

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 960 201**

51 Int. Cl.:

G16B 20/10 (2009.01)

G16B 25/10 (2009.01)

C12Q 1/6869 (2008.01)

C12Q 1/6883 (2008.01)

C12Q 1/6827 (2008.01)

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.07.2016** E 21196292 (3)

97 Fecha y número de publicación de la concesión europea: **23.08.2023** EP 3967775

54 Título: **Análisis de los patrones de fragmentación del ADN acelular**

30 Prioridad:

23.07.2015 US 201562196250 P

12.02.2016 US 201662294948 P

14.02.2016 WO PCT/CN2016/073753

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

01.03.2024

73 Titular/es:

THE CHINESE UNIVERSITY OF HONG KONG
(100.0%)

**Office of Research and Knowledge Transfer
Services, Room 301, Pi Ch'iu Building
Shatin, New Territories 999077, HK**

72 Inventor/es:

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

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 960 201 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 los patrones de fragmentación del ADN acelular

5 En estudios previos, se demostró que el ADN plasmático consta principalmente de fragmentos cortos de menos de 200 pb (Lo *et al.*, Sci Transl Med 2010; 2(61):61ra91). En la distribución por tamaño molecular del ADN plasmático, se pudo observar un pico a 166 pb. Además, se observó que la densidad de marcadores secuenciados podría variar con una periodicidad de aproximadamente 180 pb cerca de los sitios de inicio transcripcional (TSS) cuando se secuenció el ADN plasmático materno (Fan *et al.*, PNAS 2008;105:16266-71). Estos resultados son un conjunto de pruebas que
10 indican que la fragmentación del ADN plasmático puede no ser un proceso aleatorio. Sin embargo, aún no se han aclarado los patrones concretos de fragmentación del ADN en el plasma, así como los factores que determinan los patrones. Además, aún no se han evidenciado las aplicaciones prácticas del uso de la fragmentación del ADN.

15 De acuerdo con la presente invención, se proporciona un método de análisis de una muestra biológica como se define en la reivindicación 1.

Diversas realizaciones se refieren a aplicaciones (por ejemplo, aplicaciones de diagnóstico) del análisis de los patrones de fragmentación del ADN acelular, por ejemplo, ADN plasmático y ADN sérico. Las realizaciones de una aplicación pueden determinar una clasificación de una contribución proporcional de un tipo de tejido concreto en una mezcla de
20 ADN acelular de diferentes tipos de tejido. Por ejemplo, pueden determinarse a modo de clasificación los porcentajes específicos, el intervalo de porcentajes o si la contribución proporcional se encuentra por encima de un porcentaje especificado. En un ejemplo, pueden identificarse las posiciones finales preferidas para el tipo de tejido particular y puede usarse la abundancia relativa de moléculas de ADN acelular que terminan en las posiciones finales preferidas para proporcionar la clasificación de la contribución proporcional.

25 Las realizaciones de otra aplicación pueden determinar un genotipo de un tipo de tejido particular en una mezcla de ADN acelular de diferentes tipos de tejido. En un ejemplo, pueden identificarse las posiciones finales preferidas para el tipo de tejido particular y puede determinarse el genotipo utilizando moléculas de ADN acelular que terminan en las posiciones finales preferidas.

30 Las realizaciones de otra aplicación pueden identificar posiciones finales preferidas mediante la comparación de un máximo local para los extremos izquierdos de moléculas de ADN acelular con un máximo local para los extremos derechos de moléculas de ADN acelular. Las posiciones finales preferidas pueden identificarse cuando los máximos locales correspondientes están suficientemente separados. Además, pueden compararse las cantidades de moléculas de ADN acelular que terminan en un máximo local para el extremo izquierdo/derecho con una cantidad de moléculas de ADN acelular para un máximo local con baja separación, a fin de determinar una contribución proporcional de un tipo de tejido.

35 Otras realizaciones se refieren a sistemas, dispositivos portátiles de consumo, y medios legibles por ordenador asociados a los métodos descritos en el presente documento.

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

45 La FIG. 1 muestra un ejemplo ilustrativo para la definición de la probabilidad de intacto (P_i) de acuerdo con las realizaciones de la presente invención.

50 Las FIG. 2A y 2B muestran la variación de la P_i a lo largo de un segmento del cromosoma 6 utilizando 25 como valor de z , de acuerdo con las realizaciones de la presente invención.

