

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 761 624**

51 Int. Cl.:

C12Q 1/6869 (2008.01)

C12Q 1/6827 (2008.01)

C12Q 1/6809 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.03.2013 E 17209781 (8)**

97 Fecha y número de publicación de la concesión europea: **16.10.2019 EP 3354748**

54 Título: **Análisis de ADN basado en el tamaño para la clasificación del cáncer**

30 Prioridad:

08.03.2012 US 201261608623 P
06.04.2012 US 201261621451 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
20.05.2020

73 Titular/es:

THE CHINESE UNIVERSITY OF HONG KONG
(100.0%)
Knowledge Transfer Office Sha Tin New
Territories
Hong Kong, CN

72 Inventor/es:

LO, YUK MING DENNIS;
CHAN, KWAN CHEE;
ZHENG, WENLI;
JIANG, PEIYONG;
LIAO, JIAWEI y
CHIU, WAI KWUN ROSSA

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 761 624 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

Análisis de ADN basado en el tamaño para la clasificación del cáncer

5 El descubrimiento de ADN fetal sin células en el plasma materno ha abierto nuevas posibilidades para el diagnóstico prenatal no invasivo (Lo YMD y col. *Lancet* 1997; 350:485-487). Se ha descrito que la concentración fraccionaria de ADN fetal medio/mediana es aproximadamente de 3% a 10% (Lo YMD y col. *Am J Hum Genet* 1998; 62:768-775; Lun FMF y col. *Clin Chem* 2008; 54:1664-1672). La concentración fraccionaria de ADN fetal es un parámetro importante que afecta al funcionamiento de las pruebas de diagnóstico prenatal no invasivas que usan ADN plasmático materno. Por ejemplo, para el diagnóstico prenatal no invasivo de aneuploidías cromosómicas fetales (p. ej., trisomía 21, trisomía 18 o trisomía 13), cuanto mayor es la concentración fraccionaria de ADN fetal, mayor es la sobrerrepresentación de las secuencias de ADN derivadas del cromosoma aneuploide en el plasma materno. De hecho, se ha demostrado que, por cada reducción de dos veces de la concentración fraccionaria de ADN fetal en el plasma materno, el número de moléculas que deberían contarse para lograr la detección de aneuploidía sería de cuatro veces (Lo YMD y col. *Proc Natl Acad Sci USA* 2007;104:13116-13121).

Para la detección prenatal no invasiva de la trisomía fetal por secuenciación masiva en paralela aleatoria, la concentración fraccionaria de ADN fetal de una muestra afectaría a la cantidad de secuenciación que se necesitaría llevar a cabo para lograr una detección consistente (Fan HC and Quake SR. *PLoS One* 2010;5:e10439). De hecho, una serie de grupos han incluido una etapa de control de calidad en la que primero se mide la concentración fraccionaria de ADN fetal y solo las muestras que contienen una concentración fraccionaria de ADN fetal mínima serán elegidas para generar un resultado de diagnóstico (Palomaki GE y col. *Genet Med* 2011;13:913-920). Otros grupos han incluido la concentración fraccionaria de ADN fetal en su algoritmo de diagnóstico para calcular el riesgo de que se obtenga una muestra de plasma materno particular de un embarazo aneuploide (Sparks AB y col. *Am J Obstet Gynecol* 2012; 206: 319.e1-9). Fan et al., "Analysis of Size Distributions of Fetal and Maternal Cell-Free DNA by Paired-End Sequencing", *Clinical Chemistry*, vol. 56, n.º 8, 1 de agosto de 2010, pp. 1279-1286 desvela la medición de la distribución de la longitud del plasma materno sin células con resolución de base única usando la secuencia final pareada para aumentar la sensibilidad en la detección de aneuploidía fetal.

Además de la detección de aneuploidía, la concentración fraccionaria de ADN fetal también afecta de forma similar a las pruebas de diagnóstico prenatal no invasivas llevadas a cabo usando ADN plasmático materno para detectar enfermedades monogénicas, p. ej., las hemoglobinopatías (Lun FMF y col. *Proc Natl Acad Sci USA* 2008;105:19920-19925) y la hemofilia (Tsui NBY y col. *Blood* 2011;117:3684-3691). La concentración fraccionaria de ADN fetal también afecta la profundidad de la secuenciación que se necesitaría llevar a cabo para construir un mapa mutacional y genético de todo el genoma fetal, así como la secuenciación del genoma fetal completo (Lo YMD y col. *Sci Transl Med* 2010;2:61ra91 y el documento US-2011/0105353).

Se han descrito una serie de procedimientos para medir la concentración fraccionaria de ADN fetal. Un procedimiento es medir la concentración de una secuencia de herencia paterna específica fetal, que esté ausente en el genoma materno. Los ejemplos de dichas secuencias incluyen las secuencias en el cromosoma Y que están presentes en fetos de varones y las secuencias del gen *RHD* en un feto Rhesus D positivo que porta una mujer embarazada Rhesus D negativo. También se podría medir el ADN plasmático materno total usando secuencias que están presentes tanto en la madre como en el feto. Para llegar a una concentración fraccionaria de ADN fetal, se podría entonces calcular la relación de la concentración de la secuencia de herencia paterna específica fetal, frente a la concentración del ADN del plasma materno total.

Otro ejemplo de secuencias que se podrían usar incluye el uso de polimorfismos de un solo nucleótido (Lo YMD y col. *Sci Transl Med* 2010; 2:61ra91). Una desventaja de usar marcadores genéticos para la medición de la concentración fraccionaria de ADN fetal es que ningún conjunto individual de marcadores genéticos sería informativo para todas las parejas de feto-madre. Otro procedimiento más que se podría emplear, es el uso de marcadores de ADN que presentan patrones de metilación de ADN específicos del feto o placenta (Nygren AO y col. *Clin Chem* 2010;56:1627-1635). La potencial desventaja del uso de los marcadores de metilación del ADN es que puede haber variación entre individuos en el nivel de metilación del ADN. Además, los procedimientos que se usan para la detección de los marcadores de metilación del ADN típicamente son complejos, incluyendo el uso de digestión con enzimas de restricción sensibles a la metilación (Chan KCA y col. *Clin Chem* 2008; 52:2211-2218) o la conversión con bisulfito (Chim SSC y col. *Proc Natl Acad Sci USA* 2005;102:14753-14758) o la inmunoprecipitación de ADN metilado (MeDIP) (Papageorgiou EA y col. *Nat Med* 2011; 17: 510-513).

Mouliere et al., "High fragmentation characterizes tumour-derived circulating DNA", *PLOS One*, vol. 6, n.º 9, 6 de septiembre de 2011 desvela un estudio en el que del tamaño del ADN en circulación (ADNct) medida usando el análisis de Q-PCR de muestras de plasma de pacientes de cancer mostró que el ADNct presenta un perfil de cantidad específica basándose en el tamaño del ADNct y una fragmentación de ADNct significativamente superior.

De acuerdo con la presente invención, se proporciona un procedimiento de análisis de una muestra biológica de un sujeto, incluyendo la muestra biológica ADN que se origina a partir de células normales y posiblemente de células asociadas con cáncer, en donde al menos algo del ADN está libre de células en la muestra biológica,

comprendiendo el procedimiento: (a) secuenciar una pluralidad de fragmentos de ADN para obtener una pluralidad de lecturas de secuencia que comprenden los nucleótidos más externos en cada extremo de una pluralidad de fragmentos de ADN; (b) alinear las lecturas de secuencia con un genoma de referencia, obteniendo de este modo un conjunto de coordenadas genómicas que incluyen coordenadas genómicas de los nucleótidos más externos que definen un tamaño de un fragmento de ADN para cada uno de la pluralidad de fragmentos de ADN; (c) calcular un valor de un parámetro basándose en cantidades de lecturas de secuencia de fragmentos de ADN que se alinean con el conjunto de coordenadas genómicas de (b) en múltiples tamaños; (d) comparar el valor con un valor de referencia; y (e) determinar una clasificación de un nivel de cáncer en el sujeto basándose en la comparación del valor con el valor de referencia.

Otras realizaciones se dirigen a sistemas, dispositivos portátiles de consumo, y medios de lectura por ordenador asociados con los procedimientos descritos en el presente documento.

Se puede tener una mejor comprensión de la naturaleza y las ventajas de la presente invención con referencia a la siguiente descripción detallada y los dibujos que acompañan.

La figura 1 muestra una gráfica 100 de una distribución de tamaños de ADN sin células en la circulación en plasma materno de acuerdo con las realizaciones de la presente invención.

La figura 2A muestra una gráfica 200 de distribuciones de tamaños de ADN fetal en dos muestras de plasma materno (embarazos de 1^{er} trimestre) con diferentes concentraciones fraccionarias de ADN fetal de acuerdo con las realizaciones de la presente invención.

La figura 2B muestra una gráfica 250 de distribuciones de tamaños de fragmentos de ADN en dos muestras de plasma materno (embarazos de 2^o trimestre) con diferentes concentraciones fraccionarias de ADN fetal de acuerdo con las realizaciones de la presente invención.

La figura 3 es un diagrama de flujo de un procedimiento 300 que ilustra un procedimiento de cálculo de una concentración fraccionaria de ADN clínicamente relevante en una muestra biológica de acuerdo con las realizaciones de la presente invención.

La figura 4 es una gráfica 400 que muestra una distribución de tamaños (electroferograma) de ADN plasmático materno obtenido usando electroforesis de acuerdo con las realizaciones de la presente invención.

La figura 5A es una gráfica 500 que muestra una proporción de fragmentos de ADN que son de 150 pb o inferior para muestras que tienen diferentes porcentajes de ADN fetal en plasma materno de acuerdo con las realizaciones de la presente invención.

La figura 5B es una gráfica 550 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(CF(\text{tamaño} \leq 150) / \text{tamaño}(163-169))$.

La figura 6A es una gráfica 600 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-146) / \text{tamaño}(163-169))$.

La figura 6B es una gráfica 650 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-154) / \text{tamaño}(163-169))$.

La figura 7 es una gráfica 700 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(100-150) / \text{tamaño}(163-169))$.

La figura 8 es una gráfica 800 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para muestras que tienen diferentes porcentajes de ADN fetal en el plasma materno de acuerdo con las realizaciones de la presente invención.

La figura 9A es una gráfica 900 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(CF(\text{tamaño} \leq 150) / \text{tamaño}(163-169))$.

La figura 9B es una gráfica 950 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-146) / \text{tamaño}(163-169))$.

La figura 10A es una gráfica 1000 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-154) / \text{tamaño}(163-169))$.

La figura 10B es una gráfica 1005 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(100-150) / \text{tamaño}(163-169))$.

La figura 11 es una gráfica que muestra una relación de tamaños representada frente al porcentaje de ADN fetal para el tamaño de elementos de repetición de acuerdo con las realizaciones de la presente invención.

La figura 12A es un electroferograma 1200 que se puede usar para determinar una relación de tamaños de acuerdo con las realizaciones de la presente invención.

La figura 12B es una gráfica 1250 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 200 pb a 267 pb y ADN de 290 pb a 294 pb para muestras que tienen diferentes porcentajes de ADN fetal en el plasma materno de acuerdo con las realizaciones de la presente invención.

La figura 13 es un diagrama de flujo de un procedimiento 1300 para determinar puntos de datos de calibración para mediciones hechas a partir de muestras de calibración de acuerdo con las realizaciones de la presente invención.

La figura 14A es una gráfica 1400 de una relación de tamaños frente a la concentración fraccionaria de ADN fetal para el grupo de entrenamiento de acuerdo con las realizaciones de la presente invención.

La figura 14B es una gráfica 1450 de concentraciones fraccionadas deducidas (calculadas) a partir de la función

lineal 1410 de la figura 14A frente a las concentraciones fraccionarias medias usando secuencias fetales específicas de acuerdo con las realizaciones de la presente invención.

La figura 15A es una gráfica 1500 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para muestras que tienen diferentes porcentajes de ADN tumoral en el plasma de dos pacientes de carcinoma hepatocelular (HCC) antes y después de resección tumoral de acuerdo con las realizaciones de la presente invención.

La figura 15B es una gráfica 1550 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(CF(\text{tamaño} \leq 150)/\text{tamaño}(163-169))$, para dos pacientes de HCC antes y después de resección tumoral.

La figura 16A es una gráfica 1600 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-146)/\text{tamaño}(163-169))$, para dos pacientes de HCC antes y después de resección tumoral.

La figura 16B es una gráfica 1650 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-154)/\text{tamaño}(163-169))$, para dos pacientes de HCC antes y después de resección tumoral.

La figura 17 es una gráfica 1700 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(100-150)/\text{tamaño}(163-169))$, para dos pacientes de HCC antes y después de resección tumoral.

La figura 18A es una gráfica 1800 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para pacientes de HCC antes y después de resección tumoral.

La figura 18B es una gráfica 1850 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(CF(\text{tamaño} \leq 150)/\text{tamaño}(163-169))$, para pacientes de HCC antes y después de resección tumoral.

La figura 19A es una gráfica 1900 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-146)/\text{tamaño}(163-169))$, para pacientes de HCC antes y después de resección tumoral.

La figura 19B es una gráfica 1950 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-154)/\text{tamaño}(163-169))$, para pacientes de HCC antes y después de resección tumoral.

La figura 20 es una gráfica 2000 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(100-150)/\text{tamaño}(163-169))$, para pacientes de HCC antes y después de resección tumoral.

La figura 21 es un diagrama de flujo que ilustra un procedimiento 2100 para analizar una muestra biológica de un organismo para determinar una clasificación de un nivel de cáncer de acuerdo con las realizaciones de la presente invención.

La figura 22 es una tabla 2200 que muestra algunas aberraciones cromosómicas comunes vistas en diferentes tipos de cáncer.

La figura 23 muestra un diagrama de bloques de un sistema informático de ejemplo 2300 que se puede usar con el sistema y procedimientos de acuerdo con las realizaciones de la presente invención.

La expresión "muestra biológica" como se usa en el presente documento, se refiere a cualquier muestra que se extrae de un paciente (p. ej., un ser humano, tal como una mujer embarazada) y contiene una o más moléculas de ácido nucleico de interés. Los ejemplos incluyen muestras de plasma, saliva, líquido pleural, sudor, líquido ascítico, bilis, orina, suero, jugo pancreático, heces y frotis de cuello uterino. La muestra biológica se puede obtener de un ser humano, un animal, u otro organismo adecuado. Una "muestra de calibración" corresponde a una muestra biológica cuya fracción de ADN clínicamente relevante se conoce o se determina por un procedimiento de calibración, p. ej., usando un alelo específico para el ADN clínicamente relevante. Los ejemplos de ADN clínicamente relevante son ADN fetal en el plasma materno o ADN tumoral en el plasma de un paciente.

Como se usa en el presente documento, el término "locus" o su forma plural "locus" es una posición o situación de cualquier longitud de nucleótidos (o pares de bases) que tiene una variación a lo largo de los genomas. La expresión "lectura de secuencia" se refiere a una secuencia obtenida de toda o parte de una molécula de ácido nucleico, p. ej., un fragmento de ADN. En una realización, se secuencia solo un extremo del fragmento. Alternativamente, se pueden secuenciar ambos extremos (p. ej., aproximadamente 30 pb desde cada extremo) del fragmento, para generar dos lecturas de secuencias. Las lecturas de secuencias emparejadas después se pueden alinear con un genoma de referencia, que puede proporcionar una longitud del fragmento. En otra realización más, un fragmento de ADN se puede hacer circular, p. ej., mediante ligado, y se puede secuenciar la parte que abarca el sitio de ligado.

La expresión "secuenciación universal" se refiere a la secuenciación donde se añaden adaptadores al extremo de un fragmento y los cebadores para la secuenciación se unen a los adaptadores. Por lo tanto, se puede secuenciar cualquier fragmento con el mismo cebador y por lo tanto la secuenciación puede ser aleatoria.

La expresión concentración fraccionaria de ADN fetal se usa de forma intercambiable con las expresiones proporción de ADN fetal y fracción de ADN fetal, y se refiere a la proporción de las moléculas de ADN fetal que están presentes en una muestra biológica (p. ej., plasma materno o muestra de suero) que deriva del feto (Lo YMD y col. *Am J Hum Genet* 1998; 62:768-775; Lun FMF y col. *Clin Chem* 2008; 54:1664-1672). De forma similar, las

expresiones concentración fraccionaria de ADN tumoral se puede usar de forma intercambiable con las expresiones proporción de ADN tumoral y fracción de ADN tumoral, y se refieren a la proporción de moléculas de ADN tumorales que están presentes en una muestra biológica.

- 5 La expresión “perfil de tamaños” en general se refiere a los tamaños de los fragmentos de ADN en una muestra biológica. Un perfil de tamaños puede ser un histograma que proporciona una distribución de una cantidad de fragmentos de ADN de una variedad de tamaños. Se pueden usar diferentes parámetros estadísticos (también denominados parámetros de tamaño o simplemente parámetro) para distinguir un perfil de tamaños de otro. Un parámetro es el porcentaje de fragmentos de ADN de un tamaño particular o intervalo de tamaños con respecto a todos los fragmentos de ADN o con respecto a fragmentos de ADN de otro tamaño o intervalo.

15 Los ejemplos de ADN “clínicamente relevante” incluyen ADN fetal en el plasma materno y ADN tumoral en el plasma de un paciente. Otro ejemplo incluye la medición de la cantidad de ADN asociado a injerto en el plasma de un paciente de trasplante. Un ejemplo más incluye la medición de las cantidades relativas de ADN hematopoyético o no hematopoyético en el plasma de un sujeto. Esta última realización se puede usar para detectar o seguir o pronosticar procesos patológicos o lesiones que implican tejidos hematopoyéticos y/o no hematopoyéticos.

20 Un “punto de datos de calibración” incluye un “valor de calibración” y una concentración fraccionaria medida o conocida de ADN de interés (es decir, el ADN clínicamente relevante). El valor de calibración es un valor de un parámetro de tamaño determinado para una muestra de calibración, para la que se conoce la concentración fraccionaria del ADN clínicamente relevante. Los puntos de datos de calibración se pueden definir de una variedad de formas, p. ej., como puntos discretos o como una función de calibración (conocida también como curva de calibración o superficie de calibración).

25 La expresión “nivel de cáncer” se puede referir a si existe el cáncer, una etapa de un cáncer, un tamaño de tumor, cuántas eliminaciones o amplificaciones de una región cromosómica están implicadas (p. ej., duplicadas o triplicadas), y/o la medida de una gravedad de un cáncer. El nivel de cáncer podría ser un número u otros caracteres. El nivel podría ser cero. El nivel de cáncer incluye también afecciones premalignas o precancerosas asociadas con eliminaciones o amplificaciones.

30 Se sabe que las moléculas de ADN fetal sin células en el plasma materno en general son más cortas que las derivadas de la madre (Chan KCA y col. *Clin Chem* 2004; 50:88-92; Lo YMD y col. *Sci Transl Med* 2010; 2:61ra91). La presencia de ADN fetal da como resultado un desplazamiento en la distribución de tamaños global del ADN del plasma materno y el grado de desplazamiento está asociado con la concentración fraccionaria de ADN fetal. Midiendo valores particulares del perfil de tamaños de ADN del plasma materno, se pueden obtener realizaciones de concentración fraccionaria de ADN fetal en el plasma materno.

40 Aparte de las aplicaciones en diagnóstico prenatal no invasivo, se pueden usar también realizaciones para medir la concentración fraccionaria de especies de ácido nucleico clínicamente útiles de diferentes tamaños en fluidos biológicos, que pueden ser útiles para la detección del cáncer, trasplante y seguimiento médico. Se ha mostrado previamente que el ADN derivado de tumor es más corto que el ADN no derivado de cáncer en el plasma de un paciente con cáncer (Diehl F y col. *Proc Natl Acad Sci USA* 2005; 102:16368-16373). En el contexto de los trasplantes, se ha mostrado que el ADN hematopoyético es más corto que el ADN no hematopoyético (Zheng YW y col. *Clin Chem* 2012; 58:549-558). Por ejemplo, si un paciente recibe un hígado de un donante, entonces el ADN derivado del hígado (un órgano no hematopoyético en el adulto) será más corto que el ADN hematopoyético en el plasma (Zheng YW y col. *Clin Chem* 2012; 58:549-558). De forma similar, en un paciente con infarto de miocardio o accidente cerebrovascular, el ADN liberado por los órganos no hematopoyéticos dañados (es decir, el corazón y cerebro, respectivamente) se esperaría que produjera un desplazamiento en el perfil de tamaños del ADN plasmático hacia un espectro más corto.

50 I. DISTRIBUCIÓN DE TAMAÑOS

Para demostrar las realizaciones, en los siguientes ejemplos se muestra que se puede medir el perfil de tamaños, por ejemplo, mediante secuenciación masiva en paralelo de extremos emparejados o por electroforesis (p. ej., usando un Bioanalyzer). Este último ejemplo es particularmente útil porque la electroforesis usando un Bioanalyzer es un procedimiento rápido y relativamente barato. Esto permitiría llevar a cabo rápidamente este análisis como una medida de control de calidad antes de someter una muestra de ADN plasmático al procedimiento de secuenciación relativamente caro.

60 La figura 1 muestra una gráfica 100 de una distribución de tamaños de ADN sin células en la circulación en plasma materno de acuerdo con las realizaciones de la presente invención. Una distribución de tamaños se puede obtener midiendo un tamaño de fragmentos de ADN y después contando el número de fragmentos de ADN de varios tamaños, p. ej., en el intervalo de 50 bases a aproximadamente 220 bases. La gráfica 100 muestra dos distribuciones. La distribución 110 es para todos los fragmentos de ADN en la muestra de plasma materno, y la distribución 120 es solo para el ADN que procede del feto. El eje horizontal es el tamaño en pares de bases (pb) de los fragmentos de ADN. El eje vertical es el porcentaje de los fragmentos de ADN medidos.

En la figura 1, la distribución de tamaños del ADN procedente del feto en el plasma materno ha mostrado ser más corta que la del procedente de la madre (Chan KC y col. *Clin Chem* 2004; 50:88-92). Recientemente, los autores de la invención han usado el análisis de secuenciación masiva en paralelo de extremos emparejados para determinar la distribución de tamaños de alta resolución del ADN específico fetal y el ADN total (derivado principalmente de la madre) en una mujer embarazada. Han mostrado que una diferencia principal entre las dos especies de ADN es que hay una reducción en la fracción de los fragmentos de ADN de 166 pb y una mayor proporción del ADN más corto inferior a 150 pb del ADN procedente del feto (Lo YM y col. *Sci Transl Med* 2010 2:61ra91).

En el presente documento, los autores de la invención describen cómo sería útil un análisis de una distribución de tamaños de los fragmentos de ADN totales en una muestra de plasma materno (un ejemplo de una muestra biológica) para determinar la concentración fraccionaria de ADN fetal en plasma materno. La mayor concentración fraccionaria de ADN fetal en el plasma materno daría como resultado el acortamiento de la distribución de tamaños general del ADN total. En una realización, la abundancia relativa (un ejemplo de un parámetro) de los fragmentos de ADN de aproximadamente 144 pb y los fragmentos de ADN de aproximadamente 166 pb se podría usar para reflejar la concentración fraccionaria de ADN fetal. En otra realización, se pueden usar otros parámetros o combinación de parámetros relacionados con un perfil de tamaños para reflejar la distribución de tamaños del ADN plasmático.

