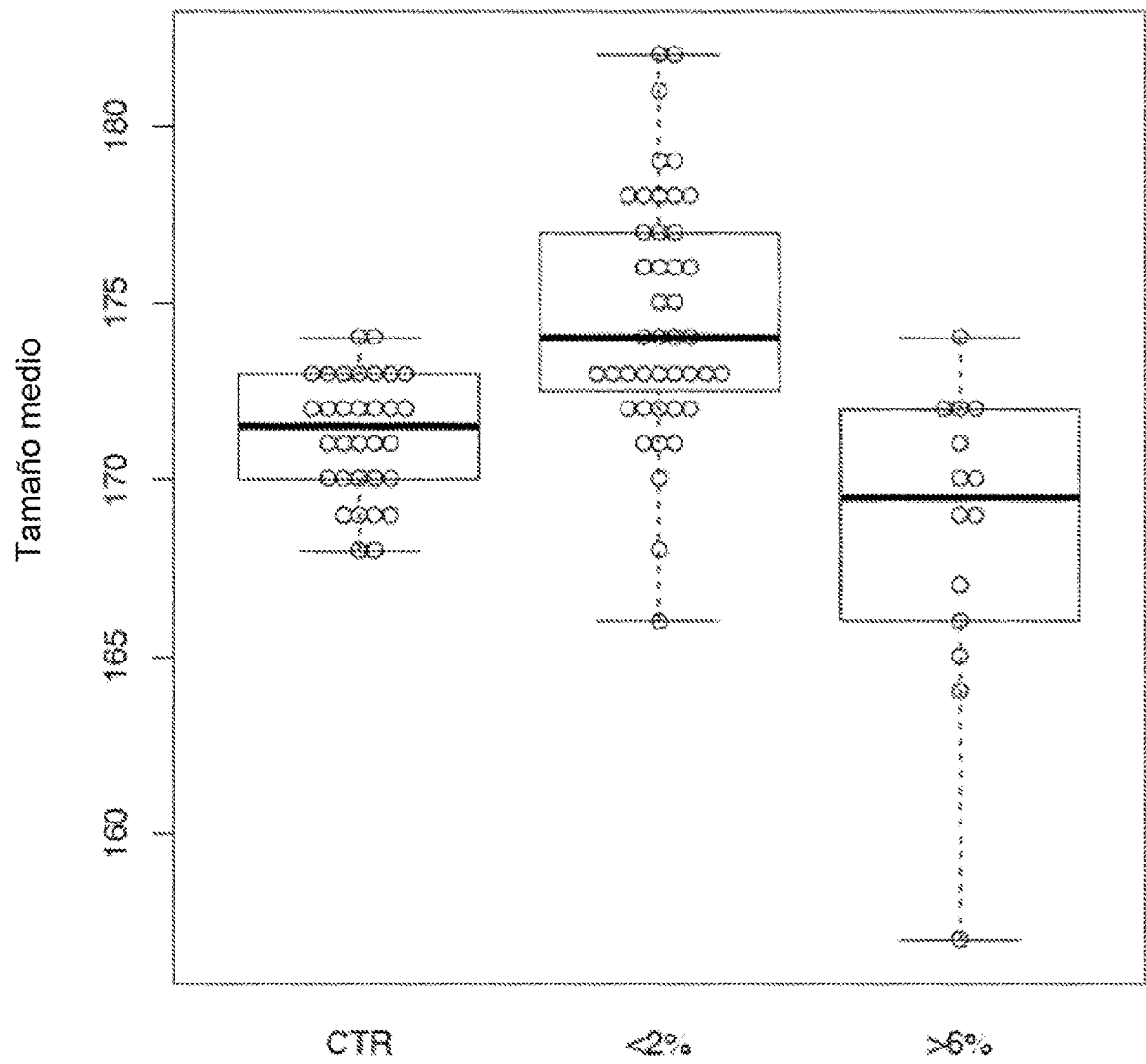


FIG. 25