La FIG. 3 muestra la ilustración de la variación sincrónica de la P_i para el ADN de origen materno y fetal en el plasma materno.

55 La FIG. 4 muestra una ilustración de la variación asincrónica de la P_i para el ADN de origen materno y fetal en el plasma materno.

La FIG. 5 es un diagrama de flujo que muestra un análisis sobre si las moléculas de ADN materno y fetal son sincrónicas en la variación de la P_i .

60 La FIG. 6 muestra un análisis de dos muestras de plasma materno (S24 y S26) para determinar la variación de la P_i para fragmentos de ADN de origen materno (rojo/gris) y fetal (azul/negro) en el plasma materno.

La FIG. 7 muestra una ilustración de la amplitud de la variación de la P_i .

65 La FIG. 8A muestra los patrones de variación de la P_i en regiones que son sitios de hipersensibilidad a DNasa, pero no TSS.

- La FIG. 8B muestra patrones de variación de la P_i en regiones que son TSS, pero no sitios de hipersensibilidad a DNasa.
- 5 La FIG. 9 muestra una ilustración del principio para la medición de la proporción de ADN liberado de diferentes tejidos.
- La FIG. 10 muestra la relación entre la FRA y la contribución proporcional del tejido A al ADN en una mezcla determinada mediante el análisis de dos o más muestras de calibración con concentraciones proporcionales conocidas de ADN del tejido A.
- 10 La FIG. 11 muestra una correlación entre la $FR_{placenta}$ y el porcentaje de ADN fetal en el plasma materno.
- La FIG. 12 muestra una correlación entre el FR_{sangre} y la concentración de ADN fetal en el plasma materno.
- 15 La FIG. 13 es un diagrama de flujo de un método 1300 de análisis de una muestra biológica para determinar una clasificación de una contribución proporcional del primer tipo de tejido de acuerdo con las realizaciones de la presente invención.
- La FIG. 14 muestra una ilustración del principio de una diferencia de los fragmentos de ADN circulante en relación con el ADN de origen tumoral o fetal.
- 20 La FIG. 15 es un diagrama de flujo de un método de análisis de una muestra biológica que incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido.
- 25 La FIG. 16 es un diagrama de Venn que muestra el número de sitios de terminación frecuentes que son específicos para el caso HCC, específicos para la mujer embarazada y compartidos por ambos casos.
- La FIG. 17 muestra una curva de calibración que muestra la relación entre la proporción de fragmentos de ADN secuenciados que terminan en posiciones finales específicas para el cáncer y la fracción de ADN tumoral en plasma para pacientes con cáncer con fracciones conocidas de ADN tumoral en plasma.
- 30 La FIG. 18 muestra un ejemplo ilustrativo de los patrones de fragmentación no aleatorios del ADN plasmático que portan un alelo específico del feto y un alelo compartido por la madre y el feto.
- 35 La FIG. 19 muestra una gráfica de la probabilidad de que una coordenada genómica sea una posición final de los fragmentos de ADN plasmático materno en una región con un polimorfismo de un solo nucleótido (SNP) informativo.
- La FIG. 20 muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático a lo largo de los SNP que eran homocigóticos en la madre y heterocigóticos en el feto.
- 40 La FIG. 21 muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático a lo largo de los SNP que eran homocigóticos en el feto y heterocigóticos en la madre.
- 45 La FIG. 22 muestra una correlación entre la abundancia relativa (Ratio (F/M)) de las moléculas de ADN plasmático con extremos fetales (Conjunto A) y maternos (Conjunto X) recurrentes y la fracción de ADN fetal.
- Las FIG. 23A-23E muestran datos relativos a las distribuciones del tamaño del ADN plasmático para los fragmentos que terminan en las posiciones finales con preferencia fetal y los fragmentos que terminan en las posiciones finales con preferencia materna.
- 50 Las FIG. 24A-24E muestran datos relativos a las distribuciones de tamaño del ADN plasmático en una muestra de ADN plasmático conjunta de 26 mujeres en el primer trimestre del embarazo para los fragmentos que terminan en las posiciones finales con preferencia fetal y los fragmentos que terminan en las posiciones finales con preferencia materna.
- 55 La FIG. 25 muestra un ejemplo ilustrativo de los patrones de fragmentación no aleatorios del ADN plasmático del paciente con HCC.
- 60 La FIG. 26 es una gráfica de la probabilidad de que una coordenada genómica sea una posición final de los fragmentos de ADN plasmático en una región con un sitio de mutación.
- La FIG. 27A muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático en las posiciones genómicas en las que había mutaciones en el tejido tumoral.
- 65