La figura 2A muestra una gráfica 200 de distribuciones de tamaños de ADN fetal en dos muestras de plasma materno (embarazos del 1^{er} trimestre) con diferentes concentraciones fraccionarias de ADN fetal de acuerdo con las realizaciones de la presente invención. Las dos mujeres embarazadas llevaban fetos de varones. Las concentraciones fraccionarias de ADN fetal se determinaron a partir de la proporción de secuencias del cromosoma Y entre los fragmentos de ADN secuenciado total. Ambas muestras se tomaron de la mujer embarazada durante el primer trimestre de sus embarazos. El caso 338 (línea negra, concentración fraccionaria de ADN fetal 10%) tenía una concentración fraccionaria de ADN fetal inferior que el caso 263 (línea de puntos, concentración fraccionaria de ADN fetal 20%). Cuando se compara con el caso 263, el caso 338 tiene un pico mayor a 166 pb, pero los picos para tamaños inferiores a 150 pb eran inferiores. En otras palabras, los fragmentos de ADN más cortos de 150 pb eran más abundantes en el caso 263, mientras que los fragmentos de aproximadamente 166 pb eran más abundantes en el caso 338. Estas observaciones están de acuerdo con la hipótesis de que las cantidades relativas de ADN corto y largo se pueden correlacionar con la concentración fraccionaria de ADN fetal.

La figura 2B muestra una gráfica 250 de distribuciones de tamaños de fragmentos de ADN en dos muestras de plasma materno (embarazos de 2^o trimestre) con diferentes concentraciones fraccionarias de ADN fetal de acuerdo con las realizaciones de la presente invención. Ambas muestras se tomaron de mujeres embarazadas durante el segundo trimestre. Ambas mujeres embarazadas llevaban fetos de varones. Las concentraciones fraccionarias de ADN fetal se determinaron a partir de la proporción de secuencias del cromosoma Y entre los fragmentos de ADN secuenciado total. De forma similar al ejemplo previo, el caso 5415 (línea de puntos, con mayor concentración fraccionaria de ADN fetal de 19%) tenía picos mayores para los tamaños inferiores a 150 pb mientras que el caso 5166 (línea negra, con menor concentración fraccionaria de ADN fetal de 12%) tenía un pico superior a 166 pb.

La correlación de diferentes valores de un parámetro de tamaño con valores de concentración fraccionaria de ADN fetal se muestra en gráficas de datos más adelante. Además, el tamaño de los fragmentos de ADN tumoral está correlacionado con el porcentaje de fragmentos de ADN tumoral en una muestra con fragmentos de ADN tumoral y fragmentos de ADN de células normales. Por lo tanto, también se puede usar el tamaño de los fragmentos tumorales para determinar el porcentaje de fragmentos tumorales en la muestra.

II. PROCEDIMIENTO

Puesto que el tamaño de los fragmentos de ADN se correlaciona con una concentración fraccionaria (también denominada un porcentaje), las realizaciones pueden usar esta correlación para determinar una concentración fraccionaria de un tipo particular de ADN (p. ej., ADN fetal o ADN de un tumor) en una muestra. El tipo particular de ADN es clínicamente relevante ya que es la concentración fraccionaria que se va a calcular. Por consiguiente, un procedimiento puede calcular una concentración fraccionaria de ADN clínicamente relevante en una muestra biológica basado en un tamaño medido de los fragmentos de ADN.

La figura 3 es un diagrama de flujo de un procedimiento 300 que ilustra un procedimiento de cálculo de una concentración fraccionaria de ADN clínicamente relevante en una muestra biológica de acuerdo con las realizaciones de la presente invención. La muestra biológica incluye el ADN clínicamente relevante y otro ADN. La muestra biológica se puede obtener de un paciente, p. ej., una mujer embarazada con un feto. En otra realización, el paciente puede tener o sospechar que tiene un tumor. En una implementación, la muestra biológica se puede recibir en una máquina, p. ej., una máquina secuenciadora, que produce datos de medición (p. ej., lecturas de secuencia) que se pueden usar para determinar tamaños de fragmentos de ADN. El procedimiento 300 se puede llevar a cabo total o parcialmente con un sistema informático, así como con otros procedimientos descritos en el presente documento.

En el bloque 310, se miden las cantidades de fragmentos de ADN que corresponden a diferentes tamaños. Para

cada tamaño de una pluralidad de tamaños, se puede medir una cantidad de una pluralidad de fragmentos de ADN de la muestra biológica que corresponden al tamaño. Por ejemplo, se puede medir el número de fragmentos de ADN que tienen una longitud de 140 pares de bases. Las cantidades se pueden guardar como un histograma. En una realización, se mide un tamaño de cada una de la pluralidad de ácidos nucleicos de la muestra biológica, que se puede hacer de forma individual (p. ej., secuenciación de una sola molécula) o como grupo (p. ej., por electroforesis). Los tamaños pueden corresponder a un intervalo. Por lo tanto, una cantidad puede ser para fragmentos de ADN que tienen un tamaño con un intervalo particular.

La pluralidad de fragmentos de ADN se puede seleccionar de forma aleatoria o preferiblemente se selecciona de una o más regiones predeterminadas de un genoma. Por ejemplo, se puede llevar a cabo enriquecimiento dirigido, como se ha descrito antes. En otra realización, los fragmentos de ADN se pueden secuenciar aleatoriamente (p. ej., usando secuenciación universal) y las lecturas de secuencias resultantes se pueden alinear con un genoma que corresponde al sujeto (p. ej., un genoma humano de referencia). Después, se pueden usar solo los fragmentos de ADN cuyas lecturas de secuencia se alinean con la una o más regiones predeterminadas para determinar el tamaño.

En varias realizaciones, el tamaño puede ser masa, longitud u otras mediciones adecuadas. La medición se puede llevar a cabo de diferentes formas, como se describe en el presente documento. Por ejemplo, se puede llevar a cabo secuenciación de extremos emparejados y alineamiento de los fragmentos de ADN, o se puede usar electroforesis. Se puede medir un número estadísticamente significativo de fragmentos de ADN para proporcionar un perfil de tamaños preciso de la muestra biológica. Los ejemplos de un número estadísticamente significativo de fragmentos de ADN incluyen más de 10.000; 1.000.000; 2.000.000, u otros valores adecuados, lo cual puede depender de la precisión requerida.

En una realización, los datos obtenidos de una medición física, tal como la secuenciación de extremos emparejados o electroforesis, se pueden recibir en un ordenador y analizar para llevar a cabo la medición de los tamaños de los fragmentos de ADN. Por ejemplo, las lecturas de secuencias de la secuenciación de extremos emparejados se pueden analizar (p. ej., por alineamiento) para determinar los tamaños. Como otro ejemplo, se puede analizar el electroferograma que resulta de la electroforesis para determinar los tamaños. En una implementación, el análisis de los fragmentos de ADN no incluye el procedimiento real de secuenciación o sometimiento de los fragmentos de ADN a electroforesis, mientras que otras implementaciones pueden llevar a cabo solo un análisis de los datos resultantes.

En el bloque 320, se calcula un primer valor de un primer parámetro basado en las cantidades de fragmentos de ADN en múltiples tamaños. En un aspecto, el primer parámetro proporciona una medición estadística de un perfil de tamaños (p. ej., un histograma) de fragmentos de ADN en la muestra biológica. El parámetro se puede denominar un parámetro de tamaño puesto que se determina a partir de los tamaños de la pluralidad de fragmentos de ADN.

El primer parámetro puede ser de diferentes formas. Dicho parámetro es una serie de fragmentos de ADN en un tamaño particular dividido entre el número total de fragmentos, que se puede obtener de un histograma (cualquier estructura de datos que proporcione recuentos absolutos o relativos de fragmentos de tamaños particulares). Como otro ejemplo, un parámetro podría ser un número de fragmentos de un tamaño particular o dentro de un intervalo particular, dividido entre un número de fragmentos de otro tamaño o intervalo. La división puede actuar como una normalización para explicar un número diferente de fragmentos de ADN que se están analizando para diferentes muestras. Se puede llevar a cabo una normalización analizando un mismo número de fragmentos de ADN para cada muestra, lo cual proporciona efectivamente un mismo resultado que dividiendo entre un número total de fragmentos analizados. Se describen otros ejemplos de parámetros en el presente documento.

En el bloque 330, se obtienen uno o más puntos de datos de calibración. Cada primer punto de datos de calibración puede especificar una concentración fraccionaria de ADN clínicamente relevante que corresponde a un valor particular (un valor de calibración) del primer parámetro. La concentración fraccionaria se puede especificar como una concentración particular o un intervalo de concentraciones. Un valor de calibración puede corresponder a un valor del primer parámetro (es decir, un parámetro de tamaño particular) determinado a partir de una pluralidad de muestras de calibración. Los puntos de datos de calibración se pueden determinar a partir de muestras de calibración con concentraciones fraccionarias conocidas, que se pueden medir mediante diferentes técnicas descritas en el presente documento. Al menos algunas de las muestras de calibración tendrían una concentración fraccionaria diferente, pero algunas muestras de calibración pueden tener una misma concentración fraccionaria.

En diferentes realizaciones, uno o más puntos de calibración se pueden definir como un punto discreto, un conjunto de puntos discretos, como una función, como un punto discreto y una función, o cualquier otra combinación de conjuntos discretos o continuos de valores. Como un ejemplo, un punto de datos de calibración se podría determinar a partir de un valor de calibración de un parámetro de tamaño (p. ej., número de fragmentos en un tamaño particular o intervalo de tamaños) para una muestra con una concentración fraccionaria particular. Se puede usar una pluralidad de histogramas, con un histograma diferente para cada muestra de calibración, donde algunas de las muestras de calibración pueden tener la misma concentración fraccionaria.

En una realización, los valores medidos de un mismo parámetro de tamaño de múltiples muestras de la misma concentración fraccionaria se podrían combinar para determinar un punto de datos de calibración para una

concentración fraccionaria particular. Por ejemplo, se puede obtener una media de los valores del parámetro de tamaño a partir de los datos de tamaño de muestras de la misma concentración fraccionaria, para determinar un punto de datos de calibración particular (o proporcionar un intervalo que corresponde al punto de datos de calibración). En otra realización, se pueden usar múltiples puntos de datos con el mismo valor de calibración para determinar una concentración fraccionaria media.

En una implementación, los tamaños de los fragmentos de ADN se miden para muchas muestras de calibración. Se determina un valor de calibración del mismo parámetro de tamaño para cada muestra de calibración, donde el parámetro de tamaño se puede representar gráficamente frente a la concentración fraccionaria conocida de la muestra. Después se puede ajustar una función a los puntos de datos de la gráfica, donde el ajuste de la función define los puntos de datos de calibración que se van a usar para determinar la concentración fraccionaria para una nueva muestra.

En el bloque 340, el primer valor se compara con un valor de calibración de al menos un punto de datos de calibración. La comparación se puede hacer en una variedad de formas. Por ejemplo, la comparación puede ser sea el primer valor superior o inferior al valor de calibración. La comparación puede implicar comparar una curva de calibración (compuesta de los puntos de datos de calibración) y, por lo tanto, la comparación puede identificar el punto sobre la curva que tiene el primer valor del primer parámetro. Por ejemplo, un valor X calculado del primer parámetro (determinado a partir de los tamaños medidos de ADN en la nueva muestra) se puede usar como entrada en una función $F(X)$, donde F es la función de calibración (curva). El resultado de $F(X)$ es la concentración fraccionaria. Se puede proporcionar un intervalo de error, que puede ser diferente para cada valor X , proporcionando así un intervalo de valores como un resultado de $F(X)$.

En la etapa 350, la concentración fraccionaria del ADN clínicamente relevante en la muestra biológica se calcula basada en la comparación. En una realización, se puede determinar si el primer valor del primer parámetro está por encima o por debajo de un valor de calibración umbral, y así determinar si la concentración fraccionaria calculada de la presente muestra está por encima o por debajo de la concentración fraccionaria correspondiente al valor de calibración umbral. Por ejemplo, si el primer valor calculado X_1 para la muestra biológica está por encima de un valor de calibración X_c , entonces la concentración fraccionaria FC_1 de la muestra biológica se puede determinar como que está por encima de la concentración fraccionaria F_c que corresponde a X_c . Esta comparación se puede usar para determinar si existe una concentración fraccionaria suficiente en la muestra biológica para realizar otros ensayos, p. ej., ensayo de una aneuploidía fetal. Esta relación de por encima y por debajo puede depender de como se define el parámetro. En dicha realización, puede ser necesario solo un punto de dato de calibración.

En otra realización, la comparación se lleva a cabo introduciendo el primer valor en la función de calibración. La función de calibración puede comparar efectivamente el primer valor con los valores de calibración identificando el punto sobre una curva que corresponde al primer valor. Después la concentración fraccionaria calculada se proporciona como el valor de salida de la función de calibración.

En una realización, se puede determinar el valor de más de un parámetro para la muestra biológica. Por ejemplo, se puede determinar un segundo valor para un segundo parámetro, que corresponde a una medición estadística diferente del perfil de tamaños de fragmentos de ADN en la muestra biológica. El segundo valor se puede determinar usando las mismas mediciones de tamaños de los fragmentos de ADN o diferentes mediciones de tamaños. Cada parámetro puede corresponder con una curva de calibración diferente. En una implementación, se pueden comparar independientemente los diferentes valores con las diferentes curvas de calibración para obtener una pluralidad de concentraciones fraccionarias calculadas, que después se pueden promediar o usar para proporcionar un intervalo como una salida.

En otra implementación, se puede usar una curva de calibración multidimensional, donde los diferentes valores de los parámetros se pueden introducir efectivamente en una única función de calibración que da como resultado la concentración fraccionaria. La función de calibración única puede ser resultado de un ajuste funcional de todos los puntos de datos obtenidos de las muestras de calibración. Por lo tanto, en una realización, los primeros puntos de datos de calibración y los segundos puntos de datos de calibración pueden ser puntos sobre una curva multidimensional, donde la comparación incluye identificar el punto multidimensional que tiene coordenadas que corresponden al primer valor y al uno o más segundos valores.

III. DETERMINACIÓN DE TAMAÑOS

La distribución de tamaños del ADN plasmático se puede determinar, por ejemplo, pero no limitado al uso de la PCR en tiempo real, electroforesis y análisis de espectrometría de masas. En diferentes realizaciones, el tamaño medido es una longitud, una masa molecular o un parámetro medido que es proporcional a la longitud o masa, tal como la movilidad en un electroforetograma y el tiempo necesario para recorrer una distancia fija en la electroforesis o espectrómetro de masas. En otro ejemplo, se puede teñir el ADN con un colorante fluorescente de intercalad, p. ej., bromuro de etidio o SYBR Green, donde la cantidad de colorante unido será proporcional a la longitud de la molécula de ADN. Se puede determinar la cantidad de colorante unido por la intensidad de la fluorescencia emitida, cuando se emite luz UV sobre la muestra. Algunas muestras para medir el tamaño y los datos resultantes se

describen más adelante.

A. Primer conjunto de muestras fetales usando secuenciación

5 La tabla 1 muestra información de la muestra y análisis de secuenciación para un ejemplo que implica una fracción de ADN fetal. Se recogieron muestras de plasma de 80 mujeres embarazadas que llevan cada una un solo feto de varón. Entre estas 80 mujeres embarazadas, 39 llevaban un feto euploide, 18 llevaban un feto con trisomía 21 (T21), 10 llevaban un feto con trisomía 18 (T18), y 13 llevaban un feto con trisomía 13 (T13). Se determinó una distribución de tamaños del ADN plasmático usando secuenciación en paralelo masiva de extremos emparejados. Se
10 construyeron bibliotecas de secuenciación de ADN plasmático materno como se ha descrito previamente (Lo YM y col. *Sci Transl Med* 2010; 2:61ra91), excepto que se introdujo un código de barras de 6 bases en las moléculas de ADN de cada muestra plasmática por una amplificación de PCR con cebadores triples.

15 Se introdujeron dos muestras en una banda de secuenciación (es decir, secuenciación 2-plex). En otras realizaciones, se pueden introducir más de dos muestras en una banda de secuenciación, p. ej., 6 o 12 o 20 o más de 20. Todas las bibliotecas se secuenciaron mediante un Genome Analyzer IIx (Illumina) usando el formato 36-bp × 2 PE. Se llevaron a cabo 7 ciclos adicionales de secuenciación para descodificar la secuencia índice en cada molécula de ADN plasmático secuenciada. Las lecturas de secuencias de 36 pb se alinearon con el genoma de referencia humano sin repeticiones enmascaradas (Hg18) (genome.ucsc.edu), usando el programa Short
20 Oligonucleotide Alignment Program 2 (SOAP2) (soap.genomics.org.cn). Se identificaron las lecturas de extremos emparejados (PE) con miembros individuales secuenciados en la misma posición de conjunto en la celda de flujo y alineados únicamente en un sitio único en el genoma humano con la orientación correcta y sin ningún emparejamiento erróneo de nucleótidos. En otras realizaciones, el alineamiento puede no ser único y pueden estar permitidos los emparejamientos erróneos.

25 Solo se recuperaron para el análisis las lecturas de PE que demostraron un tamaño de inserto ≤ 600 pb. Con estos criterios, el tamaño de los fragmentos de ADN plasmático analizados en estos experimentos estaban en el intervalo de 36 pb a 600 pb. El tamaño de cada fragmento de ADN secuenciado se infirió de las coordenadas de los nucleótidos más externos en cada extremo de los fragmentos secuenciados.
30

La tabla 1 muestra datos para muestras de diferentes estados de aneuploidía. Estos datos incluyen el número de casos, mediana e intervalo de la edad gestacional, junto con el número de la mediana e intervalo de lecturas de extremos emparejados y la fracción de ADN fetal.

Tipo de caso	Nº de casos	Edad gestacional (semanas) mediana (intervalo)	Nº de lecturas de PE (millones) mediana (intervalo)	Fracción de ADN fetal (%) mediana (intervalo)
Euploide	39	13,2 (11,3-15,1)	4,7 (1,8-12,0)	15,7 (5,9-25,7)
T21	18	13,0 (12,1-17,9)	5,2 (2,5-8,9)	13,8 (7,4-27,2)
T18	10	13,3 (12,1-14,2)	4,9 (3,6 - 6,2)	7,2 (4,8-16,7)
T13	13	12,4 (11,5-16,4)	5,3 (2,7-7,7)	7,5 (3,2-14,1)
Todos	80	13,1 (11,3-17,9)	4,9 (1,8-12,0)	13,7 (3,2-27,2)

35 Las concentraciones fraccionarias de ADN fetal en las muestras de plasma materno se dedujeron de la cantidad de secuencias que se alineaban con el cromosoma Y como se ha descrito previamente (Chiu RW y col. *BMJ* 2011;342:c7401). Esta técnica es un ejemplo de un procedimiento de calibración. Por lo tanto, la fracción de ADN fetal medida en la tabla 1 se puede usar en los puntos de datos de calibración para calcular una fracción de ADN fetal en una nueva muestra. Las muestras usadas para recoger los datos en la tabla 1 se pueden considerar
40 muestras de calibración.

B. Segundo conjunto de muestras fetales usando secuenciación dirigida

45 La tabla 2 muestra información de muestras y enriquecimiento dirigido del ADN del plasma materno de acuerdo con las realizaciones de la presente invención. Se recogieron muestras de plasma de 48 mujeres embarazadas, cada una con un solo feto. Entre estas 48 mujeres embarazadas, 21 llevaban un feto euploide, 17 llevaban un feto con trisomía 21 (T21), 9 llevaban un feto con trisomía 18 (T18), y 1 llevaba un feto con trisomía 13 (T13). Estos datos, junto con los siguientes ejemplos, ilustran que las realizaciones pueden usar técnicas dirigidas. La distribución de tamaños del ADN plasmático se puede determinar usando la secuenciación masiva en paralelo de extremos emparejados. En otras realizaciones, la distribución de tamaños del ADN plasmático se puede determinar, por
50 ejemplo, pero no limitado al uso de la PCR en tiempo real, electroforesis y análisis por espectrometría de masas.

55 Para obtener el cubrimiento de la secuenciación un número alto de veces de las regiones objetivo, se usó el Agilent SureSelect Target Enrichment System en una realización para diseñar sondas para capturar moléculas de ADN del cr7 (región de 0,9 Mb), cr13 (región de 1,1 Mb), cr18 (región de 1,2 Mb) y cr21 (región de 1,3 Mb). En la sonda, los exones en los cr7, cr13, cr18, y la región crítica del síndrome de Down en el cr21 (21q22.1-q22.3) se seleccionaron primero como regiones objetivo. Debido a que el cr13, cr18 y cr21 tienen menos regiones no exones que el cr7, se

introdujeron regiones no exones adicionales en el cr13, cr18, y la región crítica del síndrome de Down en el cr21 para equilibrar la longitud total de la región objetivo entre los cuatro cromosomas anteriores. Las regiones no exones seleccionadas eran de 120 pb de longitud, cartografiadas de forma única, con el contenido de GC cercano a 0,5 y distribuido uniformemente a lo largo de los cromosomas objetivo.

5 Las coordenadas de todas las regiones de exones y no exones anteriores se enviaron a la plataforma de Agilent eArray para el diseño de la sonda. Se incubaron 500 ng de cada biblioteca de ADN de plasma materno con las sondas de captura durante 24 h a 65°C. Después de hibridación, las moléculas de ADN objetivo se eluyeron y se amplificaron por una PCR de 12 ciclos de acuerdo con las instrucciones del fabricante. Las bibliotecas con
10 enriquecimiento del objetivo se indexaron y secuenciaron en un GA IIx (Illumina) usando el formato de 50 pb X 2 PE. Se llevaron a cabo 7 ciclos de secuenciación adicionales para descodificar la secuencia índice en cada molécula de ADN plasmático secuenciada. Las lecturas de secuencias de 50 pb se alinearon con el genoma de referencia humano sin repeticiones enmascaradas (Hg18) (genome.ucsc.edu), usando el programa Short Oligonucleotide Alignment Program 2 (SOAP2) (soap.genomics.org.cn). Se secuenciaron las lecturas de PE con miembros
15 individuales en la misma posición de conjunto en la celda de flujo y se alinearon de forma única con un sitio único en el genoma humano con la orientación correcta. Se permitieron dos emparejamientos erróneos; la complejidad de la biblioteca de secuenciación se redujo significativamente después de enriquecimiento del objetivo.

20 Solo se recuperaron para el análisis las lecturas de PE que demostraron un tamaño de inserto ≤ 600 pb. Con estos criterios, el tamaño de los fragmentos de ADN plasmático analizados en el estudio actual estaban en el intervalo de 36 pb a 600 pb. El tamaño de cada fragmento de ADN secuenciado se infirió de las coordenadas de los nucleótidos más externos en cada extremo de los fragmentos secuenciados. Las concentraciones fraccionarias de ADN fetal en las muestras de plasma materno se calcularon a partir de la relación de fragmentos que llevaban los alelos
25 específicos fetales y los alelos compartidos con las respectivas madres.

La tabla 2 muestra datos de la secuenciación dirigida para muestras de diferentes estados de aneuploidía.