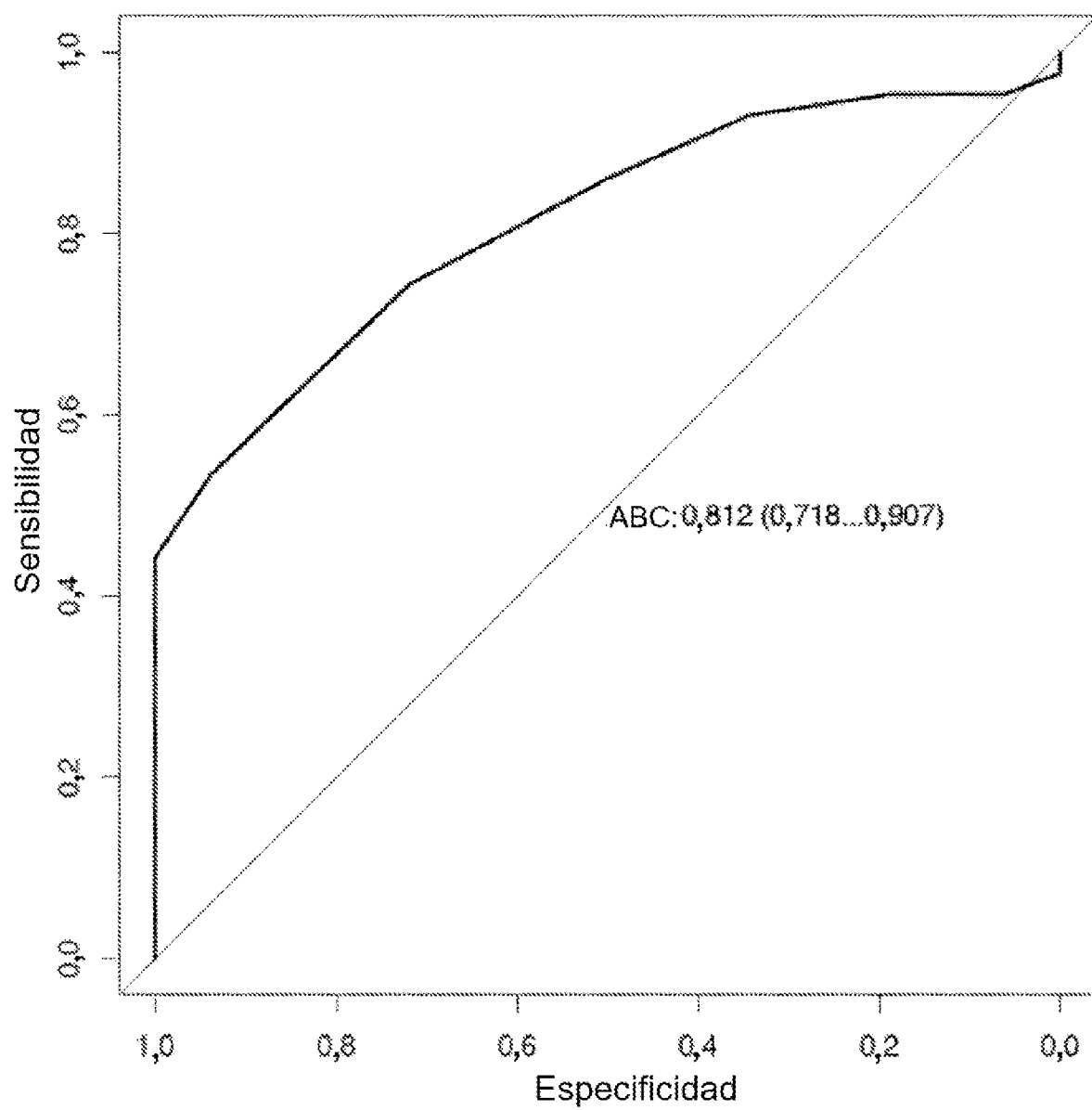


FIG. 26