- La FIG. 27B muestra una correlación entre la Relación_{M/TS} y la fracción de ADN tumoral en el plasma de 71 pacientes con HCC.
- 5 La FIG. 28A muestra el número de posiciones finales preferidas para el ADN plasmático de la mujer embarazada y del paciente con HCC. El conjunto P contenía 29 millones de posiciones finales que eran las preferidas en la mujer embarazada.
- 10 La FIG. 28B muestra que se observó una correlación positiva entre la Relación_{HCC/Emb} y la fracción de ADN tumoral en plasma para los 71 pacientes con HCC.
- La FIG. 29A muestra una ilustración del concepto de relación de terminación preferida (PETR). Cada línea representa un fragmento de ADN plasmático.
- 15 La FIG. 29B muestra una correlación entre la fracción de ADN tumoral en plasma con la PETR en las posiciones del conjunto H en 11 pacientes con HCC.
- La FIG. 30 muestra la proporción de ADN corto (<150 pb) detectado entre las moléculas de ADN plasmático que terminan con extremos preferidos por el HCC, extremos preferidos por el VHB o los extremos compartidos.
- 20 La FIG. 31A muestra una ilustración del principio de la w-PETR. El valor de w-PETR se calcula como la relación entre el número de fragmentos de ADN que terminan dentro de la ventana A y la ventana B.
- La FIG. 31B muestra una correlación entre la fracción de ADN tumoral y el valor de la w-PETR en los 11 pacientes con HCC.
- 25 La FIG. 32 muestra la proporción de posiciones finales preferidas comúnmente compartidas detectadas en muestras de plasma de cada una de las muestras estudiadas cuando se comparan con una muestra de plasma sanguíneo del cordón umbilical (cobertura del genoma haploide 210x).
- 30 La FIG. 33 muestra un diagrama de Venn en el que se indica el número de posiciones finales preferidas que se observan comúnmente en dos o más muestras, así como las que solo se observaron en una muestra.
- La FIG. 34A muestra una correlación entre la fracción de ADN fetal en el plasma y la PETR media en el conjunto de posiciones identificadas mediante la comparación entre las muestras de ADN plasmático "antes del parto" y "después del parto".
- 35 La FIG. 34B muestra una correlación entre la fracción de ADN fetal en el plasma y el promedio de w-PETR en el conjunto de posiciones identificadas a través de la comparación entre las muestras de ADN plasmático "antes del parto" y "después del parto".
- 40 La FIG. 35A muestra el millón de posiciones finales preferidas del ADN plasmático observadas con mayor frecuencia entre dos mujeres embarazadas a las 18 semanas (mujer embarazada 1) y a las 38 semanas de gestación (mujer embarazada 2).
- 45 La FIG. 35B muestra una comparación de los valores de PETR del millón de posiciones finales preferidas más frecuentemente observadas en el plasma de dos mujeres embarazadas.
- La FIG. 36 es un diagrama de flujo de un método de análisis de una muestra biológica para determinar una clasificación de una contribución proporcional del primer tipo de tejido en una mezcla de acuerdo con las realizaciones de la presente invención.
- 50 La FIG. 37 muestra las moléculas de ADN plasmático materno que portan diferentes alelos cuando se alinean con un genoma de referencia cerca de una posición final preferida por el feto.
- 55 La FIG. 38 es un diagrama de flujo de un método 3800 de análisis de una muestra biológica para determinar un genotipo del primer tipo de tejido de acuerdo con las realizaciones de la presente invención.
- 60 La FIG. 39 muestra un diagrama de bloques de un sistema informático 10 ilustrativo que puede utilizarse con el sistema y los métodos de acuerdo con las realizaciones de la presente invención.
- 65 Un "tejido" corresponde a un grupo de células que se agrupan como una unidad funcional. En un mismo tejido pueden encontrarse más de un tipo de células. Los distintos tipos de tejidos pueden estar formados por diferentes tipos de células (por ejemplo, hepatocitos, células alveolares o células sanguíneas), pero también puede corresponder a tejidos de organismos diferentes (la madre en comparación con feto) o a células sanas en comparación con células tumorales.
- Una "*muestra biológica*" se refiere a cualquier muestra que se toma de un sujeto (por ejemplo, un ser humano, tal