Tipo de caso	Nº de casos	Edad gestacional (semanas) mediana (intervalo)	Nº de lecturas de PE (millones) mediana (intervalo)	Fracción de ADN fetal (%) mediana (intervalo)
Euploide	21	13,0 (12,0-13,3)	2,2 (1,7-3,0)	13,5 (8,4-22,0)
T21	17	13,6(12,6-20,9)	2,1 (1,5-2,7)	15,4 (8,7-22,7)
T18	9	12,7(11,9-13,7)	1,9 (1,7-3,1)	10,5 (7,2-16,3)
T13	1	13	1,6	9,2
Todos	48	13,1(11,9-20,9)	2,1 (1,5-3,1)	13,4 (7,2-22,7)

C. Electroforesis de la muestra fetal

30 Además de usar la secuenciación masiva en paralelo, el análisis de la distribución de tamaños de ADN plasmático se puede lograr por electroforesis. La electroforesis mide el tiempo para que un fragmento se mueva a través del medio. Las partículas de diferentes tamaños tardan diferentes tiempos en moverse a través del medio. Por lo tanto, en una realización, se puede llevar a cabo electroforesis microfluidica de la biblioteca de secuenciación del ADN del plasma materno para determinar la distribución de tamaños del ADN del plasma materno.

35 La figura 4 es una gráfica 400 que muestra una distribución de tamaños (electroferograma) del ADN del plasma materno obtenido usando electroforesis de acuerdo con las realizaciones de la presente invención. La electroforesis microfluidica se llevó a cabo usando el Agilent 2100 Bioanalyzer. Los electroferogramas de las bibliotecas de secuenciación de dos muestras se muestran en la gráfica 400. El eje X representa la duración del tiempo que tarda el ADN en alcanzar el sensor y corresponde al tamaño de los fragmentos de ADN. El eje Y representa las unidades de fluorescencia (UF) de los fragmentos de ADN que pasan por el sensor en un tiempo particular.

40 El tiempo que tarda un fragmento de ADN en alcanzar el sensor se correlaciona positivamente con el tamaño del fragmento de ADN. El Bioanalyzer puede convertir automáticamente la duración de tiempo en un tamaño de fragmento comparando el tiempo de recorrido de la muestra de ensayo con los de una mezcla de fragmentos de ADN de longitudes conocidas (es decir, una escala de ADN). Las bibliotecas de secuenciación de ADN posteriormente se secuenciaron usando la secuenciación masiva en paralelo y se usó la fracción de secuencias del cromosoma Y para determinar las concentraciones fraccionarias de ADN fetal de estas muestras.

50 En la gráfica 400, la línea negra 410 representa la muestra UK92797 que tiene una concentración fraccionaria de ADN fetal de 8,3% y la línea de puntos 420 representa la muestra UK94884 que tiene una concentración fraccionaria de ADN fetal de 20,3%. En comparación con la muestra UK92797, la muestra UK94884 (la muestra con mayor fracción de ADN fetal) tenía una cantidad relativamente mayor de ADN en un intervalo de tiempo electroforético de 63 segundos a 73 segundos (región A) que corresponde al tamaño de ADN de 200 pb a 267 pb y una cantidad
55 relativamente menor de ADN en el tiempo electroforético de 76 s (región B), que corresponde a un tamaño de ADN de ~292 bp.

De acuerdo con el protocolo del fabricante, los conjuntos de adaptador y cebador de ADN que tienen un tamaño total de 122 pb se introdujeron en el ADN del plasma para la construcción de la biblioteca de secuenciación. Por lo tanto, la región A corresponde a fragmentos de ADN del plasma de aproximadamente 78 pb a 145 pb, y la región B corresponde a fragmentos de ADN del plasma de aproximadamente 170 pb. Dicha deducción se puede adaptar a diferentes protocolos para la construcción de bibliotecas de ADN. Por ejemplo, durante la preparación de la biblioteca de secuenciación de lectura única de Illumina, se introducirían 92 pb de conjuntos de adaptador/cebador, mientras que este tamaño sería de 119 pb para la preparación de la biblioteca de secuenciación de extremos emparejados convencional.

En otra realización, el ADN del plasma se puede amplificar mediante un sistema de amplificación de genoma completo conocido para los expertos en la técnica, p. ej., kit de Rubicon Genomics PlasmaPlex WGA (www.rubicongenomics.com/products). Después, los productos amplificados se pueden analizar mediante el Bioanalyzer. En otras realizaciones más, los productos amplificados se pueden analizar por un sistema electroforético, p. ej., de Caliper (www.caliperls.com/products/labchip-systems). En otras realizaciones más, la distribución de tamaños del ADN del plasma se puede analizar directamente, sin amplificación, usando, por ejemplo, un secuenciador basado en nanoporos (p. ej., de Oxford Nanopore Technologies (www.nanoporetech.com)), o un secuenciador de ADN de Helico (www.helicobio.com).

IV. PARÁMETROS DE TAMAÑO

Como se ha mencionado antes, varios parámetros pueden proporcionar una medida estadística de un perfil de tamaños de fragmentos de ADN en la muestra biológica. Un parámetro se puede definir usando los tamaños de todos los fragmentos de ADN analizados, o solo una parte. En una realización, un parámetro proporciona una abundancia relativa de fragmentos de ADN cortos y largos, donde el ADN corto y largo puede corresponder a tamaños o intervalos de tamaños específicos.

Para investigar si la distribución de tamaños global del ADN del plasma materno se puede usar para reflejar la concentración fraccionaria de ADN fetal, los autores de la invención han usado diferentes parámetros para cuantificar la abundancia relativa del ADN corto y largo y determinar la correlación entre estos parámetros y las concentraciones fraccionarias de ADN fetal. Los resultados de estas investigaciones se proporcionan en las siguientes secciones. Los parámetros que se usaron, para fines de ilustración, para reflejar la abundancia relativa de ADN corto incluyen:

- i. Proporción de fragmentos de ADN de 150 pb o inferior, que están marcados CF (tamaño ≤ 150). CF se refiere a la frecuencia acumulativa. Por lo tanto, CF (tamaño ≤ 150) se refiere a la frecuencia acumulativa de fragmentos menores o iguales a 150 bp;
- ii. Relación de las cantidades de fragmentos de ADN ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada (CF(tamaño ≤ 150)/tamaño(163-169));
- iii. Relación de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada (tamaño(140-146)/tamaño(163-169));
- iv. Relación de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada (tamaño(140-154)/tamaño(163-169)); y
- v. Relación de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada (tamaño(100-150)/tamaño(163-169)).

Otros ejemplos de parámetros son los contadores de frecuencia de un histograma. En una realización, se pueden usar múltiples parámetros. Por ejemplo, el valor de cada parámetro puede dar un porcentaje de diferencia y después se puede determinar un porcentaje medio. En otra realización, cada parámetro corresponde a una dimensión diferente de una función de calibración multidimensional, donde los valores de los parámetros para una nueva muestra corresponden a una coordenada en la correspondiente superficie multidimensional.

V. CORRELACIÓN DEL TAMAÑO CON LA CONCENTRACIÓN FRACCIONARIA

Los dos conjuntos de muestras que usan secuenciación, se usan para ilustrar la correlación de los diferentes parámetros de tamaño con la concentración fraccionaria. También se proporciona un análisis del tamaño de elementos que se repiten. Los datos de electroforesis también muestran una correlación entre los parámetros de tamaño y la concentración fraccionaria.

A. Primer conjunto de muestras

La figura 5A es una gráfica 500 que muestra una proporción de fragmentos de ADN que son de 150 pb o inferior, para muestras que tienen diferentes porcentajes de ADN fetal en el plasma materno de acuerdo con las realizaciones de la presente invención. Se representa la proporción de ADN ≤ 150 pb frente a la concentración fraccionaria de ADN fetal para las 80 muestras de plasma materno. Las muestras euploides se representan por círculos negros. Las muestras de trisomía 13 (T13) se representan por triángulos blancos. Las muestras de trisomía 18 (T18) se representan por rombos blancos y las muestras de trisomía 21 (T21) se representan por triángulos

blancos invertidos.

Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la proporción de ADN ≤ 150 pb para todas las muestras (coeficiente de correlación de Pearson = 0,787). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal parece que es consistente a lo largo de las muestras con diferente estado cromosómico fetal. Estos resultados sugieren que el análisis del parámetro de tamaño es útil para calcular la concentración fraccionaria de ADN fetal en una muestra de plasma materno. Por consiguiente, los puntos de datos en la figura 5 se pueden usar como puntos de datos de calibración del procedimiento 300. Después, si se determina que el parámetro CF (tamaño ≤ 150 pb) es 30 para una muestra nueva, se puede calcular que el porcentaje de ADN fetal está entre aproximadamente 7% y 16%. Los puntos de datos en la figura 5 también se pueden usar para determinar una función de calibración que se ajusta a los puntos de datos sin procesar mostrados.

La figura 5B es una gráfica 550 que muestra una relación de tamaños de las cantidades de fragmentos de ADN ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como (CF(tamaño ≤ 150)/tamaño(163-169)). La relación CF(tamaño ≤ 150)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 80 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación CF(tamaño ≤ 150)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,815). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica fetal.

La figura 6A es una gráfica 600 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(140-146)/tamaño(163-169)). La relación tamaño(140-146)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 80 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(140-146)/tamaño(163-169) ratio para todas las muestras (coeficiente de correlación de Pearson = 0.808). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica fetal.

FIG. 6B es una gráfica 650 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(140-154)/tamaño(163-169)). La relación tamaño(140-154)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 80 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(140-154)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,802). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal parece que es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica fetal.

FIG. 7 es una gráfica 700 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada (tamaño(100-150)/tamaño(163-169)). La relación tamaño(100-150)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 80 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(100-150)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,831). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica fetal.

45 B. Segundo conjunto de muestras

La figura 8 es una gráfica 800 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para muestras que tienen diferentes porcentajes de ADN fetal en el plasma materno de acuerdo con las realizaciones de la presente invención. La proporción de ADN ≤ 150 pb se representa frente a la concentración fraccionaria de ADN fetal para las 48 muestras de plasma materno que se secuenciaron de forma masiva en paralelo por extremos emparejados después de enriquecimiento de la diana. Las muestras euploides se representan por círculos negros. Las muestras de trisomía 13 (T13) se representan por triángulos blancos. Las muestras de trisomía 18 (T18) se representan por rombos blancos y las muestras de trisomía 21 (T21) se representan por triángulos blancos invertidos. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la proporción de ADN ≤ 150 pb para todas las muestras (coeficiente de correlación de Pearson = 0,816). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferentes estados cromosómicos fetales. Estos resultados sugieren que el análisis del parámetro de tamaño es útil para calcular la concentración fraccionaria de ADN fetal en una muestra de plasma materno.

La figura 9A es una gráfica 900 que muestra una relación de tamaños de las cantidades de fragmentos de ADN ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como (CF(tamaños ≤ 150)/tamaño(163-169)). La relación CF(tamaño ≤ 150)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 48 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación CF(tamaño ≤ 150)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,776). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica.

La figura 9B es una gráfica 950 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(140-146)/tamaño(163-169)). La relación tamaño(140-146)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 48 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(140-146)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,790). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica.

La figura 10A es una gráfica 1000 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(140-154)/tamaño(163-169)). La relación tamaño(140-154)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 48 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(140-154)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,793). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica.

La figura 10B es una gráfica 1005 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(100-150)/tamaño(163-169)). La relación tamaño(100-150)/tamaño(163-169) se representa frente a la concentración fraccionaria de ADN fetal para las 48 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación tamaño(100-150)/tamaño(163-169) para todas las muestras (coeficiente de correlación de Pearson = 0,798). La correlación positiva entre el parámetro de tamaño y la concentración fraccionaria de ADN fetal es consistente a lo largo de las muestras con diferente estado de ploidía cromosómica.

C. Repeticiones

Antes se ha demostrado que el tamaño de todos los fragmentos de ADN cartografiados en el plasma materno se correlaciona con la concentración fraccionaria de ADN fetal. En esta sección, se investiga si el análisis del tamaño de los elementos de repetición en el genoma también se puede usar para el cálculo de la concentración fraccionaria de ADN fetal en el plasma. En el presente ejemplo, se analizó la distribución de tamaños de los fragmentos de ADN que forman el mapa de las repeticiones *Alu* del genoma.

La figura 11 es una gráfica que muestra una relación de tamaños representada frente al porcentaje de ADN fetal para el tamaño de los elementos de repetición de acuerdo con las realizaciones de la presente invención. Este ejemplo usa la relación de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb (tamaño(100-150)/tamaño(163-169)) para reflejar la alteración en la distribución de tamaños frente al porcentaje de ADN fetal. Hay una correlación positiva entre la relación de tamaños y la concentración fraccionaria de ADN fetal (coeficiente de correlación de Pearson = 0,829). Este resultado sugiere que el análisis de tamaños de los elementos de repetición también se puede usar para determinar la concentración fraccionaria de ADN fetal en una muestra materna.

Además de usar la secuenciación masiva en paralelo, también se pueden usar otros procedimientos, p. ej., PCR, PCR en tiempo real y análisis de espectrometría de masas, para determinar la distribución de tamaños de los elementos de repetición (p. ej., repeticiones *Alu*) en el plasma materno. En una realización, el ADN en la muestra de plasma materno se puede ligar a un conector. Después, se puede llevar a cabo la PCR usando un cebador específico para las secuencias *Alu* y el otro cebador específico para el conector. Después de la PCR, los productos de la PCR se podrían analizar según sus tamaños, p. ej., por electroforesis, espectrometría de masas o secuenciación masiva en paralelo. Esto permitiría una lectura de los tamaños de las secuencias derivadas de las repeticiones *Alu* en el plasma materno. Esta estrategia se puede usar para otra secuencia diana o familia de secuencias. Además, a la PCR se puede seguir una PCR anidada que implica otro cebador específico para *Alu*, en combinación con el mismo cebador específico para el conector o un cebador anidado dentro del conector. Dicha PCR anidada tendría la ventaja de aumentar la especificidad de la amplificación hacia la secuencia de interés (que en este caso son las secuencias *Alu*).

Una ventaja de usar elementos de repetición es que tienen un número de copias relativamente alto y por lo tanto pueden ser más fáciles de analizar. Por ejemplo, se pueden usar menos ciclos de amplificación. Además, con un número mayor de copias, la precisión analítica es potencialmente mayor. La potencial desventaja es que algunas clases de elementos de repetición pueden tener números de copias que varían de un individuo a otro.

D. Electroforesis

La figura 12A es un electroferograma 1200 que se puede usar para determinar una relación de tamaños de acuerdo con las realizaciones de la presente invención. Para todas las bibliotecas de ADN analizadas, había un pico agudo a aproximadamente 292 pb, seguido de un pico secundario en el intervalo de 300 pb a 400 pb. Puesto que el área bajo la curva para un intervalo de tamaños puede representar la cantidad relativa de fragmentos de ADN de esa

región, se usó la relación del área de las regiones A (de 200 pb a 267 pb) y B (de 290 pb a 294 pb) para cuantificar la abundancia relativa de fragmentos de ADN cortos y largos. Primero se ajustó de forma manual el valor base de las unidades de fluorescencia (UF) a 0 y después se generó el área para la región seleccionada.

5 La figura 12B es una gráfica 1250 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 200 pb a 267 pb y ADN de 290 pb a 294 pb (es decir, la relación de las áreas de la región A y B mostrada en el electroferograma) para muestras que tienen diferentes porcentajes de ADN fetal en el plasma materno de acuerdo con las realizaciones de la presente invención. Había un caso T13 que mostraba un pico bajo a 292 pb con un valor de UF de 6,1, mientras que todos los demás casos mostraban un valor de $UF \geq 20$ UF. Puesto que el valor bajo de UF haría el área de medición imprecisa, este caso se ignoró del análisis. La relación de las áreas de la región A y B se representa frente a la concentración fraccionaria de ADN fetal para las otras 79 muestras de plasma materno. Hay una relación positiva entre la concentración fraccionaria de ADN fetal y la relación de áreas A y B para estas muestras (coeficiente de correlación de Pearson = 0,723).

15 VI. DETERMINACIÓN DE LOS PUNTOS DE DATOS DE CALIBRACIÓN

Como se ha mencionado antes, los puntos de datos de calibración se pueden definir en una variedad de formas. Además, los puntos de datos de calibración se pueden obtener en una variedad de formas. Por ejemplo, los puntos de datos de calibración se pueden leer simplemente de la memoria como una serie de valores de calibración de un parámetro junto con la correspondiente concentración fraccionaria. También se puede leer una función de calibración de la memoria (p. ej., una función lineal o no lineal con una forma funcional predeterminada), donde la función define los puntos de datos de calibración. En algunas realizaciones, los puntos de datos de calibración se pueden calcular a partir de los datos medidos de las muestras de calibración.

25 A. Procedimiento

La figura 13 es un diagrama de flujo de un procedimiento 1300 para determinar puntos de datos de calibración a partir de mediciones hechas de muestras de calibración de acuerdo con las realizaciones de la presente invención. Las muestras de calibración incluyen ADN clínicamente relevante y otro ADN.

30 En el bloque 1310, se reciben una pluralidad de muestras de calibración. Las muestras de calibración se pueden obtener como se describe en el presente documento. Cada muestra se puede analizar por separado por experimentos separados o por alguno medio de identificación (p. ej., marcando un fragmento de ADN con un código de barras) para identificar de que muestra era una molécula. Por ejemplo, se puede recibir una muestra de calibración en una máquina, p. ej., una máquina de secuenciación, que produce datos de medición (p. ej., lecturas de secuencias) que se pueden usar para determinar tamaños de los fragmentos de ADN, o se recibe en una máquina de electroforesis.

40 En el bloque 1320, se mide la concentración fraccionaria del ADN clínicamente relevante en cada una de la pluralidad de muestras de calibración. En diferentes realizaciones se puede usar la medición de la concentración de ADN fetal, una secuencia de herencia paterna o un marcador epigenético específico fetal. Por ejemplo, un alelo de herencia paterna estaría ausente de un genoma de la mujer embarazada y se puede detectar en el plasma materno en un porcentaje que es proporcional a la concentración fraccionaria de ADN fetal. Los marcadores epigenéticos específicos fetales pueden incluir secuencias de ADN que presentan patrones de metilación de ADN específicos del feto o placenta en el plasma materno.

50 En el bloque 1330, se miden cantidades de fragmentos de ADN de cada muestra de calibración para diferentes tamaños. Los tamaños se pueden medir como se describe en el presente documento. Los tamaños se pueden contar, representar gráficamente, usar para crear un histograma u otros procedimientos de clasificación para obtener datos relacionados con un perfil de tamaños de la muestra de calibración.

En el bloque 1340, se calcula un valor de calibración para un parámetro basado en las cantidades de fragmentos de ADN en múltiples tamaños. Se puede calcular un valor de calibración para cada muestra de calibración. En una realización, se usa el mismo parámetro de muestra para cada valor de calibración. Sin embargo, las realizaciones pueden usar múltiples parámetros como se describe en el presente documento. Por ejemplo, la fracción acumulativa de fragmentos de ADN menores de 150 bases se puede usar como el parámetro, y las muestras con diferente concentración fraccionaria tendrían probablemente diferentes valores de calibración. Se puede determinar para cada muestra un punto de dato de calibración, donde el punto de datos de calibración incluye el valor de calibración y la concentración fraccionaria medida para la muestra. Estos puntos de datos de calibración se pueden usar en el procedimiento 300, o se pueden usar para determinar los puntos de datos de calibración finales (p. ej., como se define por un ajuste de función).

65 En el bloque 1350, se determina una función que aproxima los valores de calibración a lo largo de una pluralidad de concentraciones fraccionarias. Por ejemplo, se podría ajustar una función lineal a los valores de calibración como función de la concentración fraccionaria. La función lineal puede definir los puntos de datos de calibración que se van a usar en el procedimiento 300.

En algunas realizaciones, se pueden calcular valores de calibración para múltiples parámetros para cada muestra. Los valores de calibración para una muestra pueden definir una coordenada multidimensional (donde cada dimensión es para cada parámetro) que junto con la concentración fraccionaria puede proporcionar un punto de dato. Por lo tanto, en una implementación, se puede ajustar una función multidimensional a todos los puntos de datos multidimensionales. Por consiguiente, se puede usar una curva de calibración multidimensional, donde los diferentes valores de los parámetros se pueden introducir de forma efectiva en una sola función de calibración que da como resultado la concentración fraccionaria. Y, la única función de calibración puede ser resultado de un ajuste funcional de todos los puntos de datos obtenidos de las muestras de calibración.

B. Medición de la concentración de ADN tumoral

Como se ha mencionado, también se pueden aplicar realizaciones a la concentración de ADN tumoral en una muestra biológica. A continuación, se da un ejemplo que implica la determinación de la concentración fraccionaria de ADN tumoral.

Se recogieron muestras de plasma de dos pacientes que padecían carcinoma hepatocelular (HCC) antes y después de resección quirúrgica de los tumores. Se llevó a cabo el análisis de tamaños usando secuenciación masiva en paralelo de extremos emparejados (PE). Se construyeron bibliotecas de secuenciación del ADN del plasma materno como se ha descrito previamente (Lo YM y col. *Sci Transl Med* 2010; 2:61ra91). Todas las bibliotecas se secuenciaron mediante un HiSeq 2000 (Illumina) usando el formato de 50 pb x 2 PE. Las lecturas de secuencias de 50 pb se alinearon con el genoma de referencia humano sin repeticiones enmascaradas (Hg18) (<http://genome.ucsc.edu>), usando el Short Oligonucleotide Alignment Program 2 (SOAP2) (soap.genomics.org.cn). El tamaño de cada uno de los fragmentos secuenciados se infirió de las coordenadas de los nucleótidos más externos en cada extremo de los fragmentos alineados.

Se llevó a cabo el genotipo del ADN extraído de las células sanguíneas y la muestra de tumor de los pacientes con HCC usando el sistema de micromatriz Affymetrix SNP6.0. Para cada caso, las regiones que demuestran pérdida de heterocigosidad (LOH) en el tejido tumoral se identificaron usando Affymetrix Genotyping Console v4.0 basado en las intensidades de los diferentes alelos de los locus de SNP. Las concentraciones fraccionarias del ADN derivadas del tumor (F) se calcularon a partir de la diferencia en cantidades de secuencias que llevan los alelos eliminados y no eliminados en las regiones LOH usando la siguiente fórmula: $F = (A-B) / A \times 100\%$, donde A es el número de lecturas de secuencias que llevan alelos no eliminados en los SNP heterocigotos en las regiones LOH, y B es el número de lecturas de secuencias que llevan alelos eliminados para los SNP heterocigotos en las regiones LOH. La tabla 3 muestra los resultados.

La tabla 3 muestra información de secuenciación y concentración fraccionaria medida del ADN tumoral en muestras de plasma.

Caso N°	Tiempo de toma de muestra	N° de lecturas secuenciadas	Concentración fraccionaria de ADN tumoral en el plasma (%)
1	antes de resección tumoral	448M	51,60
	después de resección tumoral	486M	0,90
2	antes de resección tumoral	479M	5,60
	después de resección tumoral	542M	0,90

En otra realización, se puede usar un locus que presenta duplicación. Por ejemplo, un tumor puede presentar una ganancia de una copia de uno de los dos cromosomas homólogos de modo que un alelo está duplicado. Entonces, se puede determinar una primera cantidad A de lecturas de secuencias que tienen un alelo no duplicado en el uno o más locus heterocigotos (p. ej., SNP) y una segunda cantidad B de lecturas de secuencias que tienen un alelo duplicado en el locus heterocigoto. La concentración fraccionaria F del ADN clínicamente relevante se puede calcular como una relación de la primera cantidad y la segunda cantidad usando una relación $(B-A)/A$.