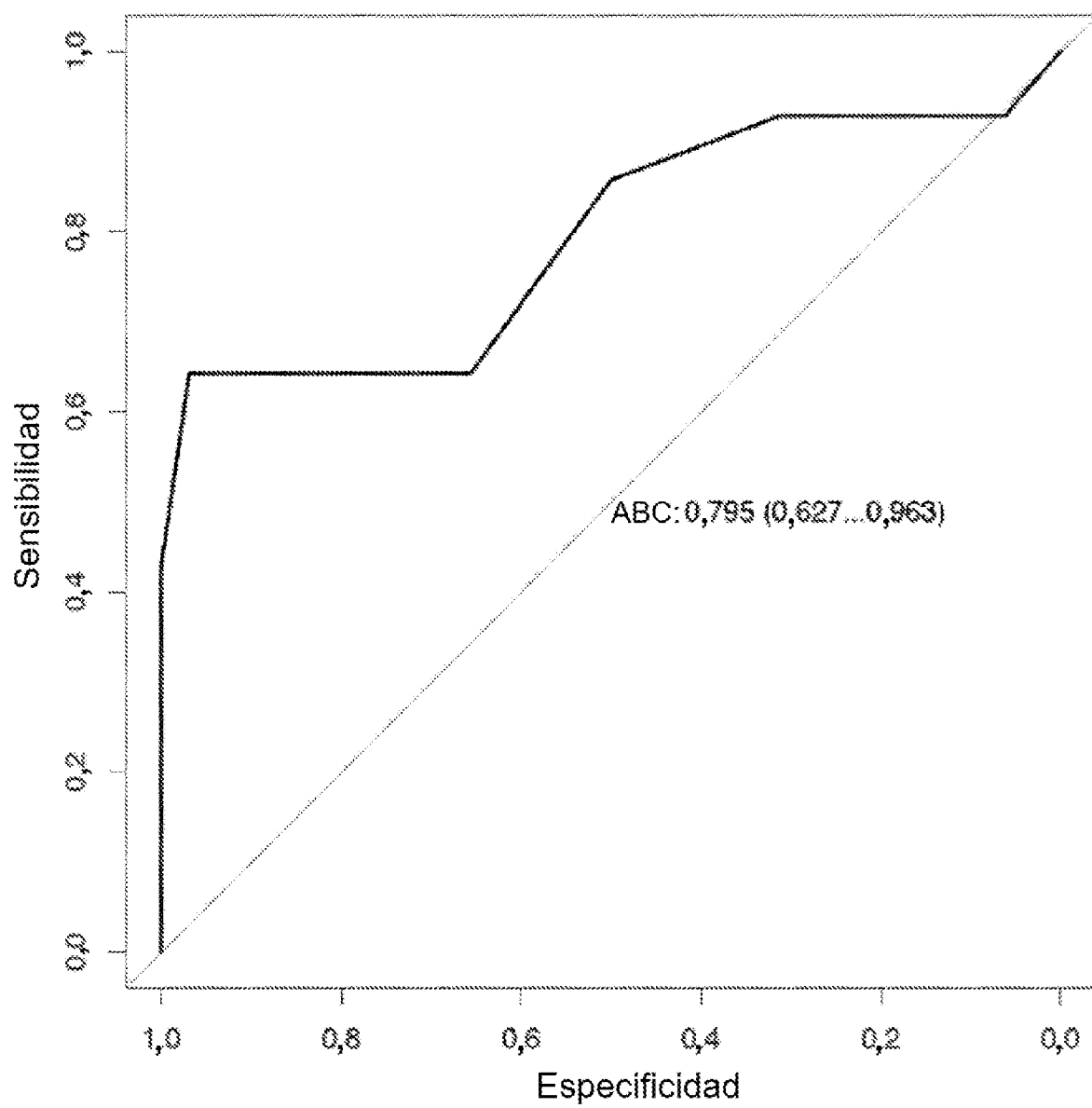


FIG. 27

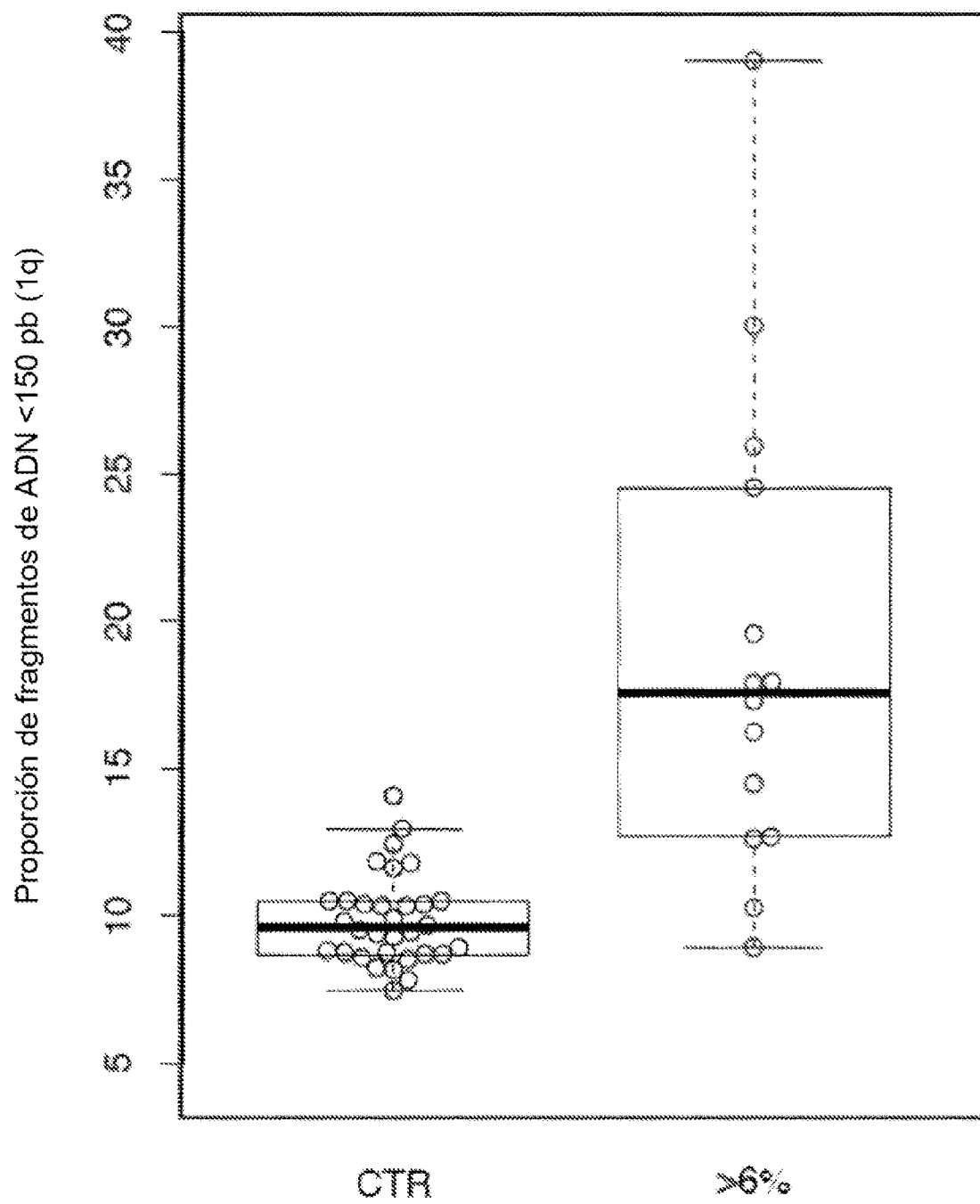