como una mujer embarazada, una persona con cáncer o una persona que se sospecha que tiene cáncer, un receptor de un trasplante de órganos o un sujeto que se sospecha que padece una enfermedad que afecta a un órgano (por ejemplo, el corazón en el infarto de miocardio, o el cerebro en el ictus, o el sistema hematopoyético en la anemia) y que contiene una o más moléculas de ácido nucleico de interés. La muestra biológica puede ser un tejido corporal, tal como sangre, plasma, suero, orina, fluido vaginal, líquido de un hidrocele (por ejemplo, de los testículos), líquidos de un lavado vaginal, líquido pleural, líquido ascítico, líquido cefalorraquídeo, saliva, sudor, lágrimas, esputo, líquido de lavado broncoalveolar, líquido de descarga del pezón, líquido de aspiración de diferentes partes del cuerpo (por ejemplo, tiroides, mama), etc. También pueden utilizarse muestras de heces. En diversas realizaciones, la mayor parte del ADN de una muestra biológica enriquecida en ADN acelular (por ejemplo, una muestra de plasma obtenida mediante un protocolo de centrifugación) puede estar libre de células, por ejemplo, más del 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o 99 % del ADN puede estar libre de células. El protocolo de centrifugación puede incluir, por ejemplo, 3.000 g x 10 minutos, obtener la parte líquida, y volver a centrifugar, por ejemplo, 30.000 g durante otros 10 minutos para eliminar las células residuales.

Los "*cambios asociados al cáncer*" o "*cambios específicos del cáncer*" incluyen, pero sin limitación, mutaciones derivadas del cáncer (incluidas mutaciones de un solo nucleótido, deleciones o inserciones de nucleótidos, deleciones de segmentos genéticos o cromosómicos, translocaciones, inversiones), amplificación de genes, segmentos genéticos o cromosómicos, secuencias asociadas a virus (por ejemplo, episomas víricos e inserciones víricas), perfiles de metilación aberrantes o firmas de metilación específicas del tumor, perfiles de tamaño del ADN acelular aberrantes, marcas aberrantes de modificación de histonas y otras modificaciones epigenéticas, y ubicaciones de los extremos de los fragmentos de ADN acelular que están asociados al cáncer o son específicos del cáncer.

Un "*fragmento de ADN de cáncer informativo*" corresponde a un fragmento de ADN portador de uno o más de los cambios o mutaciones asociados al cáncer o específicos del cáncer. Un "*fragmento de ADN fetal informativo*" corresponde a un fragmento de ADN fetal portador de una mutación que no se encuentra en ninguno de los genomas de los progenitores. Un "*fragmento de ADN informativo*" puede referirse a cualquiera de los tipos de fragmentos de ADN mencionados.

Una "*lectura de secuencia*" se refiere a una cadena de nucleótidos secuenciada de cualquier parte o de una molécula completa de ácido nucleico. Por ejemplo, una lectura de secuencia puede ser una cadena corta de nucleótidos (por ejemplo, 20-150) secuenciados a partir de un fragmento de ácido nucleico, una cadena corta de nucleótidos en uno o ambos extremos de un fragmento de ácido nucleico, o la secuenciación de todo el fragmento de ácido nucleico que existe en la muestra biológica. Una lectura de secuencia puede obtenerse de diversos modos, por ejemplo, mediante técnicas de secuenciación o utilizando sondas, por ejemplo, en matrices de hibridación o sondas de captura, o técnicas de amplificación, como la reacción en cadena de la polimerasa (PCR) o la amplificación lineal con un único cebador o la amplificación isotérmica.