En otra realización, se pueden usar uno o más locus homocigotos. Por ejemplo, se pueden identificar uno o más locus donde el paciente es homocigoto y donde está presente una mutación de un solo nucleótido en el tejido tumoral. Entonces, se puede determinar una primera cantidad A de lecturas de secuencias que tienen un alelo de tipo natural en el uno o más locus homocigotos. Y se puede determinar una segunda cantidad B de lecturas de secuencias que tienen un alelo mutado en el uno o más locus homocigotos. La concentración fraccionaria F del ADN clínicamente relevante se puede calcular como una relación de la primera cantidad y la segunda cantidad usando una relación $(B-A)/A$.

C. Ejemplo de ajuste funcional de puntos de datos

Se describe ahora un ejemplo de llevar a cabo un ajuste funcional de los valores de parámetros determinados a partir de las muestras de calibración. Se analizaron muestras de plasma de 80 mujeres embarazadas que llevan un solo feto de varón. Entre estas 80 mujeres embarazadas, 39 llevaban fetos euploides, 13 llevaban fetos con trisomía

13 (T13), 10 llevaban trisomía 18 (fetos T18) y 18 llevaban fetos con trisomía 21 (T21). La edad gestacional mediana de las mujeres embarazadas era 13 semanas y 1 día. Se extrajo ADN de muestras de plasma y se secuenciaron usando la plataforma Illumina HiSeq2000 como se ha descrito (Zheng YW y col. *Clin Chem.* 2012; 58:549-58.) excepto que la secuenciación se hizo en un formato 8-plex. Para cada molécula de ADN, se secuenciaron 50 nucleótidos de cada uno de los dos extremos y se alinearon con un genoma de referencia (hg28).

Después, se dedujo el tamaño de cada molécula secuenciada de las coordenadas de los nucleótidos más externos en ambos extremos. Para cada muestra, una mediana de 11,1 millones de fragmentos se secuenciaron y alinearon únicamente con el genoma de referencia. Se calculó una relación dividiendo la proporción de moléculas de ADN con tamaños de 100 pb a 150 pb entre la proporción de moléculas de ADN con tamaños de 163 pb a 169 pb y esta relación se denomina relación de tamaños. Puesto que todos los 80 embarazos llevaban un feto de varón, la proporción de lecturas de secuencia que se alineaban únicamente con el cromosoma Y se usó para determinar la concentración fraccionaria de ADN fetal en cada muestra de ADN de plasma.

Las muestras se dividieron aleatoriamente en dos grupos, en concreto el grupo de entrenamiento y el grupo de validación. La relación entre la concentración fraccionaria de ADN fetal y la relación de tamaños se estableció basada en las muestras en el grupo de entrenamiento usando regresión lineal. Después, la relación de tamaños se usó para deducir la concentración fraccionaria de ADN fetal para la validación usando la fórmula de regresión lineal. La validación se describe en la siguiente sección.

La figura 14A es una gráfica 1400 de una relación de tamaños frente a la concentración fraccionaria de ADN fetal para el grupo de entrenamiento de acuerdo con las realizaciones de la presente invención. Como se ha mencionado antes, la relación de tamaños se calcula dividiendo la proporción de moléculas de ADN con tamaños de 100 pb a 150 pb entre la proporción de moléculas de ADN con tamaños de 163 pb a 169 pb. La relación de tamaños se representa gráficamente frente a la concentración fraccionaria de ADN fetal, como se muestra mediante los puntos de datos 1405. Los círculos blancos representan los casos euploides. Los símbolos negros representan los casos aneuploides (cuadrados para T13, círculos para T18 y triángulos para T21). La recta de regresión lineal 1410 resulta del ajuste funcional a los puntos de datos. El ajuste funcional se puede llevar a cabo por cualquier técnica adecuada, p. ej., por mínimos cuadrados. La recta 1410 se puede usar para calcular valores de parámetros medidos para otras muestras, no en el grupo de entrenamiento. Cada parte de la recta 1410 se puede considerar un punto de datos de calibración.

VII. COMPARACIÓN DE LOS PUNTOS DE DATOS DE CALIBRACIÓN

Como se ha mencionado antes, los puntos de datos de calibración se pueden usar para determinar la concentración fraccionaria de ADN clínicamente relevante. Por ejemplo, se pueden usar los puntos de datos sin procesar 1405 en la figura 14A para proporcionar un intervalo de concentración fraccionaria de ADN para un valor de calibración particular (relación de tamaños marcada en la fig. 14A), donde el intervalo se puede usar para determinar si la concentración fraccionaria está por encima de una cantidad umbral. En lugar de un intervalo, se puede usar una media de las concentraciones fraccionarias en una relación de tamaños particular. Por ejemplo, la concentración fraccionaria que corresponde a una medición de 1,3 como la relación de tamaños en una muestra nueva se puede determinar como la concentración mediana calculada a partir de los dos puntos de datos en 1,3. En una realización, se puede usar un ajuste funcional (p. ej., recta 1410).

La figura 14B es una gráfica 1450 de concentraciones fraccionarias deducidas (calculadas) a partir de la función lineal 1410 de la fig. 14A frente a las concentraciones fraccionarias medias usando secuencias específicas fetales de acuerdo con las realizaciones de la presente invención. Usando la ecuación de regresión (es decir, la recta 1410) determinada basada en los datos del grupo de entrenamiento, la relación de tamaños determinadas para una muestra de validación se usó para deducir la concentración fraccionaria de ADN fetal para las muestras del grupo de validación. Las concentraciones fraccionarias medidas corresponden a la proporción de secuencias del cromosoma Y en la muestra de ADN plasmático (es decir, la proporción de lecturas de secuencias que se alinean con el cromosoma Y).

La recta 1460 representa la correlación perfecta entre los dos grupos de valores. La desviación de un punto de dato 1455 indica cómo de preciso es el cálculo, siendo perfectamente precisos los puntos sobre la recta 1460. Como se indica en el presente documento, el cálculo no tiene que ser perfectamente preciso, puesto que el ensayo deseado puede ser simplemente determinar si hay un porcentaje suficiente de ADN clínicamente relevante en la muestra biológica. Los círculos blancos representan los casos euploides. Los símbolos negros representan los casos aneuploides (cuadrados para T13, círculos para T18 y triángulos para T21). La diferencia mediana entre la concentración fraccionaria de ADN fetal deducida de la relación de tamaños y la medida a partir de la proporción de secuencias de cromosoma Y era 2,1%. La diferencia era menor que 4,9% en 90% de las muestras.

Se usaron muestras con diferente estado de ploidía tanto en el grupo de calibración como en el grupo de validación. Como se muestra en la figura 14A, la relación entre la relación de tamaños y la concentración fraccionaria de ADN fetal era consistente a lo largo de las muestras con diferente estado de ploidía. Como resultado, la concentración fraccionaria de ADN fetal se puede deducir de la relación de tamaños de la muestra sin un conocimiento previo del

estado de ploidía de las muestras como se ilustra en la fig. 14B. Se usó una curva de calibración para muestras con diferente estado de ploidía y, por lo tanto, no es necesario conocer el estado de ploidía de la muestra antes de usar realizaciones para determinar la concentración fraccionaria de ADN fetal.

5 VIII. CÁNCER

Como se describe en el presente documento, se pueden usar realizaciones para calcular la concentración fraccionaria de ADN tumoral en una muestra biológica. Como en los ejemplos fetales, se pueden usar muestras de calibración para determinar puntos de datos de calibración, p. ej., mediante ajuste de una función (p. ej., una función lineal) a los puntos de datos que muestran una correlación entre un valor de un parámetro de tamaño y una concentración fraccionaria medida.

A. Correlación del tamaño con la concentración de ADN tumoral

15 La fig. 15A es una gráfica 1500 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para muestras que tienen diferentes porcentajes de ADN tumoral en el plasma de dos pacientes con HCC antes y después de resección tumoral de acuerdo con las realizaciones de la presente invención. La proporción de ADN \leq 150 pb se representa frente a las concentraciones fraccionarias de ADN tumoral para los dos pacientes con HCC antes (círculos negros) y después (círculos blancos) de resección tumoral. Los dos círculos blancos están en posiciones muy cercanas uno de otro (efectivamente uno encima de otro). Estos resultados sugieren que el análisis del parámetro de tamaño es útil para calcular la concentración fraccionaria de ADN tumoral en la muestra de plasma de pacientes con HCC. Hay una reducción tanto en la concentración fraccionaria de ADN tumoral como en la proporción de fragmentos de ADN \leq 150 pb después de resección tumoral. El círculo negro 1505 corresponde a una muestra con mucho menor porcentaje de ADN tumoral, que está relacionado con un tamaño menor del tumor. En otras palabras, el paciente con un tumor mayor tiene una proporción mayor de ADN corto, que se refleja en una mayor relación de CF (\leq 150 pb) comparado con el paciente con un tumor más pequeño.

La fig. 15B es una gráfica 1550 que muestra una relación de tamaños de las cantidades de fragmentos de ADN \leq 150 pb y ADN de 163 pb a 169 pb, que está marcada como (CF(tamaño \leq 150)/tamaño(163-169)), para dos pacientes con HCC antes y después de resección tumoral. La relación de CF(tamaño \leq 150)/tamaño(163-169) se representa frente a las concentraciones fraccionarias de ADN tumoral para los dos pacientes con HCC antes (círculos negros) y después (círculos blancos) de resección tumoral. Los dos círculos blancos están en posiciones muy cercanas uno de otro. Hay una reducción tanto en la concentración fraccionaria de ADN tumoral como en la relación de tamaños después de resección tumoral.

La fig. 16A es una gráfica 1600 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada (tamaño(140-146)/tamaño(163-169)), para dos pacientes de HCC antes y después de resección tumoral. La relación tamaño(140-146)/tamaño(163-169) se representa frente a las concentraciones fraccionarias de ADN tumoral para los dos pacientes con HCC antes (círculos negros) y después (círculos blancos) de resección tumoral. Hay una reducción tanto de la concentración fraccionaria de ADN tumoral como de la relación de tamaños después de resección tumoral.

La fig. 16B es una gráfica 1650 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(140-154)/tamaño(163-169)), para dos pacientes de HCC antes y después de resección tumoral. La relación tamaño(140-154)/tamaño(163-169) se representa frente a las concentraciones fraccionarias de ADN tumoral para los dos pacientes con HCC antes (círculos negros) y después (círculos blancos) de resección tumoral. Hay una reducción tanto de la concentración fraccionaria de ADN tumoral como de la relación de tamaños después de resección tumoral.

La fig. 17 es una gráfica 1700 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 pb, que está marcada como (tamaño(100-150)/tamaño(163-169)), para dos pacientes de HCC antes y después de resección tumoral. La relación tamaño(100-150)/tamaño(163-169) se representa frente a las concentraciones fraccionarias de ADN tumoral para los dos pacientes con HCC antes (círculos negros) y después (círculos blancos) de resección tumoral. Hay una reducción tanto de la concentración fraccionaria de ADN tumoral como de la relación de tamaños después de resección tumoral.

B. Disminución de tamaño debido al tratamiento

La fig. 18A es una gráfica 1800 que muestra una proporción de fragmentos de ADN de 150 pb o inferior para pacientes con HCC antes y después de resección tumoral. El par de muestras del mismo paciente de cáncer se representan mediante símbolos iguales conectados mediante una línea de trazos. Hay una disminución general en la proporción de ADN \leq 150 pb para el ADN del plasma en pacientes de cáncer después de resección tumoral.

La separación en los valores de la proporción para el pretratamiento y el postratamiento ilustran una correlación entre la existencia de un tumor y el valor del parámetro de tamaño. La separación en los valores para el pretratamiento y el postratamiento se pueden usar para determinar cómo de satisfactorio era el tratamiento, p. ej.,

comparando la proporción con un umbral, donde una proporción por debajo del umbral puede indicar éxito. En otro ejemplo, se puede comparar una diferencia entre el pretratamiento y el postratamiento con un umbral.

5 La proporción (o cualquier otro valor de un parámetro de tamaño) también se puede usar para detectar la aparición de un tumor. Por ejemplo, se puede determinar un valor base para un parámetro de tamaño. Después, en un tiempo posterior, se puede medir de nuevo un valor para el parámetro de tamaño. Si el valor del parámetro de tamaño muestra un cambio significativo, entonces el paciente puede estar en riesgo mayor de tener un tumor. Si el valor del parámetro de tamaño no varía mucho entre individuos, lo que la fig. 18A indica es que la proporción tampoco lo hace (es decir, puesto que los valores de postratamiento son los mismos), entonces se puede usar el mismo valor base para otros pacientes. Por lo tanto, no es necesario tomar un valor base para cada paciente.

15 La fig. 18B es una gráfica 1850 que muestra una relación de tamaños de las cantidades de fragmentos de ADN ≤ 150 pb y ADN de 163 pb a 169 pb, que está marcada como $(CF(\text{tamaño} \leq 150)/\text{tamaño}(163-169))$, para pacientes con HCC antes y después de resección tumoral. El par de muestras del mismo paciente de cáncer se representan por símbolos iguales conectados por una línea de trazos. Hay una disminución en esta relación de tamaños para los dos casos después de resección tumoral.

20 La fig. 19A es una gráfica 1900 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 146 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-146)/\text{tamaño}(163-169))$, para pacientes con HCC antes y después de resección tumoral. El par de muestras del mismo paciente de cáncer se representan por símbolos iguales conectados por una línea de trazos. Hay una disminución en esta relación de tamaños para los dos casos después de resección tumoral.

25 La fig. 19B es una gráfica 1950 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 140 pb a 154 pb y ADN de 163 pb a 169 pb, que está marcada como $(\text{tamaño}(140-154)/\text{tamaño}(163-169))$, para pacientes de HCC antes y después de resección tumoral. El par de muestras del mismo paciente de cáncer se representan por símbolos iguales conectados por una línea de trazos. Hay una disminución en esta relación de tamaños para los dos casos después de resección tumoral.

30 La fig. 20 es una gráfica 2000 que muestra una relación de tamaños de las cantidades de fragmentos de ADN de 100 pb a 150 pb y ADN de 163 pb a 169 bp, que está marcada como $(\text{tamaño}(100-150)/\text{tamaño}(163-169))$, para pacientes con HCC antes y después de resección tumoral. El par de muestras del mismo paciente de cáncer se representan por símbolos iguales conectados por una línea de trazos. Hay una disminución en esta relación de tamaños para los dos casos después de resección tumoral.

35 C. Procedimiento

La fig. 21 muestra un diagrama de flujo que ilustra un procedimiento 2100 para analizar una muestra biológica de un organismo para determinar una clasificación de un nivel de cáncer de acuerdo con las realizaciones de la presente invención. El procedimiento 2100 puede analizar la muestra biológica del organismo (p. ej., un ser humano). La muestra biológica incluye ADN procedente de células normales y potencialmente de células asociadas con cáncer. Al menos algo del ADN está sin células en la muestra biológica. Se pueden usar aspectos de los procedimientos 300 y 1300 con realizaciones del procedimiento 2100.

45 En el bloque 2110, se miden cantidades de fragmentos de ADN que corresponden a diferentes tamaños. Para cada tamaño de una pluralidad de tamaños, se puede medir una cantidad de una pluralidad de fragmentos de ADN de la muestra biológica que corresponde al tamaño, como se describe para el procedimiento 300. La pluralidad de fragmentos de ADN se puede seleccionar de forma aleatoria o preferiblemente de una o más regiones predeterminadas de un genoma. Por ejemplo, se puede llevar a cabo el enriquecimiento dirigido o se puede usar la selección de lecturas de secuencias que son de regiones particulares del genoma, p. ej., como se ha descrito antes.

50 En el bloque 2120, se calcula un primer valor de un primer parámetro basado en las cantidades de fragmentos de ADN en múltiples tamaños. En un aspecto, el primer parámetro proporciona una medida estadística de un perfil de tamaños (p. ej., un histograma) de fragmentos de ADN en la muestra biológica. El parámetro se puede denominar un parámetro de tamaño puesto que se determina a partir de los tamaños de la pluralidad de fragmentos de ADN. Se proporcionan ejemplos de parámetros en el presente documento. Se pueden usar múltiples parámetros, como se describe también en el presente documento.

60 En el bloque 2130, el primer valor se compara con un valor de referencia. Los ejemplos de un valor de referencia incluyen un valor normal y un valor de corte que es una distancia especificada desde un valor normal (p. ej., en unidades de desviación estándar). El valor de referencia se puede determinar a partir de una muestra diferente del mismo organismo (p. ej., cuando se sabía que el organismo era sano). Por lo tanto, el valor de referencia puede corresponder a un valor del primer parámetro determinado a partir de una muestra cuando se supone que el organismo no tiene cáncer. En una realización, la muestra biológica se obtiene del organismo después de tratamiento y el valor de referencia corresponde a un valor del primer parámetro determinado a partir de una muestra tomada antes del tratamiento (p. ej., ilustrado anteriormente). El valor de referencia también se puede determinar a

partir de muestras de otros organismos sanos.

En el bloque 2140, se determina una clasificación de un nivel de cáncer en el organismo basándose en la comparación. En diferentes realizaciones, la clasificación puede ser numérica, de texto o cualquier otro indicador. La clasificación puede proporcionar un resultado binario de si o no respecto al cáncer, una probabilidad u otra puntuación, que puede ser un valor absoluto o relativo, p. ej., relativo respecto a una clasificación previa del organismo en un tiempo anterior. En una implementación, la clasificación es que el organismo no tiene cáncer o que el nivel de cáncer ha disminuido. En otra implementación, la clasificación es que el organismo tiene cáncer o que ha aumentado un nivel de cáncer.

Como se describe en el presente documento, el nivel de cáncer puede incluir una existencia de cáncer, una etapa del cáncer, o un tamaño de un tumor. Por ejemplo, si el primer valor sobrepasa (p. ej., es mayor o menor que, dependiendo de cómo se define el primer parámetro) se puede usar para determinar si existe el cáncer, o al menos una probabilidad (p. ej., un porcentaje de probabilidad). La extensión por encima del umbral puede proporcionar una mayor probabilidad, lo que puede conducir al uso de múltiples umbrales. Además, la extensión anterior puede corresponder a un nivel diferente de cáncer, p. ej., más tumores o tumores mayores. Por lo tanto, las realizaciones pueden diagnosticar, determinar el estado, pronosticar o seguir la evolución de un nivel de cáncer en el organismo.

D. Determinación de la distribución de tamaños para regiones particulares

Como con otras realizaciones, el primer conjunto de fragmentos de ADN puede corresponder a una o más regiones predeterminadas de un genoma del organismo. Por lo tanto, el análisis de tamaños también se puede llevar a cabo para regiones seleccionadas, p. ej., cromosomas específicos, brazos de cromosomas o múltiples regiones (depósitos) de la misma longitud, p. ej., 1 Mb. Por ejemplo, se puede poner el foco en regiones que están habitualmente alteradas en un tipo de cáncer de interés. La tabla 2200 de la fig. 22 muestra algunas aberraciones de cromosomas comunes vistas en diferentes tipos de cáncer. La ganancia se refiere a una amplificación de un cromosoma con una o más copias adicionales en un segmento particular, y pérdida se refiere a eliminaciones de uno o ambos cromosomas homólogos en un segmento particular.

En una realización, se pueden identificar conjuntos adicionales de fragmentos de ADN de la muestra biológica. Cada conjunto de fragmentos de ADN puede corresponder a diferentes regiones predeterminadas, tales como las regiones especificadas en la tabla 2200. Las regiones que no están asociadas con el cáncer también se podrían usar, p. ej., para determinar un valor de referencia. Se puede determinar la cantidad de fragmentos de ADN que corresponden a diferentes tamaños y se puede determinar el valor de tamaño de un parámetro para cada conjunto adicional de fragmentos de ADN, como se describe en el presente documento. Por lo tanto, se puede determinar un valor de tamaño diferente para cada región genómica, donde hay una correspondencia uno-uno entre un conjunto de fragmentos de ADN y una región genómica.

Cada uno de los valores de tamaños se puede comparar con un valor de referencia respectivo. Se pueden identificar regiones predeterminadas donde el correspondiente valor de tamaño es estadísticamente diferente del respectivo valor de referencia. Cuando un valor de referencia es un valor normal, la determinación de la diferencia estadística se puede hacer comparando un valor de tamaño con un valor de corte (p. ej., donde el valor de corte es un número específico de desviaciones estándar del valor normal, basado en una distribución estadística supuesta o medida). Los valores de referencia respectivos pueden ser iguales o diferentes para regiones diferentes. Por ejemplo, diferentes regiones pueden tener diferentes valores normales para el tamaño.

En una realización, el número de regiones estadísticamente diferentes del valor de referencia, se puede usar para determinar la clasificación. Por lo tanto, se puede determinar el número de regiones predeterminadas de identificación donde el correspondiente valor de tamaño es estadísticamente diferente del respectivo valor de referencia. El número se puede comparar con un número umbral de regiones para determinar la clasificación del nivel de cáncer en el organismo. El número umbral se puede determinar basándose en una variación dentro de las muestras normales y dentro de las muestras de cáncer.

Como se destaca en la tabla 2200, diferentes cánceres están asociados con diferentes partes del genoma. Por lo tanto, aquellas regiones que son estadísticamente diferentes se pueden usar para determinar uno o más posibles tipos de cáncer, cuando los posibles tipos de cáncer están asociados con las regiones identificadas. Por ejemplo, si un valor de tamaño para fragmentos de ADN del segmento de cromosoma 7p se encuentra que es significativamente menor que un valor normal (p. ej., determinado por un valor de corte), entonces el cáncer colorrectal se puede identificar como un cáncer probable cuando la clasificación indica que existe cáncer. Obsérvese que el valor de tamaño para el segmento de cromosoma 7p se puede usar como un solo indicador para determinar la clasificación, o se pueden usar múltiples regiones. En una realización, solo si una clasificación global indica cáncer se usaría el valor de tamaño para el segmento de cromosoma 7p para identificar el cáncer colorrectal como un cáncer probable.

IX. SISTEMA INFORMÁTICO

Cualquiera de los sistemas informáticos mencionados en el presente documento puede usar cualquier número adecuado de subsistemas. Ejemplos de dichos subsistemas se muestran en la fig. 23 en el aparato de ordenador 2300. En algunas realizaciones, un sistema informático incluye un solo aparato de ordenador, donde los subsistemas pueden ser los componentes del aparato de ordenador. En otras realizaciones, un sistema informático puede incluir múltiples aparatos de ordenador, siendo cada uno un subsistema con componentes internos.

Los subsistemas mostrados en la fig. 23 están interconectados por un bus de sistema 2375. Se muestran subsistemas adicionales tales como una impresora 2374, teclado 2378, disco fijo 2379, pantalla 2376, que está acoplada a un adaptador de pantalla 2382, y otros. Los dispositivos periféricos y de entrada/salida (I/O), que se acoplan al controlador I/O 2371, pueden estar conectados al sistema informático por una serie de medios conocidos en la técnica, tal como un puerto serie 2377. Por ejemplo, el puerto serie 2377 o la interfaz externa 2381 (p. ej., Ethernet, Wi-Fi, etc.) se puede usar para conectar el sistema informático 2300 a una red de área extensa tal como Internet, un dispositivo de entrada de ratón, o un escáner. La interconexión por el bus de sistema 2375 permite que el procesador central 2373 se comuniquen con cada subsistema y controle la ejecución de las instrucciones de la memoria del sistema 2372 o el disco fijo 2379, así como el intercambio de información entre subsistemas. La memoria del sistema 2372 y/o el disco fijo 2379 puede incorporar un medio de lectura por ordenador. Cualquiera de los valores mencionados en el presente documento puede salir 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, p. ej., conectados entre sí por la interfaz externa 2381 o por una interfaz interna. En algunas realizaciones, los sistemas informáticos, subsistemas o aparatos pueden comunicarse por una red. En dichos casos, un ordenador se puede considerar un cliente y otro ordenador un servidor, donde cada uno puede ser parte de un mismo sistema informático. Un cliente y un servidor pueden incluir cada uno múltiples sistemas, subsistemas o componentes.