FIG. 28

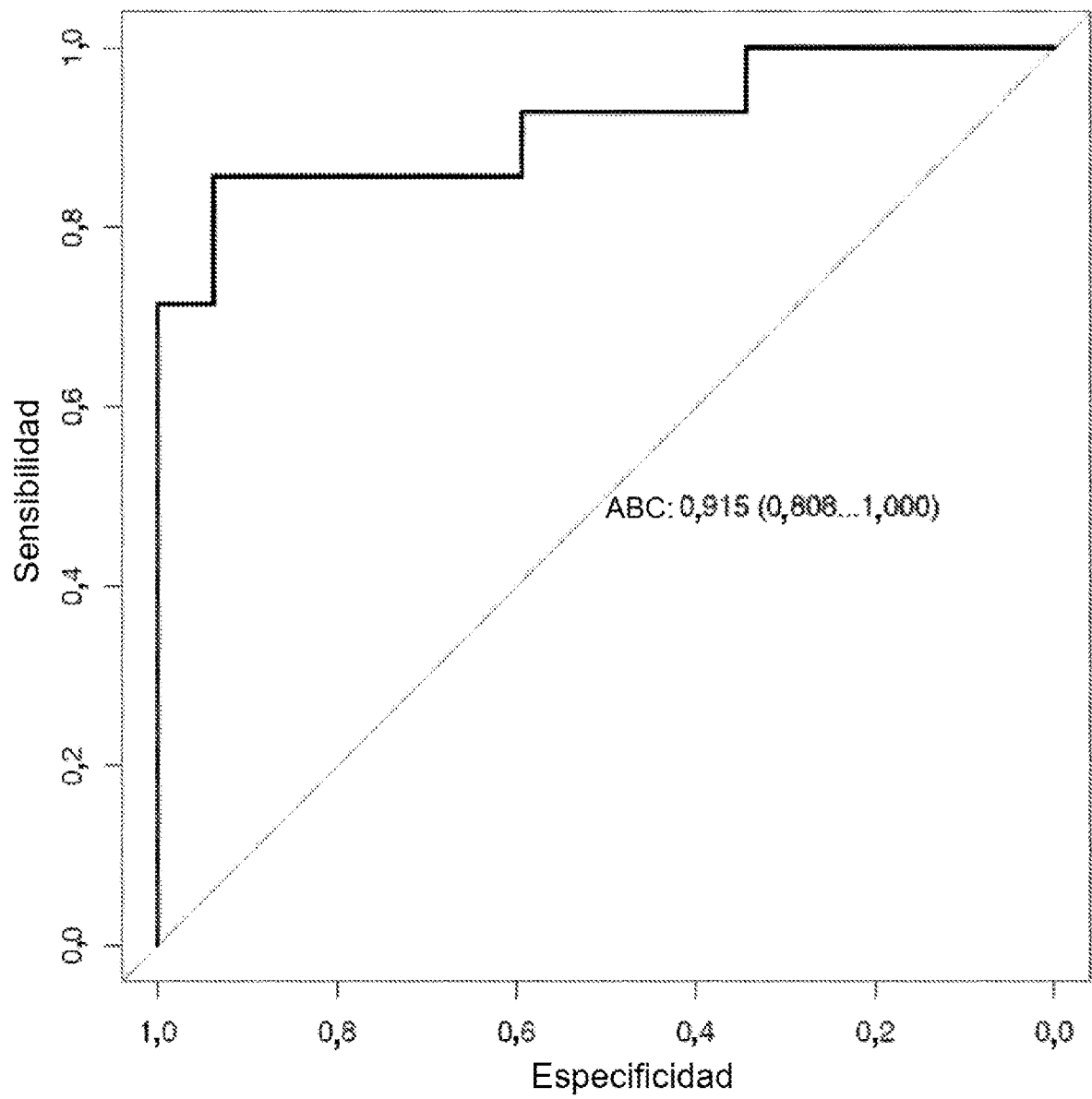