Una "*posición final*" (o simplemente "*final*") puede referirse a la coordenada genómica o a la identidad genómica o de nucleótidos de la base más externa, es decir, en los extremos, de una molécula de ADN acelular, por ejemplo, la molécula de ADN plasmático. La posición final puede corresponder a cualquier extremo de una molécula de ADN. De esta manera, si se hace referencia a un inicio y un final de una molécula de ADN, ambos corresponderían a una posición final. En la práctica, una posición final es la coordenada genómica o la identidad de los nucleótidos de la base más externa de un extremo de una molécula de ADN acelular que se detecta o determina mediante un método analítico, como, por ejemplo, secuenciación paralela masiva o secuenciación de última generación, secuenciación de molécula única, protocolos de preparación de bibliotecas de secuenciación de ADN monocatenario o bicatenario, reacción en cadena de la polimerasa (PCR) o micromatriz. Estas técnicas *in vitro* pueden alterar los extremos auténticos *in vivo* de las moléculas de ADN acelulares. Por tanto, cada extremo detectable puede representar el extremo biológicamente auténtico o el extremo es uno o más nucleótidos hacia adentro o uno o más nucleótidos extendidos desde el extremo original de la molécula, por ejemplo, el despunte en 5' y el relleno en 3' de los salientes de las moléculas de ADN bicatenario no despuntadas por el fragmento Klenow. La identidad genómica o la coordenada genómica de la posición final podría derivarse de los resultados de la alineación de las lecturas de la secuencia con un genoma humano de referencia, por ejemplo, hg19. Podría derivarse de un catálogo de índices o códigos que representan las coordenadas originales del genoma humano. Podría referirse a una posición o identidad de nucleótidos en una molécula de ADN acelular que se lee mediante, pero sin limitación, sondas específicas para la diana, minisequenciación, amplificación de ADN.

Un "*extremo preferido*" (o una "*posición final recurrente*") se refiere a un extremo que está más representado o es más prevalente (por ejemplo, medido por una tasa) en una muestra biológica que tiene un estado fisiológico (por ejemplo, embarazo) o patológico (enfermedad) (por ejemplo, cáncer) que una muestra biológica que no tiene dicho estado o que, en diferentes puntos temporales o etapas del mismo estado patológico o fisiológico, por ejemplo, antes o después del tratamiento. Por lo tanto, un extremo preferido tiene una mayor probabilidad de ser detectado en el estado fisiológico o patológico pertinente en comparación con otros estados. El aumento de la probabilidad puede compararse entre el estado patológico y un estado no patológico, por ejemplo, en pacientes con y sin cáncer y cuantificados como cociente de probabilidad o probabilidad relativa. El cociente de probabilidad puede determinarse en función de la probabilidad de detectar al menos un número umbral de extremos preferidos en la muestra analizada o en función de la probabilidad de detectar los extremos preferidos en pacientes con dicha afección que en pacientes sin dicha

afección. Los ejemplos para los umbrales de los cocientes de probabilidad incluyen, entre otros, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,8, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5, 6, 8, 10, 20, 40, 60, 80 y 100. Estos cocientes de probabilidad pueden medirse comparando los valores de abundancia relativa de las muestras con y sin la afección correspondiente. Dado que la probabilidad de detectar un extremo preferido en un estado fisiológico o patológico relevante es mayor, esas posiciones finales preferidas se verían en más de un individuo con ese mismo estado fisiológico o patológico. Con el aumento de la probabilidad, puede detectarse más de una molécula de ADN acelular como terminada en una misma posición final preferida, incluso cuando el número de moléculas de ADN acelular analizadas es mucho menor que el tamaño del genoma. Por tanto, las posiciones finales preferidas o recurrentes también se denominan "posiciones finales frecuentes". En algunas realizaciones, se puede utilizar un umbral cuantitativo para exigir que los extremos se detecten al menos varias veces (por ejemplo, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o 50) dentro de la misma muestra o de la misma alícuota de muestra para ser considerado como un extremo preferente. Un estado fisiológico relevante puede incluir un estado cuando una persona está sana, libre de enfermedad, o libre de una enfermedad de interés. De manera similar, una "ventana final preferida" corresponde a un conjunto contiguo de posiciones finales preferidas.

La "frecuencia" de moléculas de ADN que terminan en una posición se refiere a la frecuencia con la que una molécula de ADN termina en la posición. La frecuencia puede estar basada en un número de moléculas de ADN que terminan en la posición normalizada frente a un número de moléculas de ADN analizadas. Por consiguiente, la frecuencia corresponde a una frecuencia del número de moléculas de ADN que terminan en una posición, y no está relacionado con una periodicidad de posiciones que tienen un máximo local en el número de moléculas de ADN que terminan en la posición.

Una "muestra de calibración" puede corresponder a una muestra biológica cuya fracción de ADN específica del tejido se conoce o se determina mediante un método de calibración, por ejemplo, utilizando un alelo específico para el tejido. Como otro ejemplo, una muestra de calibración puede corresponder a una muestra a partir de la cual se pueden determinar las posiciones finales preferidas. Una muestra de calibración puede utilizarse para ambos fines.