Debe entenderse que cualquiera de las realizaciones de la presente invención se puede implementar en forma de lógica de control usando hardware (p. ej., un circuito integrado específico de aplicación o matriz de puertas programable in situ) y/o usando software informático con un procesador generalmente programable de una forma modular o integrada. Basándose en la descripción y enseñanzas proporcionadas en el presente documento, un experto en la técnica conocerá y apreciará otras formas y/o procedimientos de implementar realizaciones de la presente invención usando hardware y una combinación de hardware y software.

Cualquiera de los componentes de software o funciones descritos en esta solicitud se pueden implementar como código de software para ser ejecutado por un procesador usando cualquier lenguaje informático adecuado tal como, por ejemplo, Java, C++ o Perl, usando, por ejemplo, técnicas convencionales u orientadas a objetos. El código de software se puede almacenar en una serie de instrucciones o comandos en un medio de lectura por ordenador para el almacenamiento y/o transmisión, los medios adecuados incluyen memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM), un medio magnético tal como un disco duro o un disco flexible, o un medio óptico tal como un disco compacto (CD) o DVD (disco versátil digital), memoria flash, y similares. El medio de lectura por ordenador puede ser cualquier combinación de dichos dispositivos de almacenamiento o transmisión.

Dichos programas también pueden ser codificados y transmitidos usando señales vehículo adaptadas para la transmisión por redes con cable, ópticas y/o inalámbricas conforme a una variedad de protocolos, incluyendo internet. Como tal, un medio de lectura por ordenador de acuerdo con una realización de la presente invención, se puede crear usando una señal de datos codificada con dichos programas. El medio de lectura por ordenador codificado con el código de programa puede estar empaquetado con un dispositivo compatible o se puede proporcionar por separado de otros dispositivos (p. ej., por descarga de internet). Cualquiera de dichos medios de lectura por ordenador puede residir sobre o en un solo producto de programa informático (p. ej., un disco duro, un CD o un sistema informático entero), y puede estar presente sobre o en diferentes productos de programas informáticos dentro de un sistema o red. Un sistema informático puede incluir una pantalla, impresora, u otra presentación adecuada para proporcionar cualquiera de los resultados mencionados en el presente documento a un usuario.

Cualquiera de los procedimientos descritos en el presente documento se puede llevar a cabo total o parcialmente con un sistema informático que incluye uno o más procesadores, que se pueden configurar para llevar a cabo las etapas. Por lo tanto, las realizaciones pueden estar dirigidas a sistemas informáticos configurados para realizar las etapas de cualquiera de los procedimientos descritos en el presente documento, potencialmente con diferentes componentes que llevan a cabo las respectivas etapas o un grupo de etapas respectivo. Aunque presentado como etapas numeradas, las etapas de los procedimientos del presente documento se pueden llevar a cabo en el mismo tiempo o en diferente orden. Además, partes de estas etapas se pueden usar con partes de otras etapas de otros procedimientos. Además, todas o partes de una etapa pueden ser opcionales. Además, cualquiera de las etapas de cualquiera de los procedimientos se puede llevar a cabo con módulos, circuitos u otros medios para llevar a cabo estas etapas.

Los detalles específicos de realizaciones particulares se pueden combinar de cualquier manera adecuada sin salirse del alcance de la invención. Sin embargo, otras realizaciones de la invención se pueden dirigir a realizaciones específicas relacionadas con cada aspecto individual o combinaciones específicas de estos aspectos individuales.

5 La descripción anterior de realizaciones de ejemplo de la invención se ha presentado con fines ilustrativos y de descripción. No se pretende ser exhaustivo o limitar la invención a la forma precisa descrita, y son posibles muchas modificaciones y variaciones a la luz de la enseñanza anterior. Las realizaciones se seleccionaron y describieron con el fin de explicar mejor los principios de la invención y sus aplicaciones prácticas para así permitir que otros expertos en la técnica usen mejor la invención en diferentes realizaciones y con diferentes modificaciones según sea adecuado para el uso particular contemplado.

10 Con una relación de “un”, “una” o “el”, “la” se quiere indicar “uno o más” salvo que se indique específicamente lo contrario.

REIVINDICACIONES

- 5 1. Un procedimiento de análisis de una muestra biológica de un sujeto, incluyendo la muestra biológica ADN que se origina a partir de células normales y posiblemente de células asociadas con cáncer, en donde al menos algo del ADN está libre de células en la muestra biológica, comprendiendo el procedimiento:
 - (a) secuenciar una pluralidad de fragmentos de ADN para obtener una pluralidad de lecturas de secuencia que comprenden los nucleótidos más externos en cada extremo de una pluralidad de fragmentos de ADN;
 - 10 (b) alinear las lecturas de secuencia de un genoma de referencia, obteniendo de este modo un conjunto de coordenadas genómicas que incluyen coordenadas genómicas de los nucleótidos más externos que definen un tamaño de un fragmento de ADN para cada una de la pluralidad de fragmentos de ADN;
 - (c) calcular un valor de un parámetro basado en cantidades de lecturas de secuencia de fragmentos de ADN que se alinean con el conjunto de coordenadas genómicas de (b) en múltiples tamaños;
 - (d) comparar el valor con un valor de referencia; y
 - 15 (e) determinar una clasificación de un nivel de cáncer en el sujeto basándose en la comparación del valor con el valor de referencia.
- 20 2. El procedimiento de la reivindicación 1, en donde la muestra biológica se selecciona del grupo que consiste en sangre, plasma, suero, orina y saliva.
3. El procedimiento de la reivindicación 1 o 2, en donde el conjunto de coordenadas genómicas se corresponde on una región del genoma de referencia.
- 25 4. El procedimiento de la reivindicación 3, en donde la región del genoma de referencia está predeterminado.
5. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde la secuenciación comprende la secuenciación de extremos pareados.
- 30 6. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde cada lectura de secuencia de las lecturas de secuencia comprende al menos aproximadamente 30 nucleótidos de al menos un extremo del fragmento de ADN.
- 35 7. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde el valor se determina usando un número de lecturas de secuencia que tiene extremos que se alinean con las coordenadas genómicas del conjunto de coordenadas genómicas para tamaños de los fragmentos de ADN.
- 40 8. El procedimiento de una cualquiera de las reivindicaciones 1 a 6, en donde el valor se determina usando un primer número de lecturas de secuencia que tiene extremos que se alinean con coordenadas genómicas del conjunto de coordenadas genómicas normalizadas por un segundo número de lecturas de secuencia que tiene diferentes coordenadas genómicas para tamaños de los fragmentos de ADN.
9. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde la clasificación se corresponde con un tamaño o una existencia de cáncer, un estadio del cáncer o un tamaño de un tumor.
- 45 10. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde al menos una parte de la pluralidad de los fragmentos de ADN provienen de un tumor.
- 50 11. El procedimiento de una cualquiera de las reivindicaciones anteriores, en donde la muestra biológica se obtiene del sujeto tras un tratamiento, y en donde el valor de referencia se corresponde con un valor determinado de una muestra tomada del sujeto antes del tratamiento.
12. El procedimiento de una cualquiera de las reivindicaciones 1 a 10, en donde el valor de referencia se corresponde con un valor determinado de una muestra cuando se presume que el sujeto no tiene cáncer.
- 55 13. El procedimiento de una cualquiera de las reivindicaciones 1 a 10, en donde el valor de referencia se establece a partir de una o más muestras biológicas obtenidas de uno o más sujetos sanos.