FIG. 29

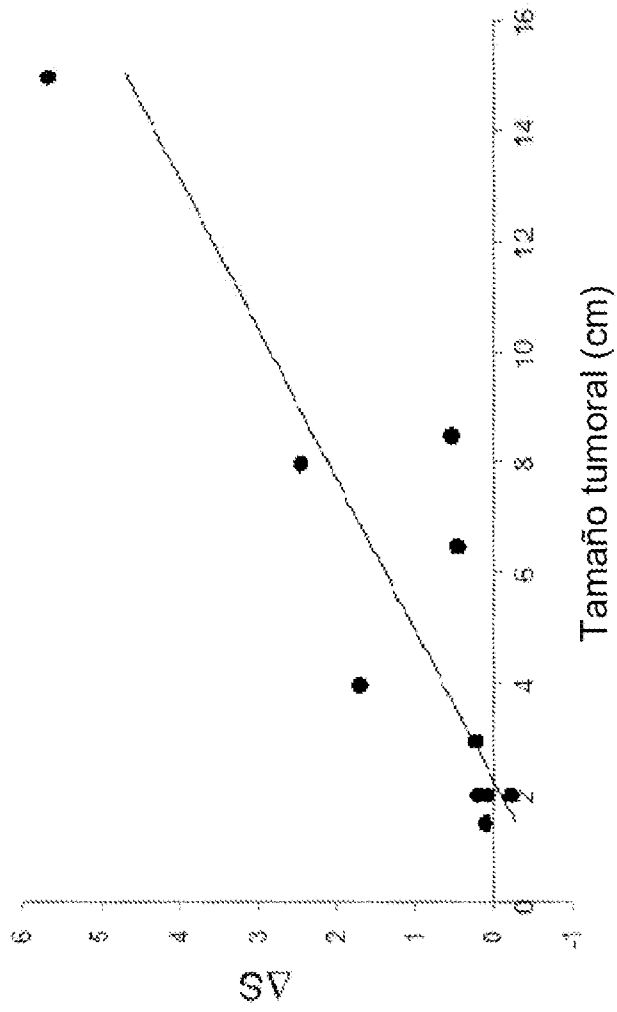