Un "punto de datos de calibración" incluye un "valor de calibración" y una distribución proporcional medida o conocida del ADN de interés (es decir, el ADN de un tipo de tejido concreto). El valor de calibración puede ser una abundancia relativa determinada para una muestra de calibración, para el que se conoce la distribución proporcional del tipo de tejido. Los puntos de datos de calibración pueden definirse de varias maneras, por ejemplo, como puntos individuales o como una función de calibración (también denominada curva de calibración o superficie de calibración). La función de calibración podría derivarse de una transformación matemática adicional de los puntos de datos de calibración.

La expresión "profundidad de secuenciación" se refiere al número de veces que un locus está cubierto por una lectura de secuencia alineada con el locus. El locus puede ser tan pequeño como un nucleótido, o tan grande como un brazo cromosómico, o tan grande como todo el genoma. La profundidad de secuenciación puede expresarse como 50x, 100x, etc., donde "x" se refiere al número de veces que se cubre un locus con una lectura de secuencia. La profundidad de secuenciación también puede aplicarse a múltiples locus o a todo el genoma, en cuyo caso x puede referirse al número medio de veces que se secuencian los locus, el genoma haploide o todo el genoma. La secuenciación ultra profunda puede referirse a una profundidad de secuenciación de al menos 100x.

Un "valor de separación" corresponde a una diferencia o a un cociente entre dos valores. El valor de separación puede ser una simple diferencia o frecuencia. A modo de ejemplo, una relación directa de x/y es un valor de separación, así como $x/(x+y)$. El valor de separación puede incluir otros factores, por ejemplo, factores multiplicativos. Como otro ejemplo, se puede utilizar una diferencia o relación de funciones de los valores, por ejemplo, una diferencia o ratio de los logaritmos naturales (ln) de los dos valores. Un valor de separación puede incluir una diferencia y una proporción.

Una "abundancia relativa" es un tipo de valor de separación que relaciona una cantidad (un valor) de moléculas de ADN acelular que terminan dentro de una ventana de posición genómica con una cantidad (otro valor) de moléculas de ADN acelular que terminan dentro de otra ventana de posiciones genómicas. Las dos ventanas pueden superponerse, pero serían de diferentes tamaños. En otras implementaciones, las dos ventanas no se superpondrían. Además, las ventanas pueden tener una amplitud de un nucleótido y, por tanto, equivaler a una posición genómica.

El término "clasificación", como se usa en el presente documento, se refiere a cualquier número u otros caracteres que se asocian a una propiedad particular de una muestra. Por ejemplo, un símbolo "+" (o la palabra "positivo") podría significar que una muestra está clasificada como con deleciones o amplificaciones. La clasificación puede ser binaria (por ejemplo, positiva o negativa) o tener más niveles de clasificación (por ejemplo, una escala de 1 a 10 o de 0 a 1). Los términos "corte" y "umbral" se refieren a números predeterminados utilizados en una operación. Por ejemplo, un tamaño de corte puede referirse a un tamaño por encima del cual se excluyen los fragmentos. Un valor umbral puede ser un valor por encima o por debajo del cual se aplica una determinada clasificación. Cualquiera de estos términos puede utilizarse en cualquiera de estos contextos.

La expresión "nivel de cáncer" puede referirse a si existe el cáncer (es decir, presencia o ausencia), un estadio de un cáncer, un tamaño del tumor, si hay metástasis, la carga tumoral total del cuerpo, y/u otra medida de la gravedad de un cáncer (por ejemplo, la recurrencia del cáncer). El nivel de cáncer puede ser un número u otros indicios, tales como símbolos, letras del alfabeto y colores. El nivel podría ser cero. El nivel de cáncer también incluye las afecciones

(estados) premalignas o precancerosas asociadas a mutaciones o a un número de mutaciones. El nivel de cáncer puede usarse de varias maneras. Por ejemplo, el cribado puede comprobar si el cáncer está presente en una persona que no se sabe ha tenido cáncer con anterioridad. La evaluación puede investigar a una persona a la que se le ha diagnosticado un cáncer para controlar la evolución del mismo a lo largo del tiempo, estudiar la eficacia de las terapias o determinar el pronóstico. En una realización, el pronóstico puede expresarse como la probabilidad de que un paciente muera de cáncer, o la probabilidad de que el cáncer progrese después de una duración o tiempo específicos, o la probabilidad de que el cáncer metastatice. Detección puede significar "cribado" o puede significar comprobar si alguien, con características indicativas de cáncer (por ejemplo, síntomas u otras pruebas positivas), tiene cáncer.