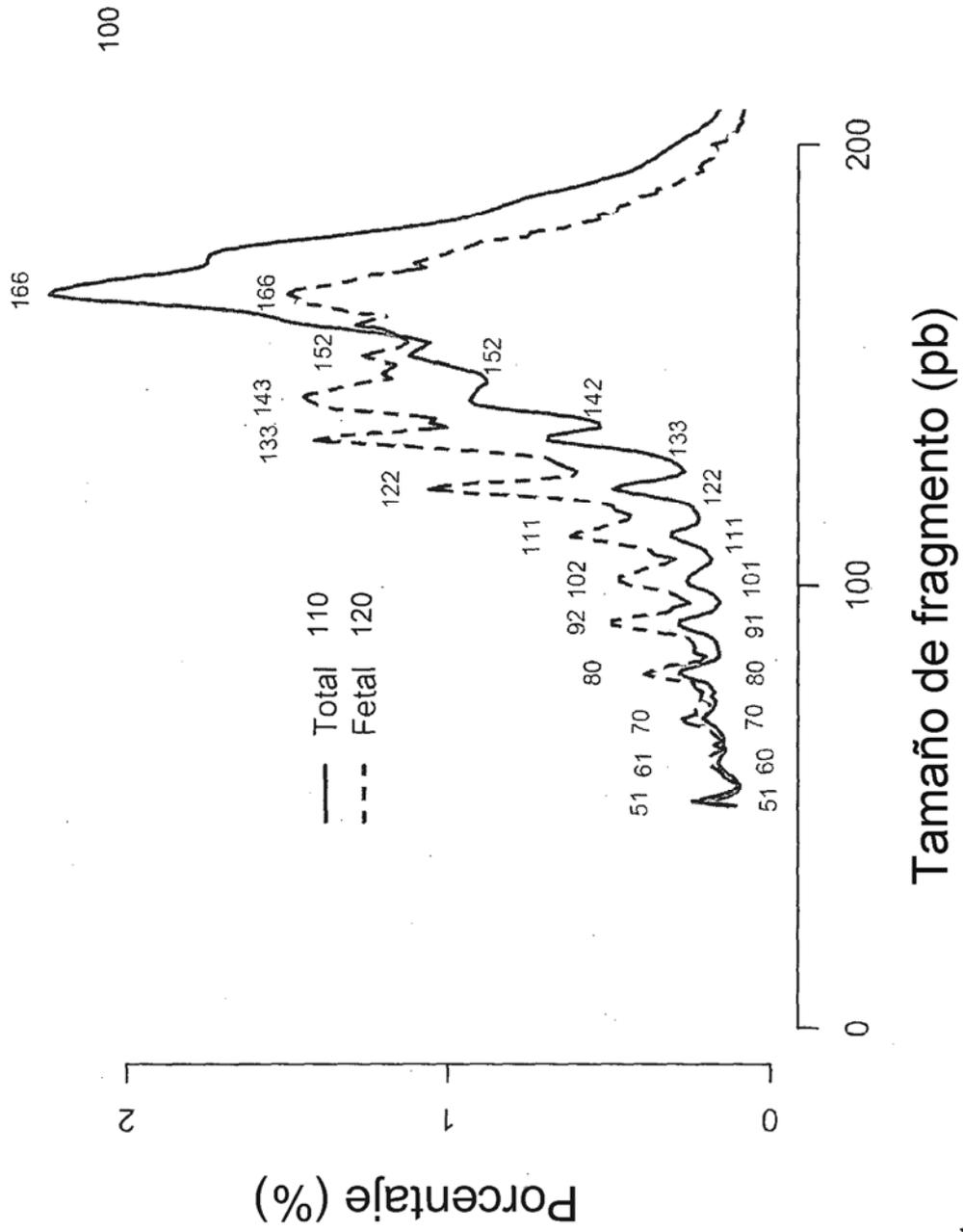


FIG. 1

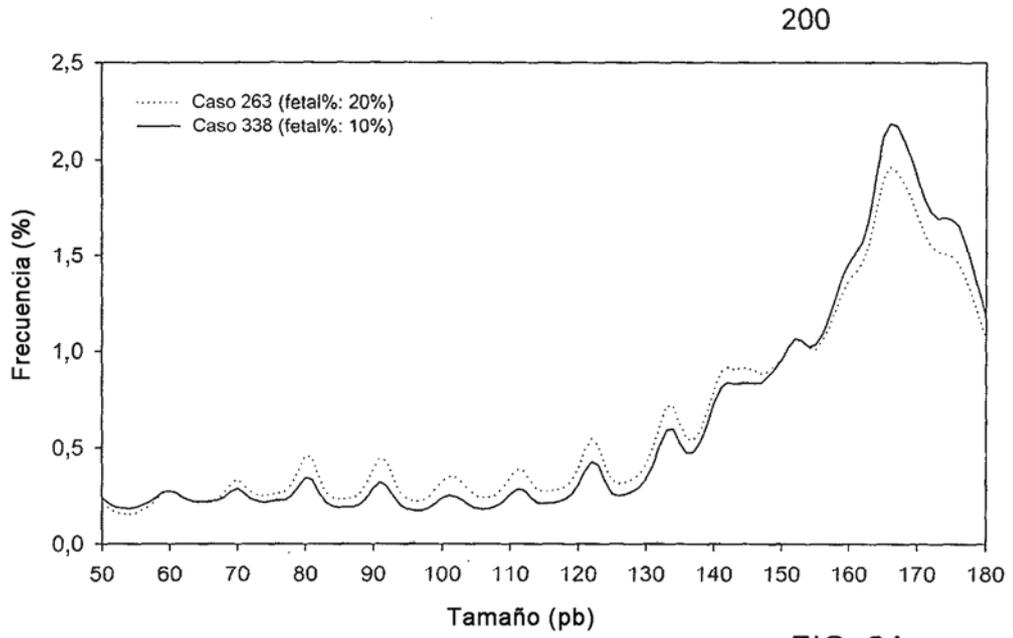


FIG. 2A

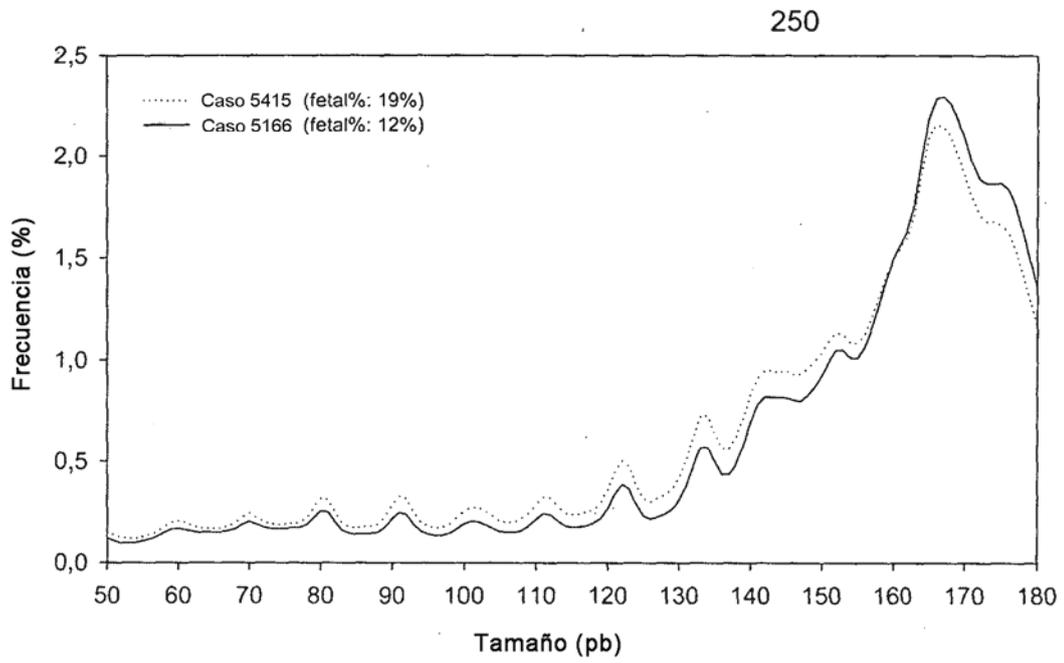


FIG. 2B

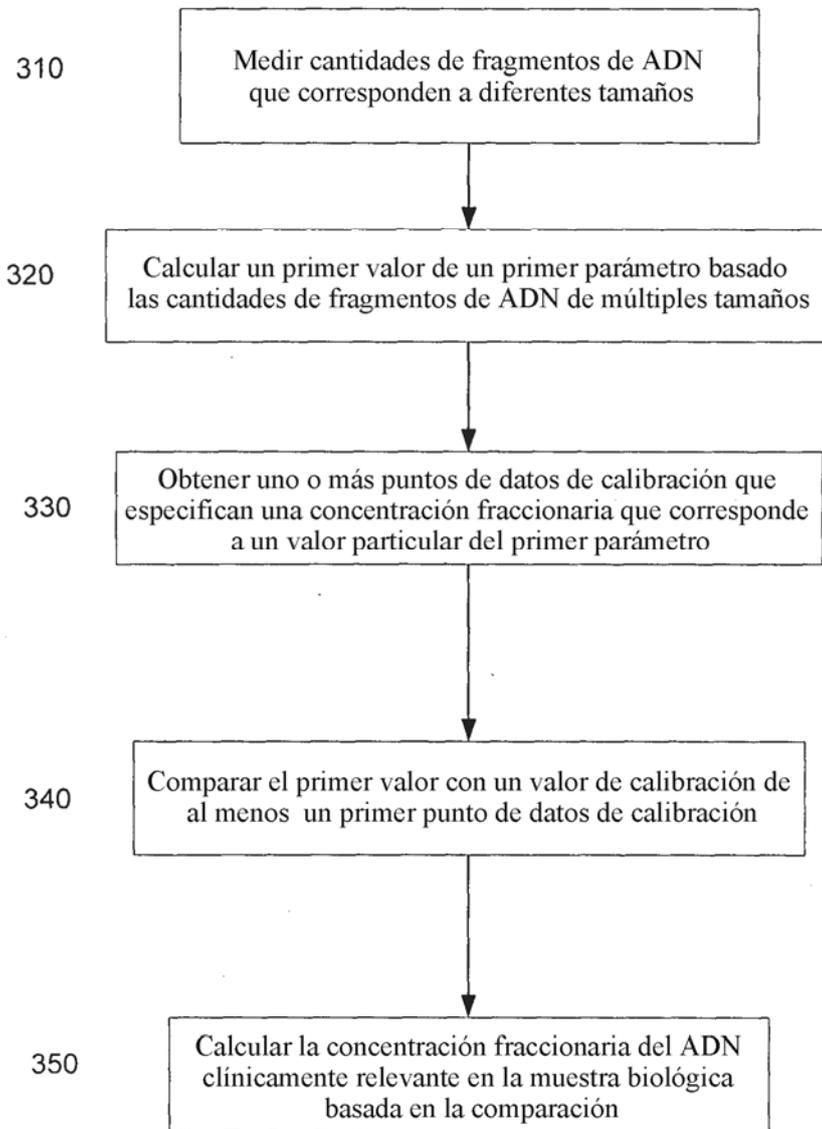


FIG. 3

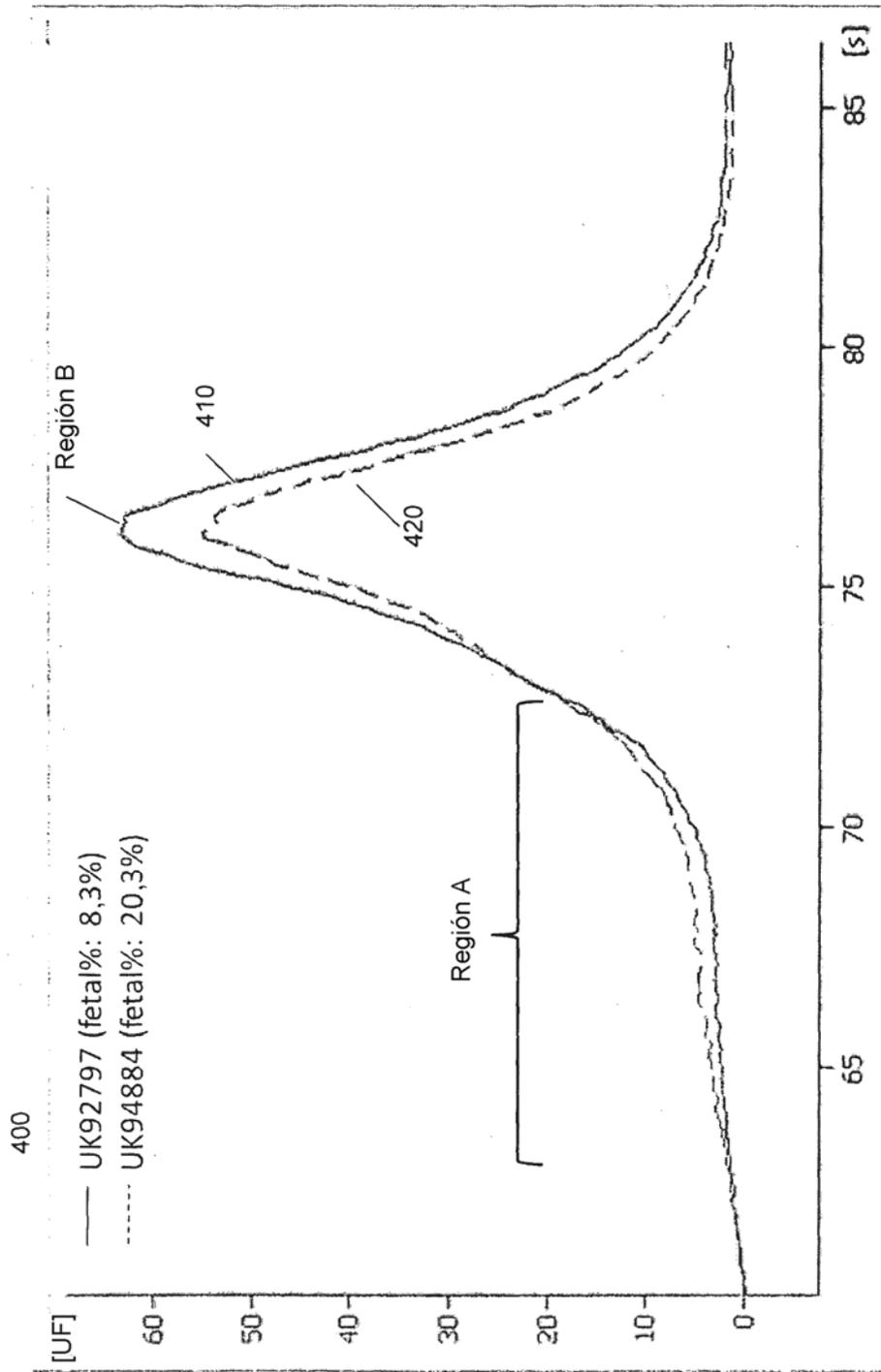
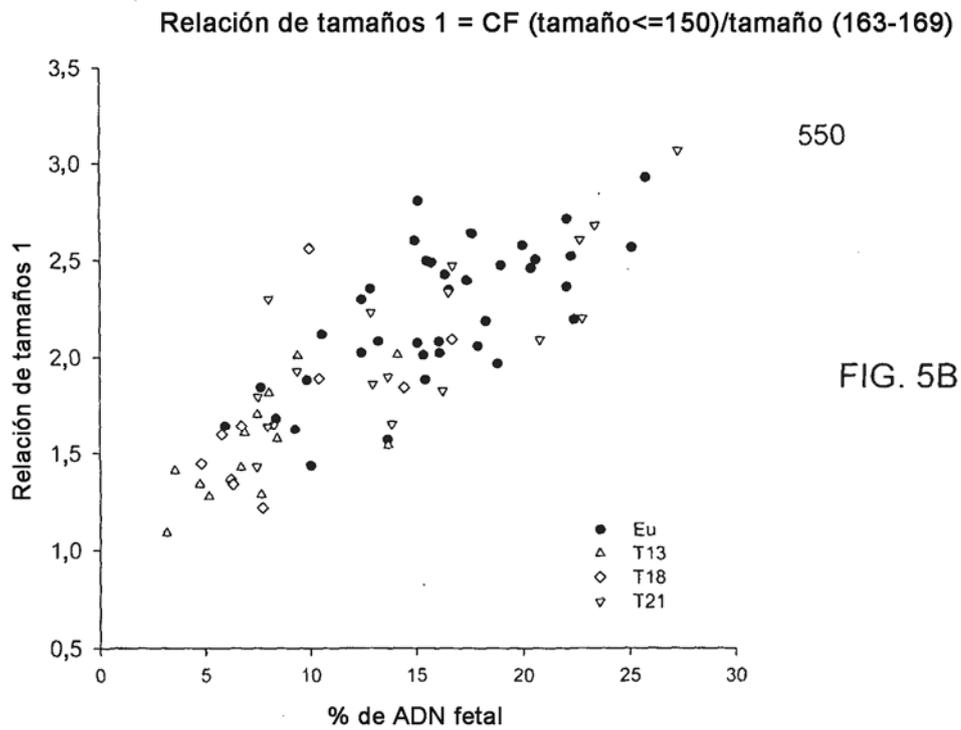
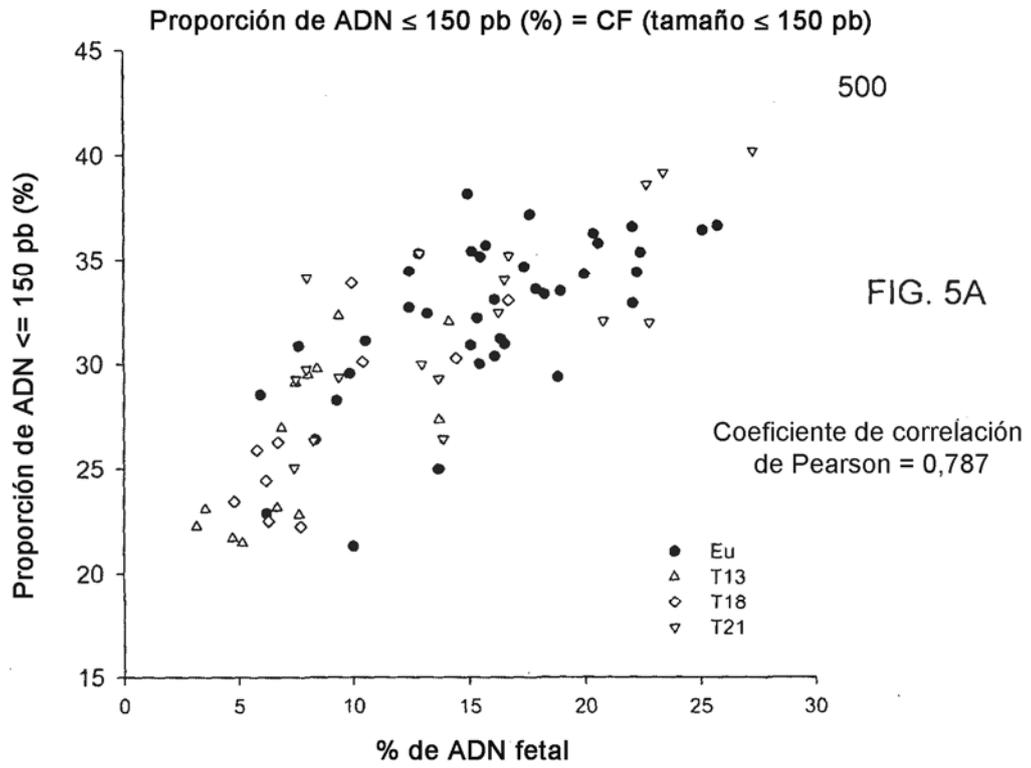
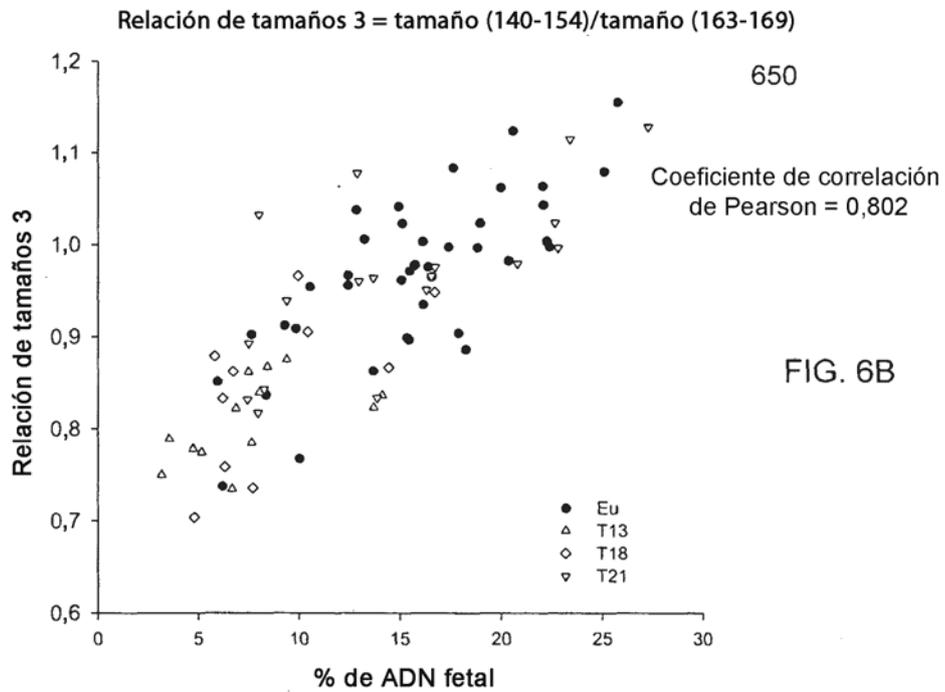
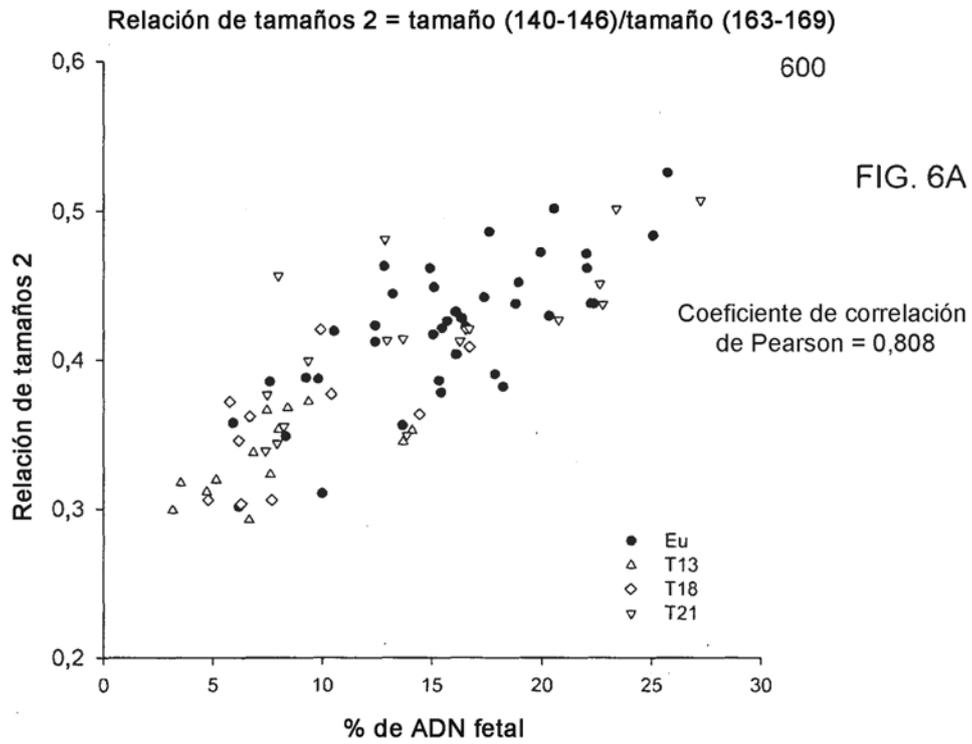


FIG. 4





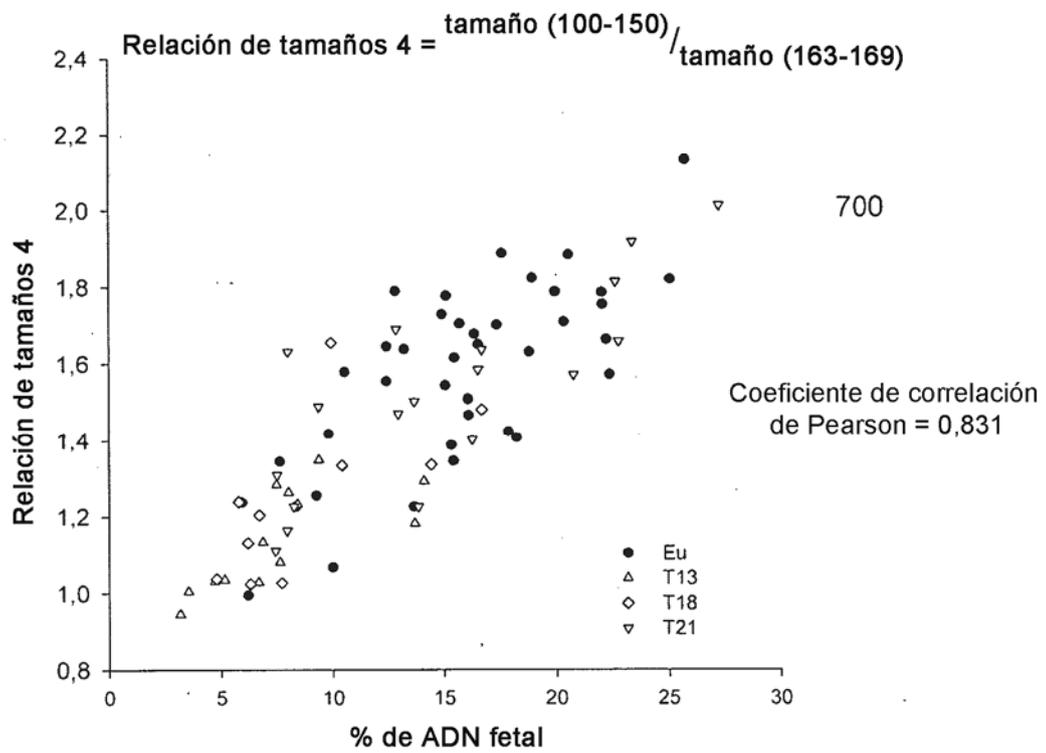


FIG. 7

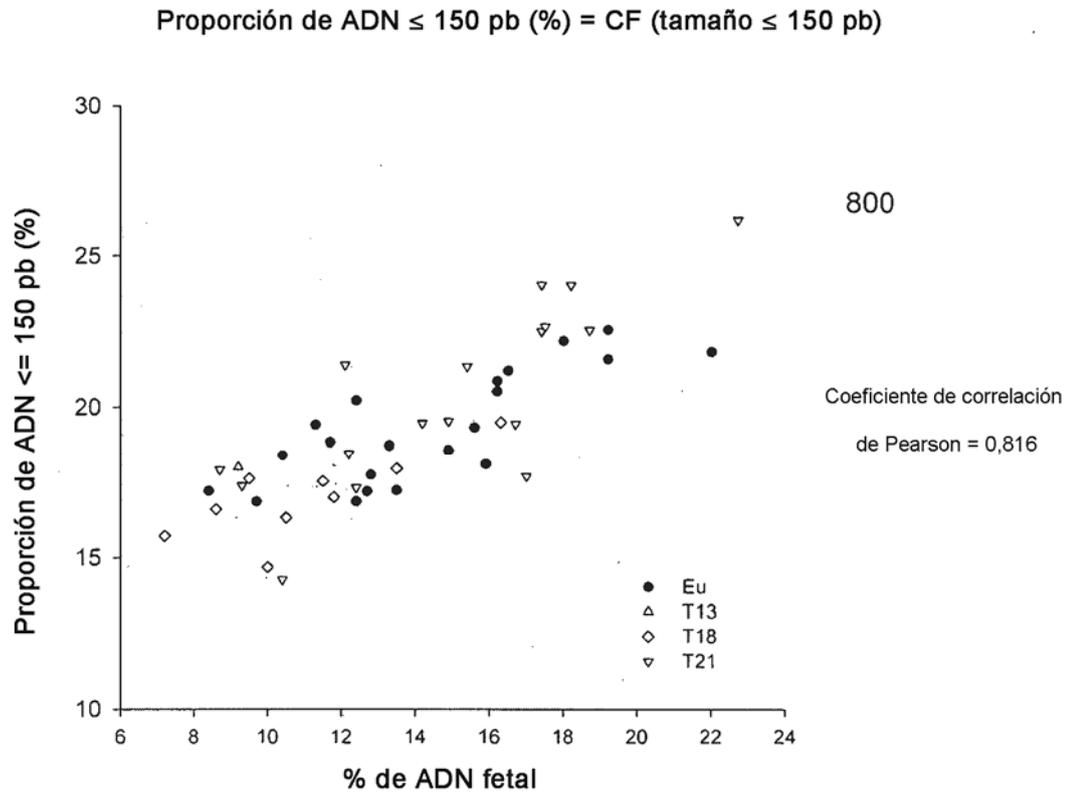
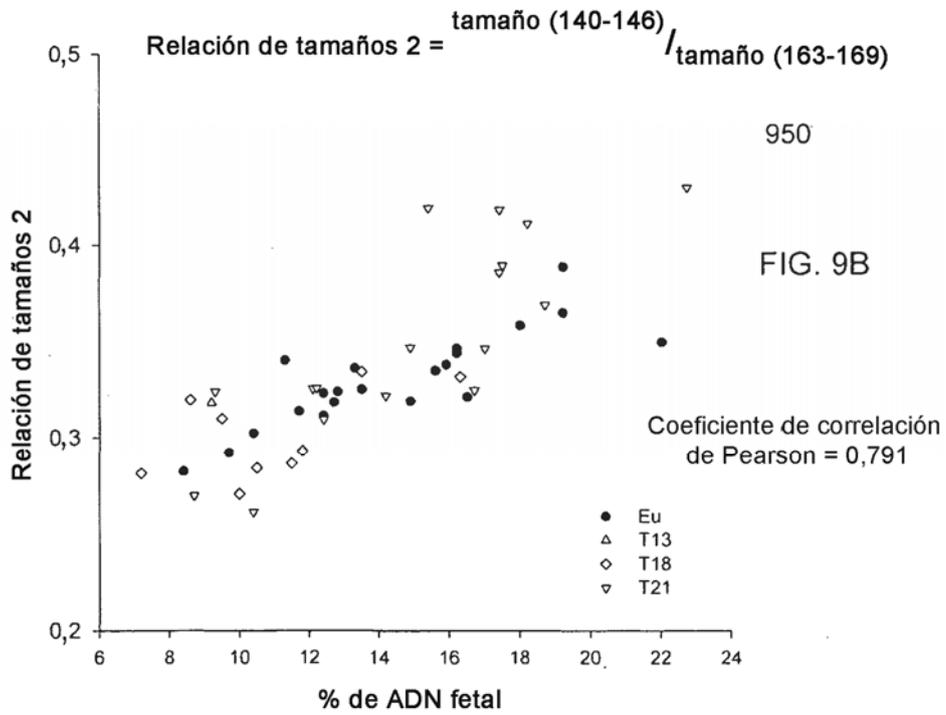
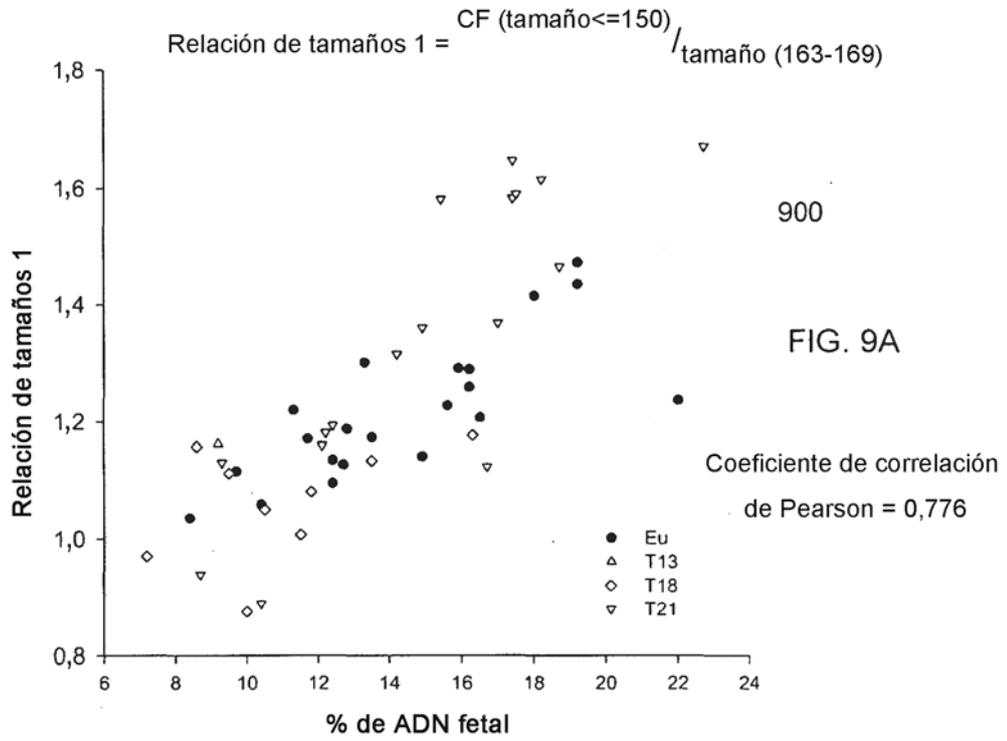
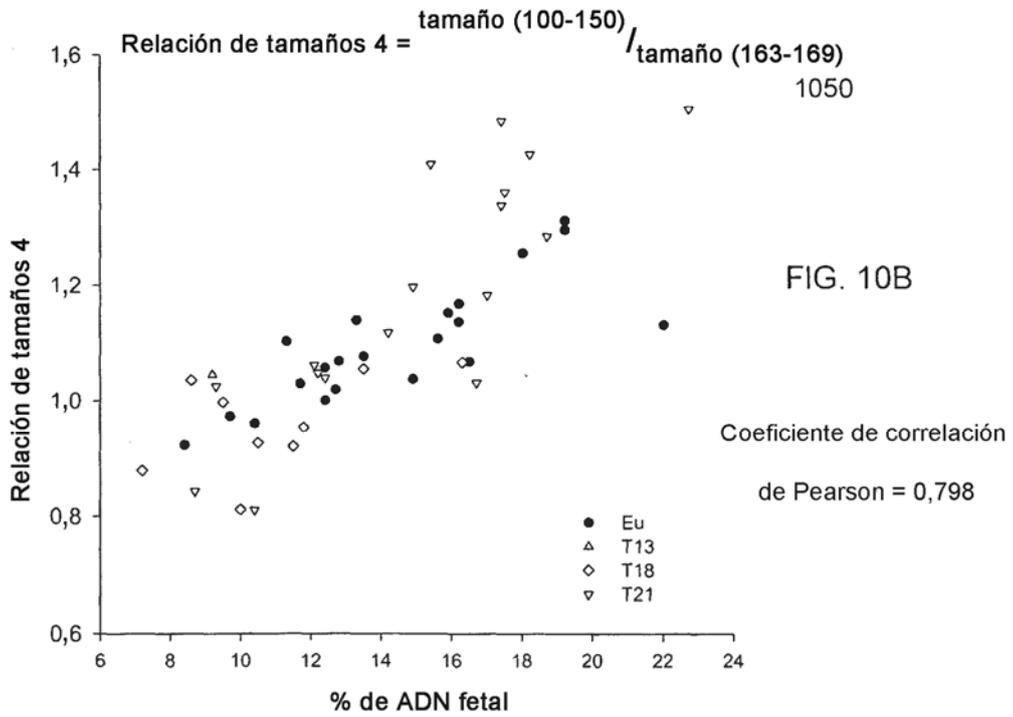
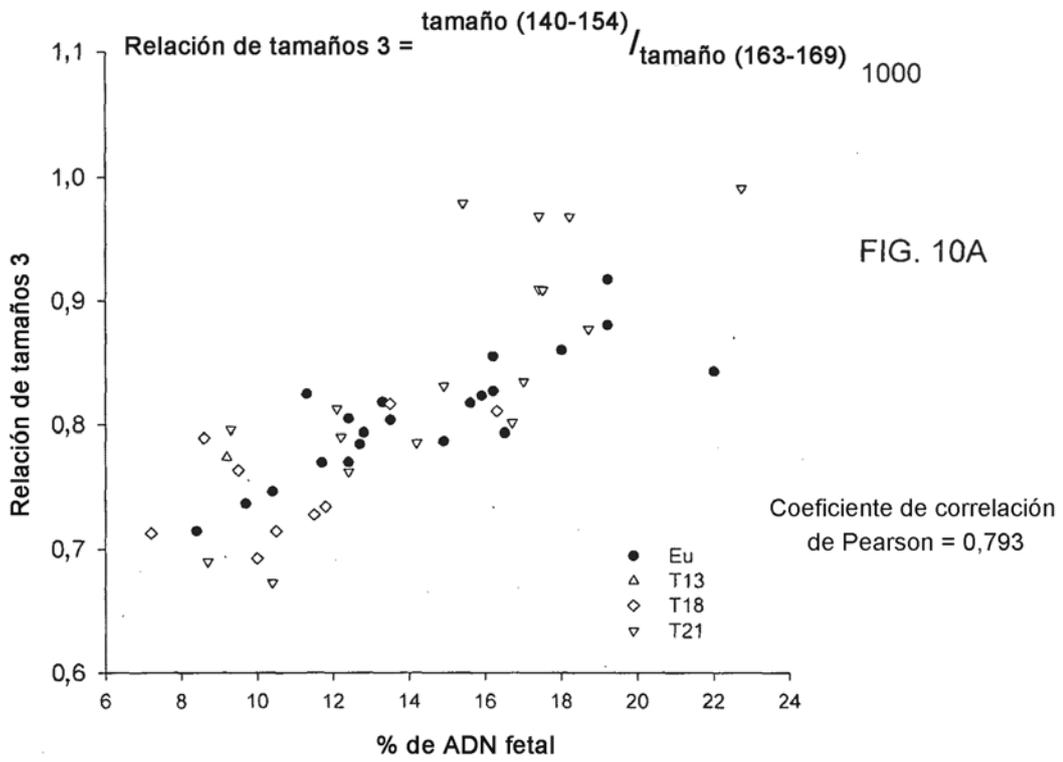