FIG. 30

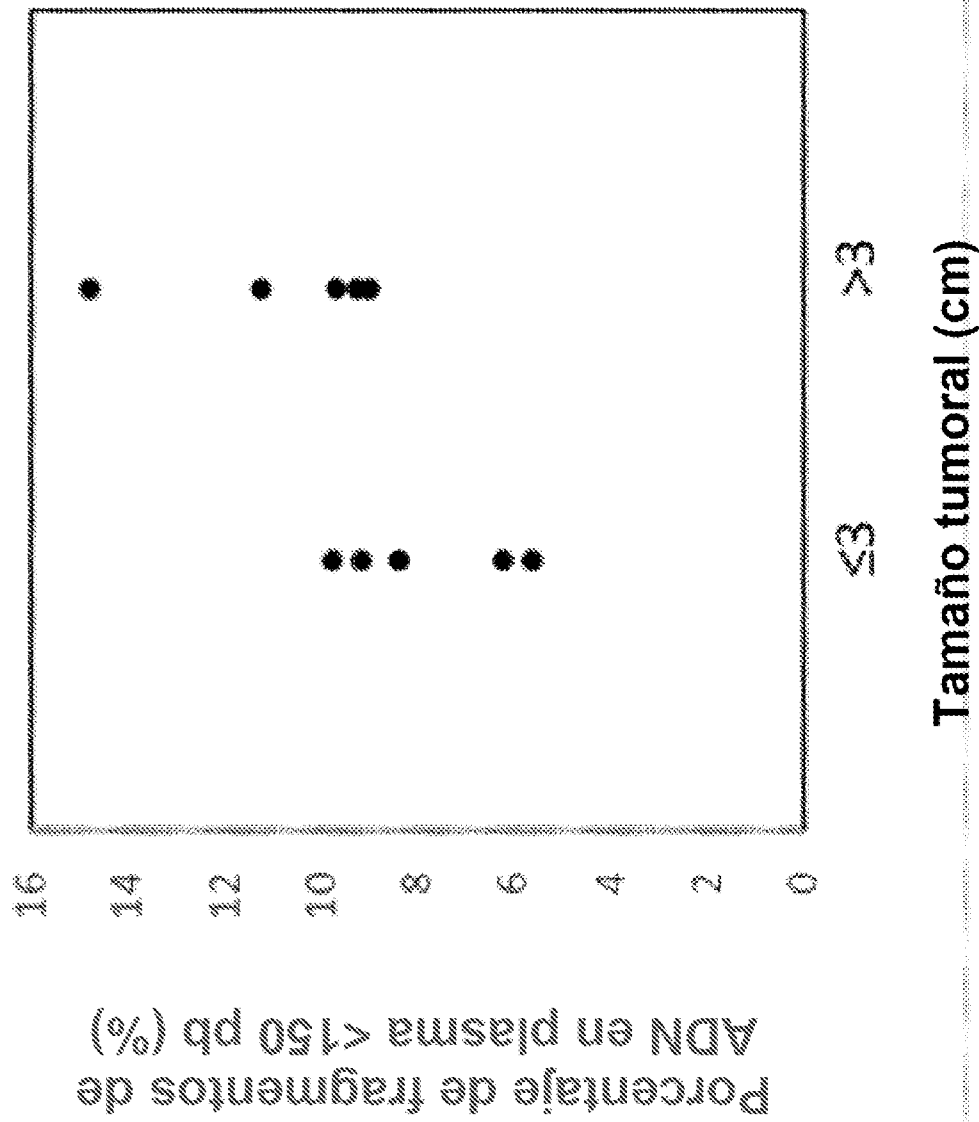