Un "*máximo local*" puede referirse a una posición genómica (por ejemplo, un nucleótido) en la que se obtiene el mayor valor del parámetro de interés cuando se compara con las posiciones vecinas, o puede referirse al valor del parámetro de interés en dicha posición genómica. A modo de ejemplo, las posiciones vecinas pueden variar de 50 pb a 2000 pb. Entre los ejemplos del parámetro de interés se incluyen, pero sin limitación, el número de fragmentos que terminan en una posición genómica, el número de fragmentos que se solapan con la posición o la proporción de fragmentos que cubren la posición genómica y que son mayores que un tamaño umbral. Pueden producirse muchos máximos locales cuando el parámetro de interés tiene una estructura periódica. Un máximo global es uno específico de los máximos locales. De manera similar, un "*mínimo local*" puede referirse a una posición genómica en la que se obtiene el menor valor del parámetro de interés cuando se compara con las posiciones vecinas, o puede referirse al valor del parámetro de interés en dicha posición genómica.

Se describen factores que afectan al patrón de fragmentación del ADN acelular (por ejemplo, el ADN plasmático) y aplicaciones del análisis de los patrones de fragmentación del ADN acelular, por ejemplo, en diagnóstico molecular. Varias aplicaciones pueden utilizar una propiedad de un patrón de fragmentación para determinar una contribución proporcional de un tipo de tejido concreto, para determinar el genotipo de un tipo de tejido concreto (por ejemplo, tejido fetal en una muestra materna o tejido tumoral en una muestra de un paciente con cáncer), y/o para identificar las posiciones finales preferidas para un tipo de tejido concreto, que puede utilizarse para determinar la contribución proporcional de un tipo de tejido concreto. En algunas realizaciones, las posiciones finales preferidas para un tejido concreto también pueden utilizarse para medir la contribución absoluta de un tipo de tejido concreto en una muestra, por ejemplo, en número de genomas por unidad de volumen (por ejemplo, por mililitro).

Los ejemplos de clasificación de una contribución proporcional incluyen porcentajes específicos, intervalos de porcentajes o puede determinarse como clasificación si la contribución proporcional se encuentra por encima de un porcentaje. Para determinar la clasificación de una contribución proporcional, algunas realizaciones pueden identificar las posiciones finales preferidas correspondientes a un tipo de tejido concreto (por ejemplo, tejido fetal o tejido tumoral). Estas posiciones finales preferidas pueden determinarse de varias maneras, por ejemplo, analizando la velocidad con la que las moléculas de ADN acelular terminan en posiciones genómicas, comparando dichas tasas con otras muestras (por ejemplo, que no tengan una afección relevante), y comparando conjuntos de posiciones genómicas con altas frecuencias de aparición de extremos de moléculas de ADN acelular para diferentes tejidos y/o diferentes muestras que difieran en una afección. La abundancia relativa de las moléculas de ADN acelular que terminan en las posiciones finales preferidas en relación con las moléculas de ADN acelular que terminan en otras posiciones genómicas puede compararse con uno o más valores de calibración determinados a partir de una o más muestras biológicas de calibración cuya contribución proporcional del tipo de tejido particular se conoce. Los datos proporcionados en el presente documento muestran una relación positiva entre varias medidas de abundancia relativa y una contribución proporcional de varios tejidos en una muestra.

Para determinar la clasificación de una contribución proporcional, algunas realizaciones pueden utilizar una amplitud en un patrón de fragmentación (por ejemplo, el número de moléculas de ADN acelular que terminan en una posición genómica). Por ejemplo, pueden identificarse uno o más mínimos locales y uno o más máximos locales analizando el número de moléculas de ADN acelular que terminan en una pluralidad de posiciones genómicas. Un valor de separación (por ejemplo, una relación) de un primer número de moléculas de ADN acelular en uno o más máximos locales y un segundo número de moléculas de ADN acelular en uno o más mínimos locales se muestra positivamente relacionado con una contribución proporcional del tipo de tejido particular.

En algunas realizaciones, se podría medir la concentración del tejido de interés en relación con el volumen o el peso de las muestras de ADN acelular. Por ejemplo, la PCR cuantitativa podría utilizarse para medir el número de moléculas de ADN acelular que terminan en uno o más extremos preferidos en una unidad de volumen o de peso de la muestra de ADN acelular extraída. Se pueden realizar mediciones similares para las muestras de calibración y, por lo tanto, se puede determinar la contribución proporcional, ya que la contribución es una concentración por unidad de volumen o de peso.