FIG. 8





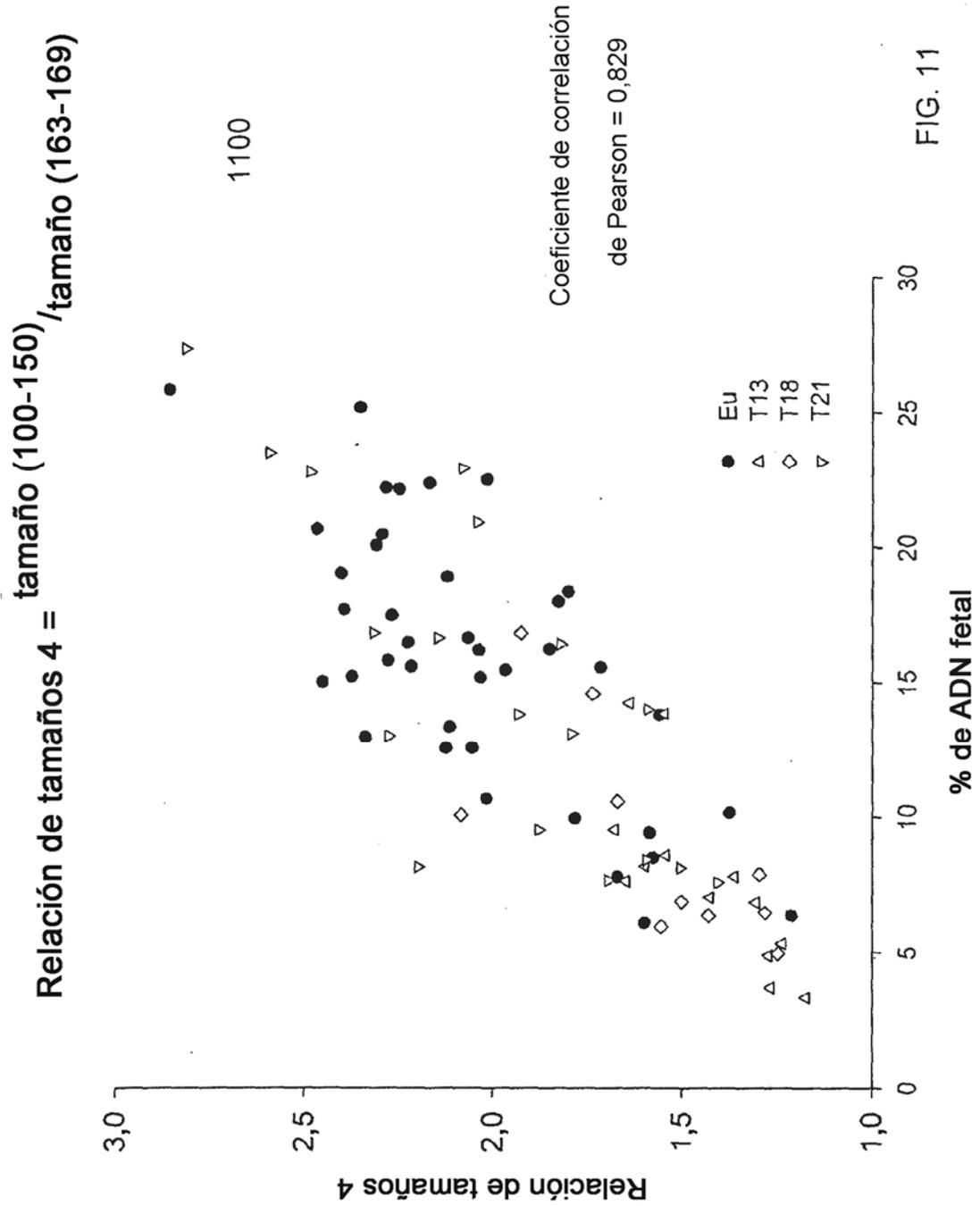


FIG. 11

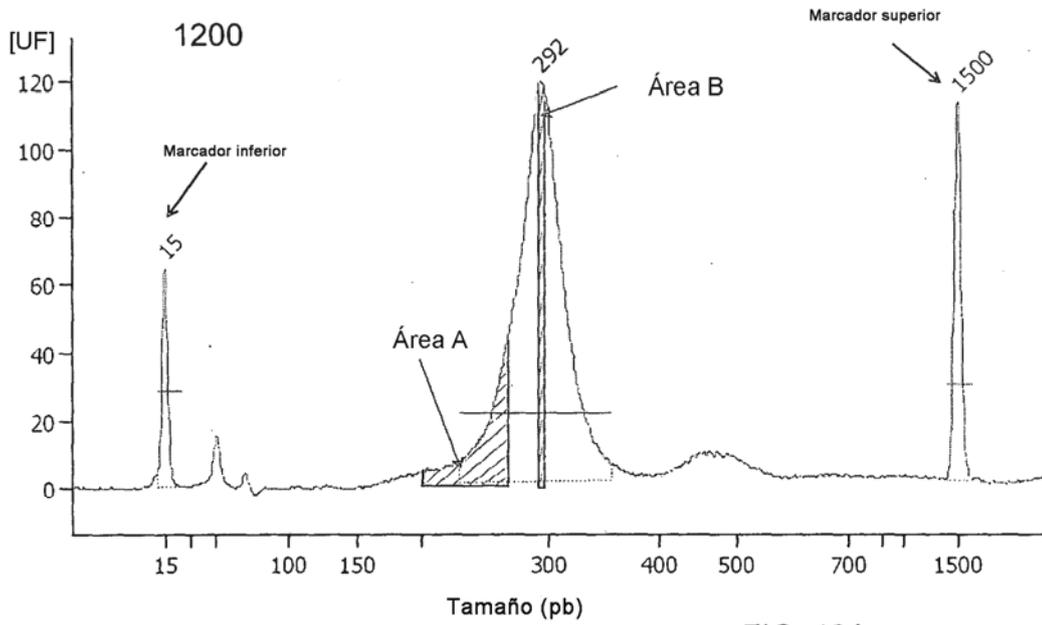


FIG. 12A

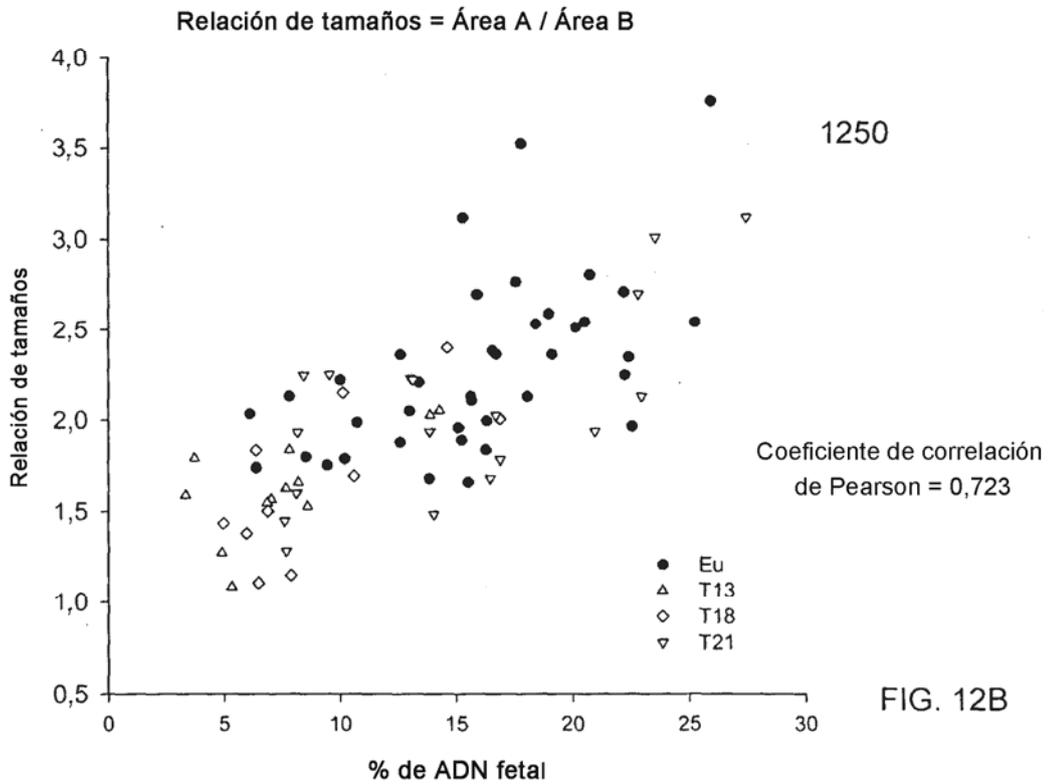


FIG. 12B

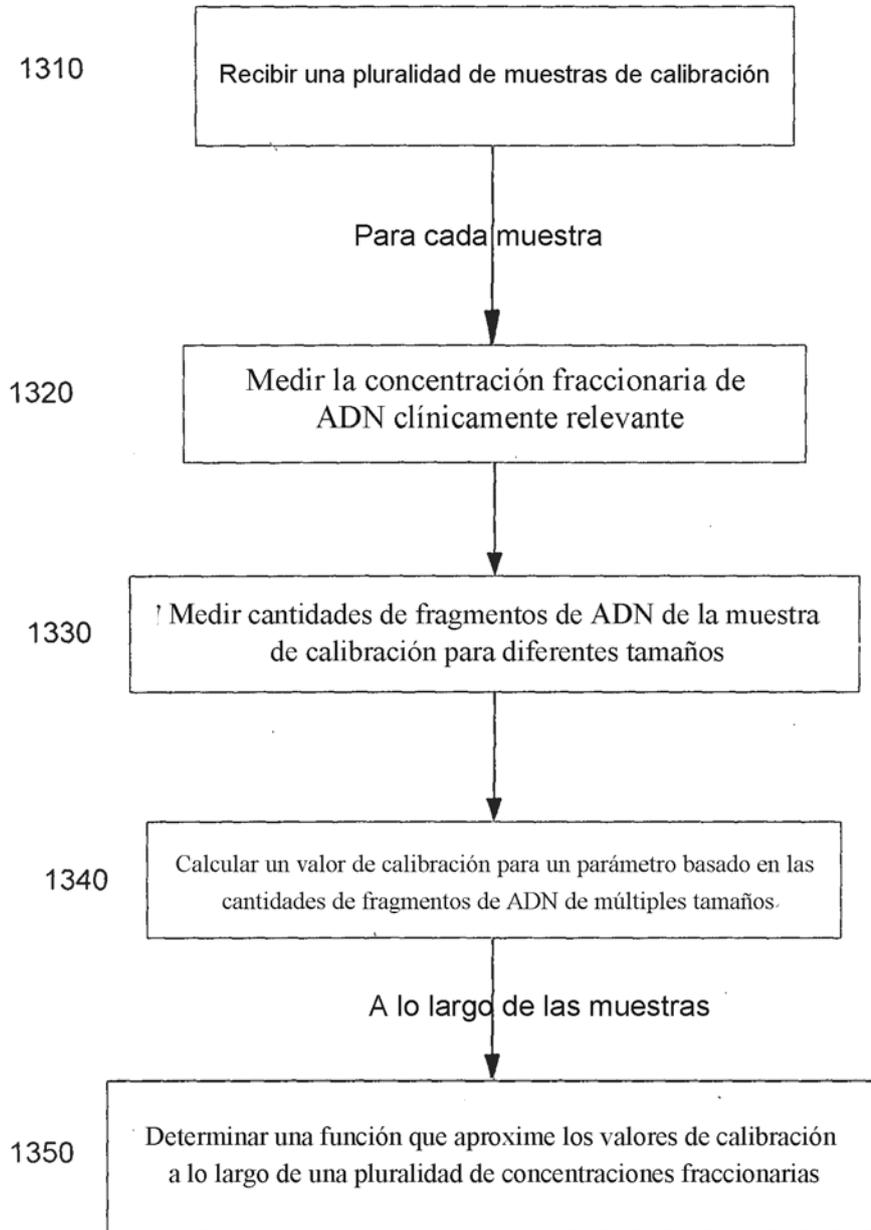
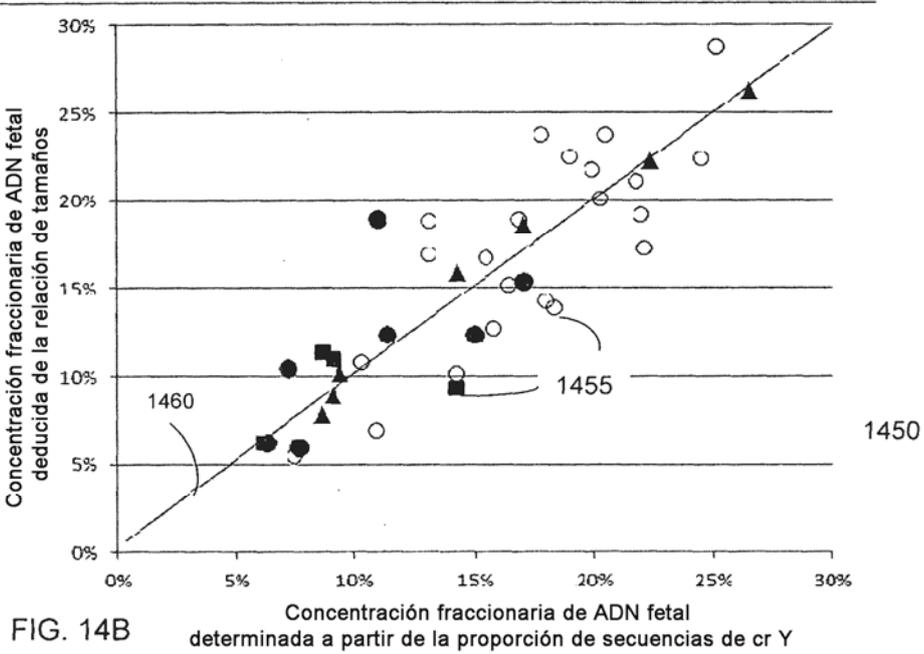
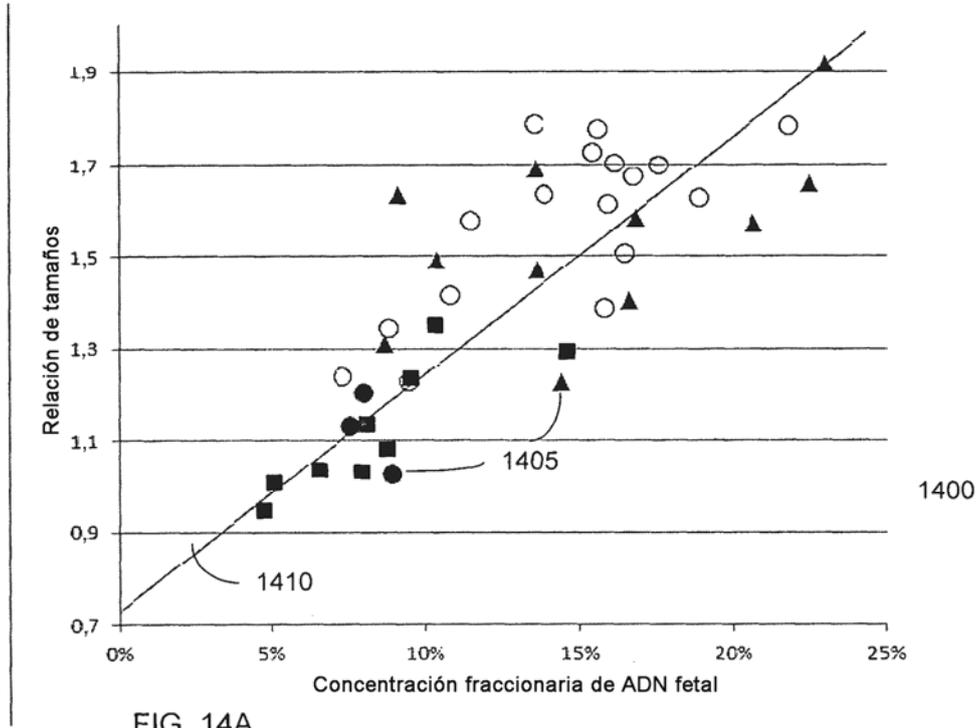


FIG. 13



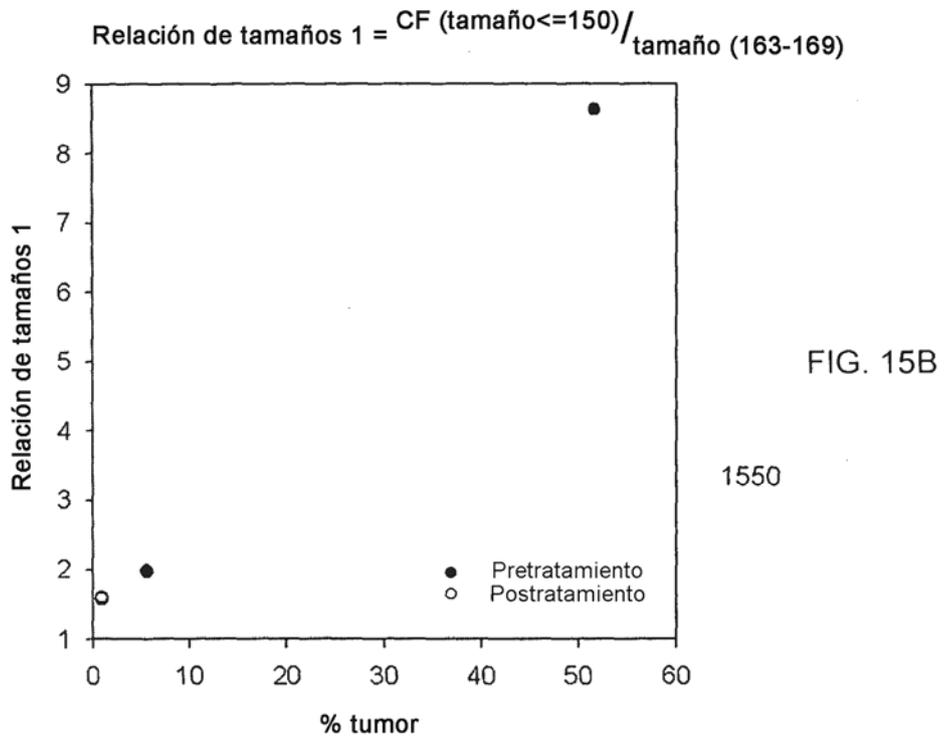
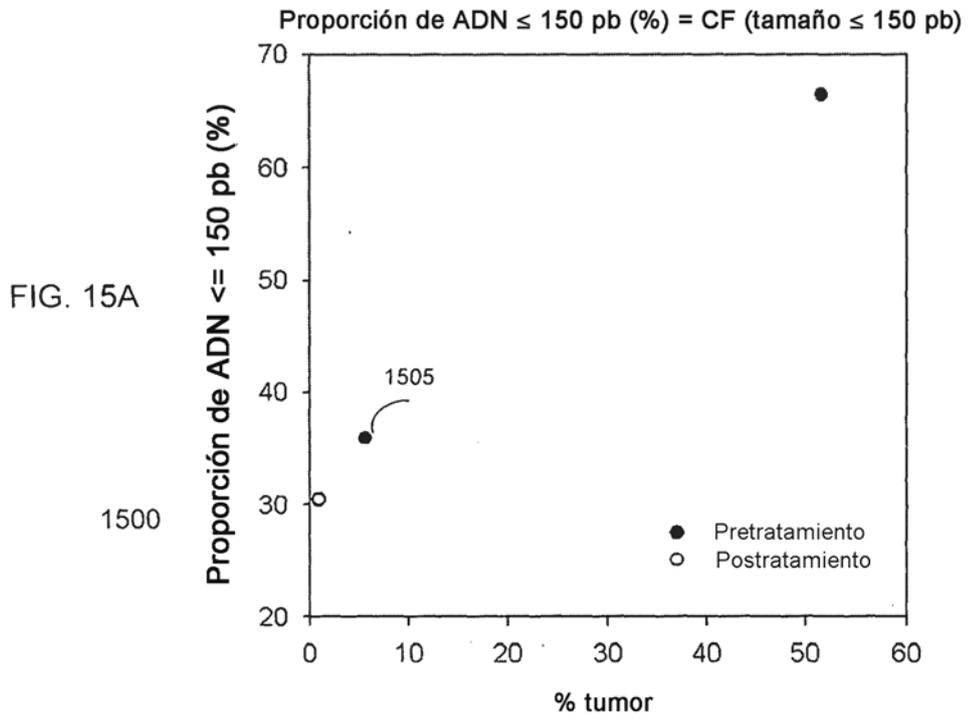
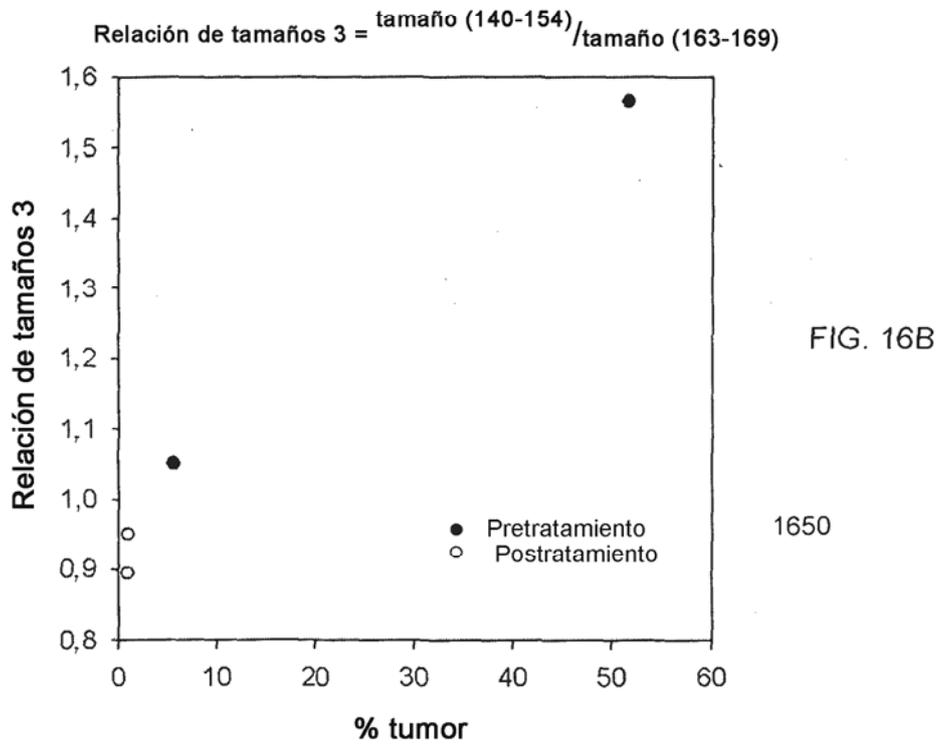
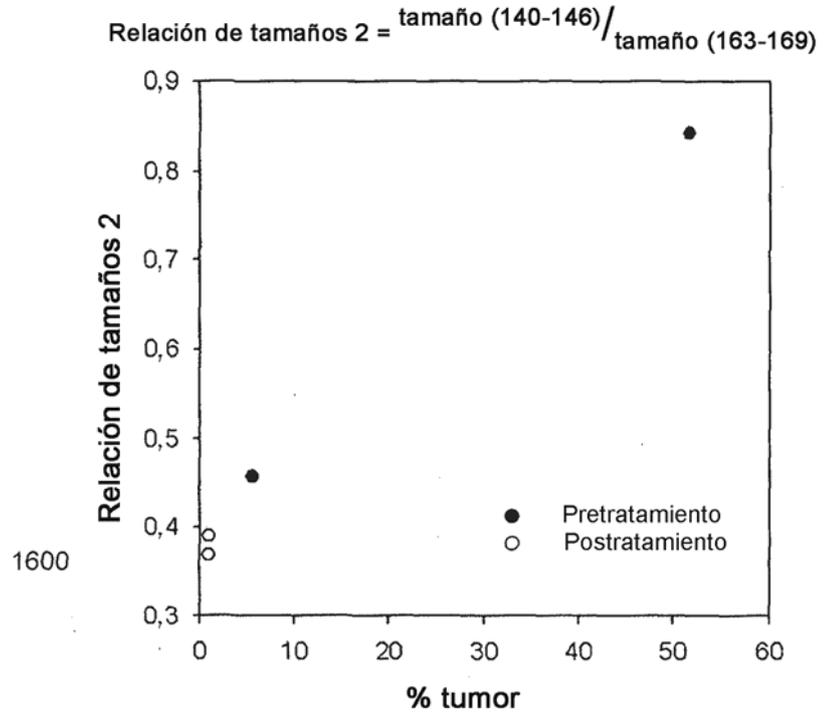