FIG. 31

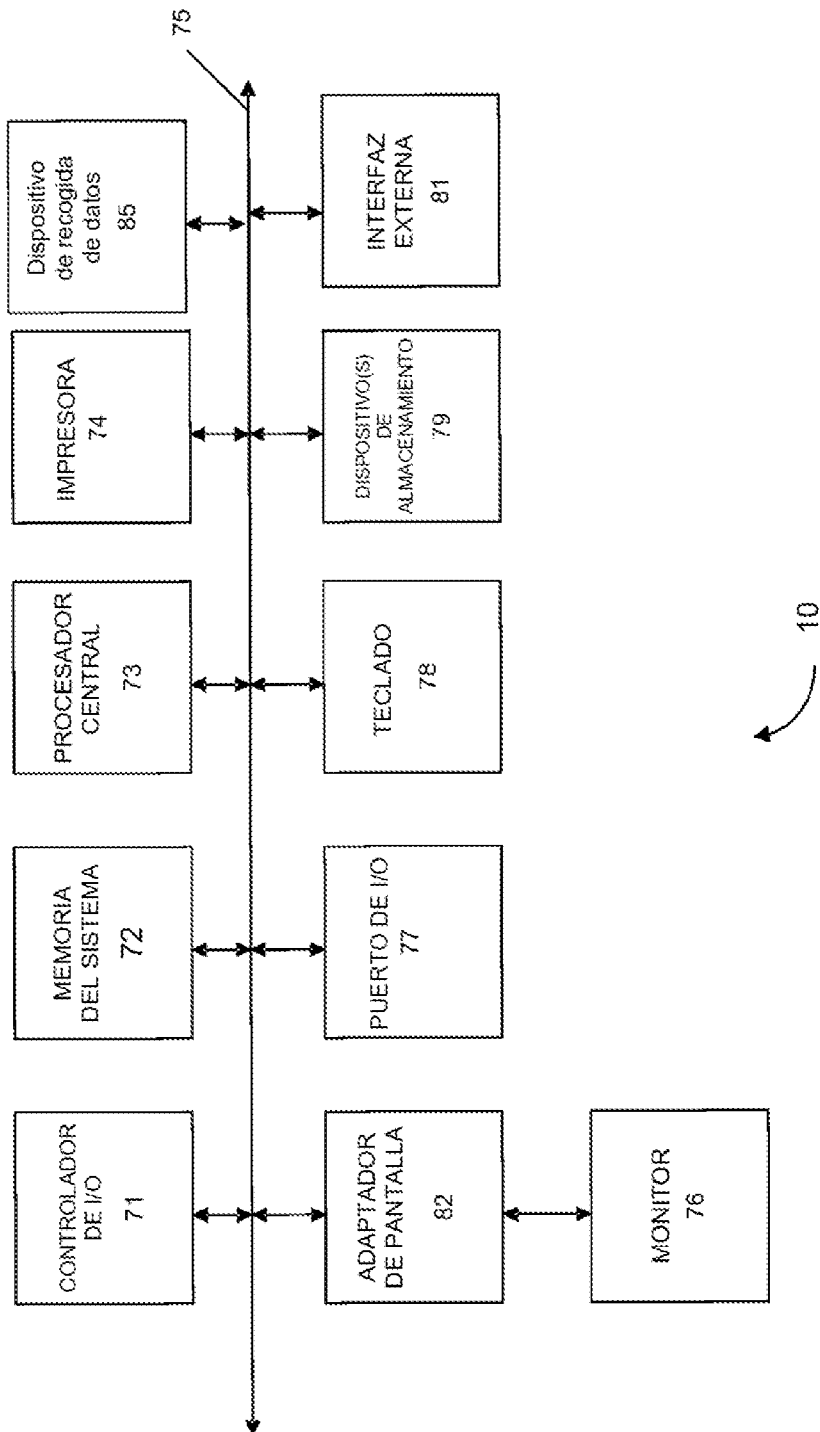


FIG. 32