FIG. 16A



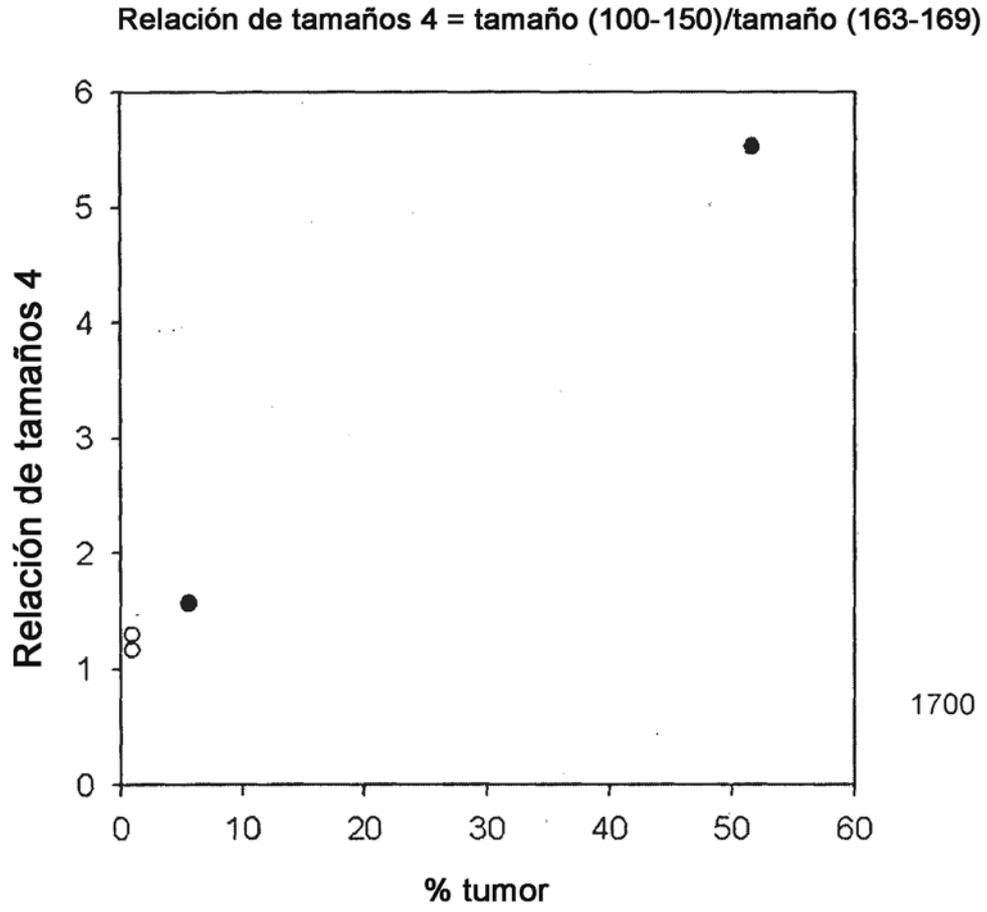


FIG. 17

FIG. 18A

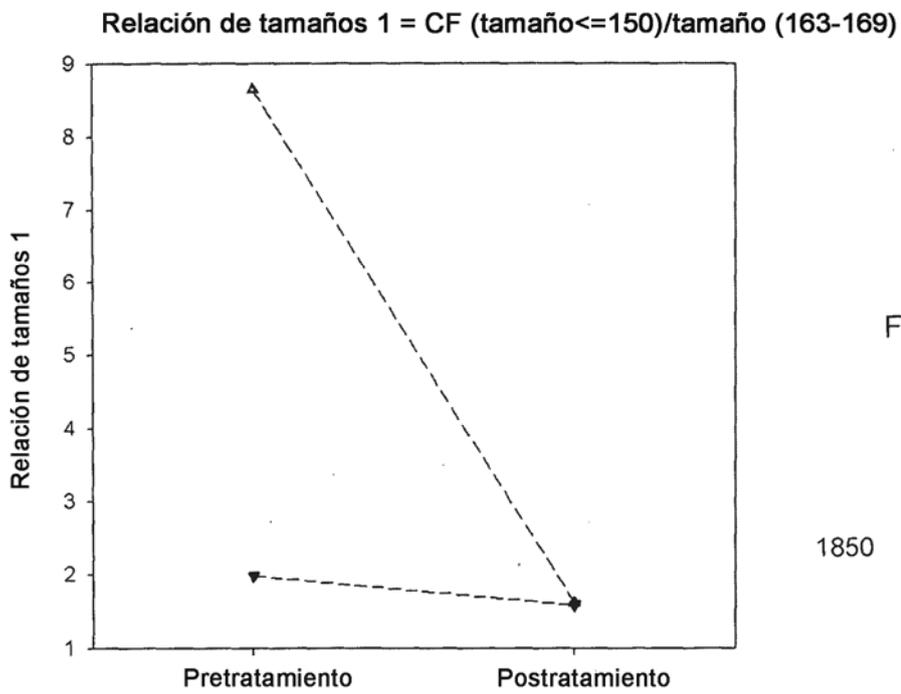
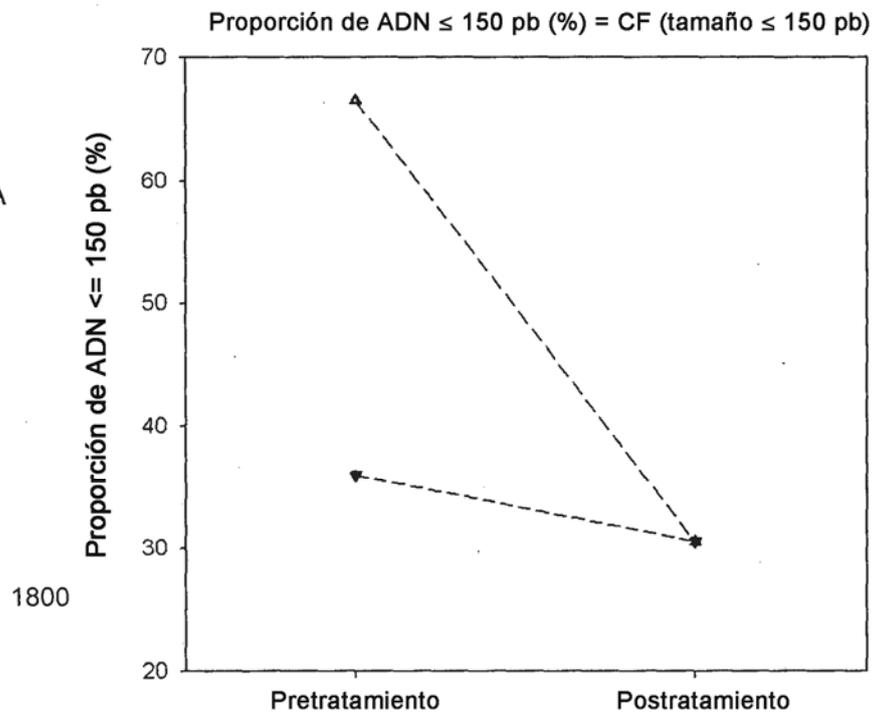
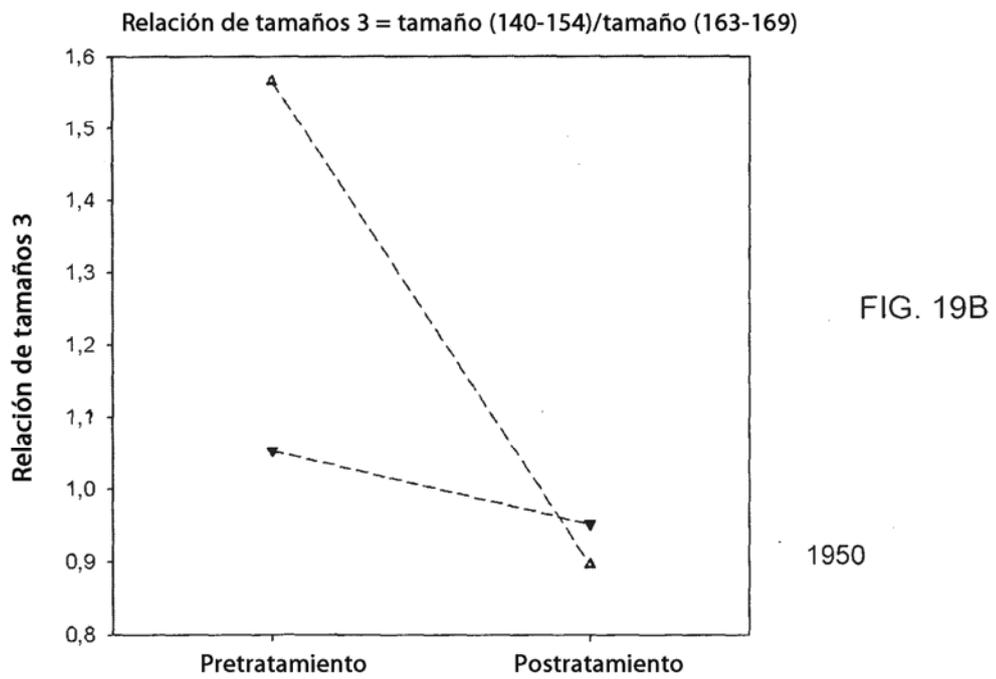
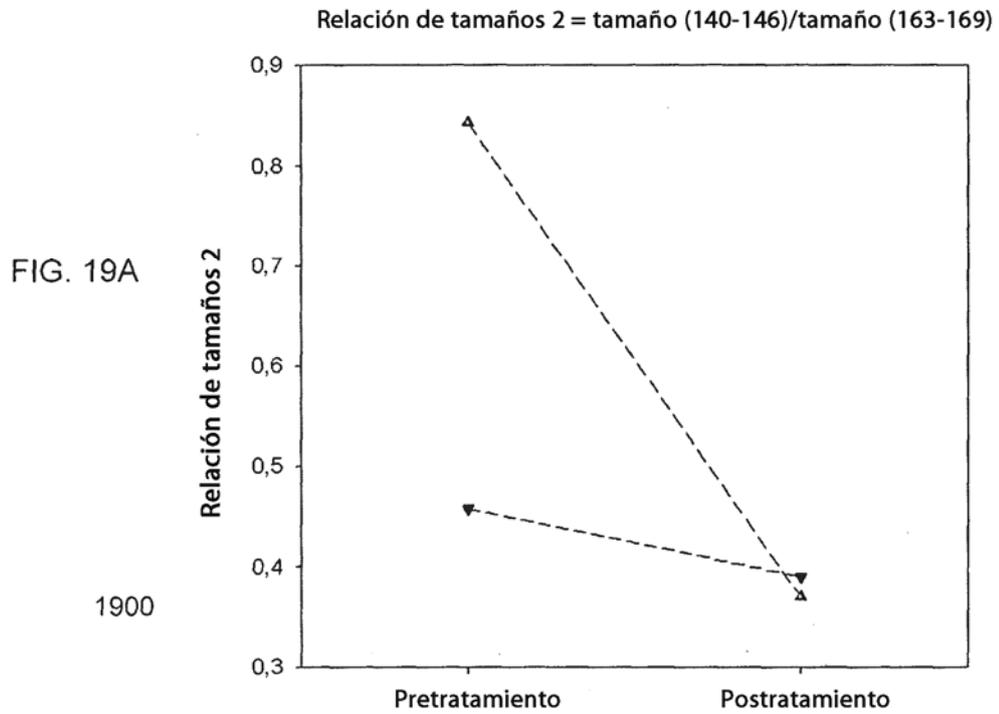


FIG. 18B

1850



$$\text{Relación de tamaños 4} = \frac{\text{tamaño (100-150)}}{\text{tamaño (163-169)}}$$

2000

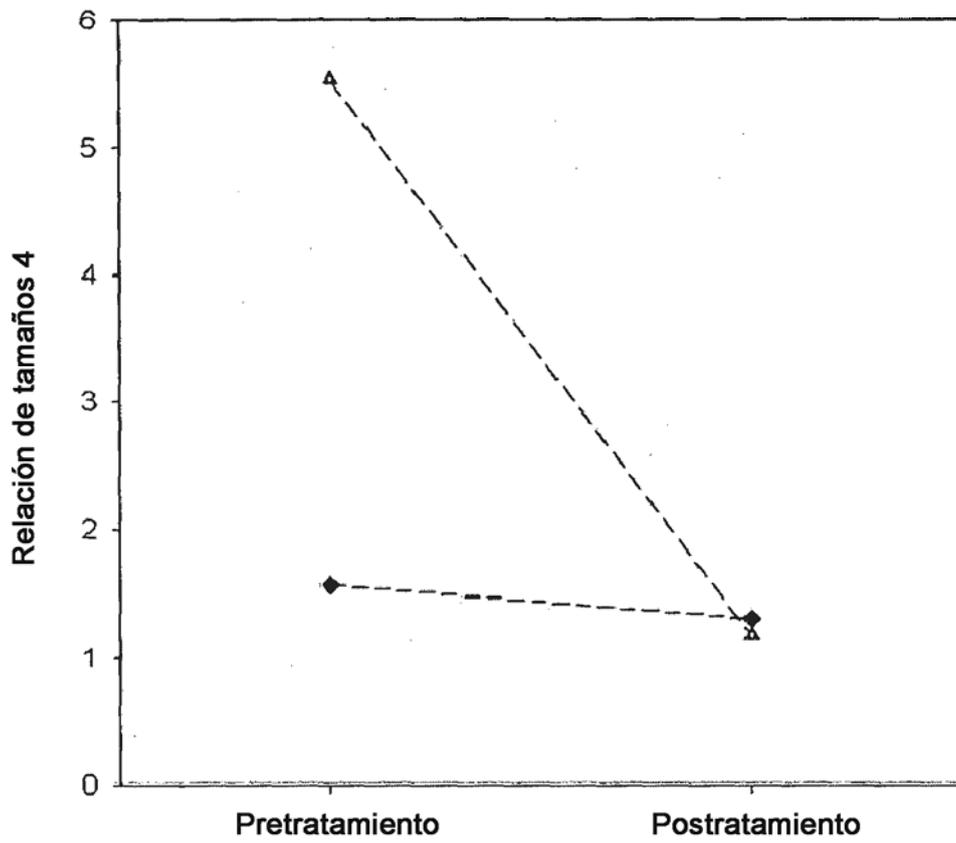


FIG. 20

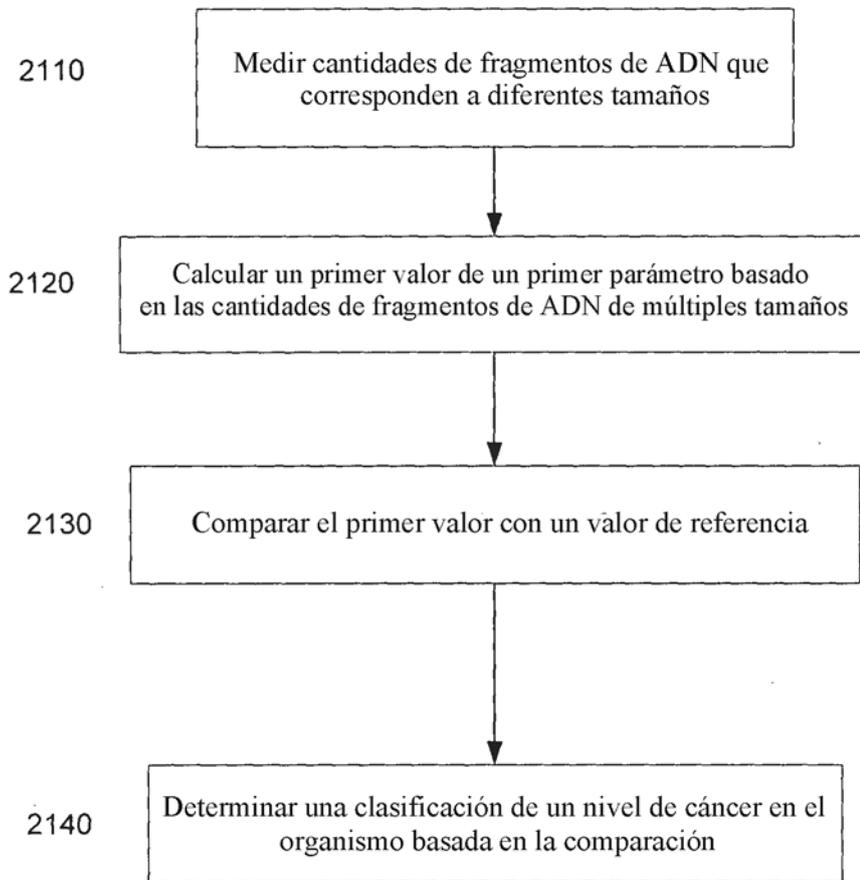


FIG. 21

	Ganancia	Pérdida	Referencia
Colorrectal	7p, 7q, 8q, 11q, 13q, y 20q,	5q, 8p, 17p, 18p, 18q y 20p,	(Nakao et al. Carcinogenesis 2004;25: 1345-1357.) (Tsafrir et al. Cancer Res 2006; 66: 2129-2137)
Mama	1q, 6p, 8q, 11q, 16p, 17q, 19, y 20q	6q, 13q, 16q, 17p, y 22q	(Tirkkonen et al. Gene Chromosome Canc 1998; 21: 177-184) (Richard et al. Int J Cancer 2000; 89: 305-310) (Pinkel et al. Nat Genet 1998; 20: 207-211) (Persson et al. Gene Chromosome Canc 1999; 25: 115-122) (Nishizaki et al. Int J Cancer 1997; 74: 513 - 517)
Pulmón	1q, 3q, 5p, y 8q	3p, 6q, 8p, 9p, 13q, y 17p	(Berrieman et al. Brit J Cancer 2004; 90: 900-905) (Luk et al. Cancer Genet Cytogen 2001; 125: 87 - 99) (Petersen et al. Cancer Res 1997; 57: 2331-2335) (Pei et al. Gene Chromosome Canc 2001; 31: 282-287)
HCC	1q, 8q, 17q y 20q	4q, 6q, 8p, 13q, 16q y 17p	(Kusano et al. Cancer 2002; 94: 746-751) (Laurent-Puig et al. Gastroenterology 2001; 120: 1763-1773) (Moinzadeh et al. Brit J Cancer 2005; 92: 935-941)
Ovario	20q, 3q, 1q, 8q, 12p, 11q, y 17q	Xp, 18q, 4q, 9p, y 13q	(Taetle et al. Gene Chromosome Canc 1999; 25: 290-300) (Schraml et al. Am J Pathol 2003; 163: 985 - 992) (Sonoda et al. Gene Chromosome Canc 1997; 20: 320-328)

FIG. 22

2200

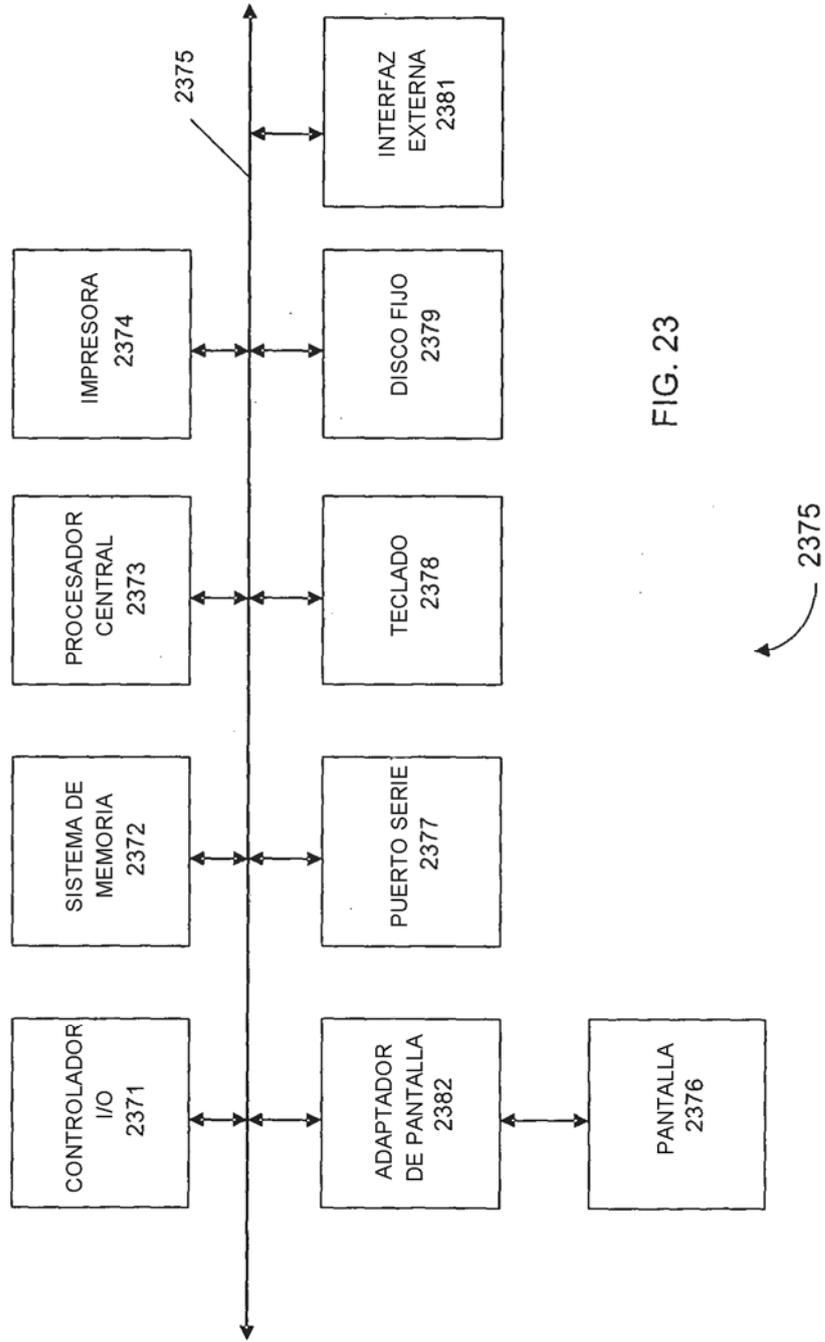


FIG. 23