Para determinar el genotipo de un tipo de tejido concreto (por ejemplo, tejido fetal o tejido tumoral) en una mezcla de ADN acelular de diferentes tipos de tejido, algunas realizaciones pueden identificar una posición final preferida para el tipo de tejido concreto. Para cada molécula de ADN acelular de un conjunto de moléculas de ADN acelular que termina en la posición final preferida, puede determinarse una base correspondiente que ocurra en la posición final preferida. Las bases correspondientes pueden utilizarse para determinar el genotipo en la posición final preferida, por ejemplo, basándose en los porcentajes de las diferentes bases observadas. En diversas implementaciones, un alto porcentaje

de una sola base (por ejemplo, por encima del 90 %) puede indicar que el genotipo es homocigoto para esa base, mientras que dos bases con porcentajes similares (por ejemplo, entre el 30-70 %) pueden llevar a determinar que el genotipo es heterocigoto.

5 Para identificar posiciones finales preferidas, algunas realizaciones pueden comparar un máximo local para los extremos izquierdos de las moléculas de ADN acelular con un máximo local para los extremos derechos de moléculas de ADN acelular. Las posiciones finales preferidas pueden identificarse cuando los máximos locales correspondientes están suficientemente separados. Además, pueden compararse las cantidades de moléculas de ADN acelular que terminan en un máximo local para el extremo izquierdo/derecho con una cantidad de moléculas de ADN acelular para un máximo local con baja separación, a fin de determinar una contribución proporcional de un tipo de tejido.

10 En la siguiente descripción, se presenta en primer lugar una descripción general de la fragmentación y de las técnicas, seguida de aspectos específicos de los patrones de fragmentación y de los ejemplos de la cuantificación de los mismos y una descripción adicional relacionada con la determinación de una contribución proporcional, la identificación de posiciones finales preferidas y la determinación de un genotipo.

I. DESCRIPCIÓN GENERAL DE LA FRAGMENTACIÓN Y LAS TÉCNICAS

20 En la presente divulgación, los presentes inventores demuestran que existe un proceso de fragmentación no aleatorio del ADN acelular. El proceso de fragmentación no aleatoria tiene lugar en cierta medida en varios tipos de muestras biológicas que contienen ADN acelular, por ejemplo, plasma, suero, orina, saliva, líquido cefalorraquídeo, líquido pleural, líquido amniótico, líquido peritoneal y líquido ascítico. El ADN acelular se produce de forma natural en forma de fragmentos cortos. La fragmentación del ADN acelular se refiere al proceso por el que se escinde, se rompe o se digiere en fragmentos cortos el ADN de alta masa molecular (tal como el ADN del núcleo de una célula) cuando se generan o liberan moléculas de ADN acelular.

30 No todas las moléculas de ADN acelular tienen la misma longitud. Algunas moléculas son más cortas que otras. Se ha demostrado que el ADN acelular, tal como ADN plasmático, es generalmente más corto y está menos intacto, es decir, tiene menos probabilidades de mantenerse intacto o tiene menos integridad, dentro de los dominios de cromatina abiertos, incluidos alrededor de los sitios de inicio de la transcripción, y en lugares entre los núcleos nucleosómicos, como en las posiciones de los enlaces (Straver et al., Prenat Diagn 2016, 36:614-621). Cada tejido diferente tiene su perfil de expresión génica característico que, a su vez, está regulado por medios que incluyen la estructura de la cromatina y el posicionamiento nucleosómico. Por tanto, los patrones de ADN acelular con probabilidad de estar intactos o íntegros en determinadas ubicaciones genómicas, como las del ADN plasmático, son firmas o distintivos del origen tisular de dichas moléculas de ADN. De manera similar, cuando un proceso patológico, por ejemplo, un cáncer, altera el perfil de expresión génica y la función del genoma de una célula, el perfil de probabilidad de ADN acelular intacto procedente de las células con enfermedad sería un reflejo de esas células. El perfil de ADN acelular, por tanto, proporcionaría pruebas de la presencia de la enfermedad o es distintivo de su presencia.

40 Algunas realizaciones mejoran aún más la resolución para estudiar el perfil de la fragmentación del ADN acelular. En lugar de limitarse a sumar las lecturas sobre un tramo de nucleótidos para identificar regiones con mayor o menor probabilidad de integridad o intactas, los presentes inventores estudiaron las posiciones finales reales o terminales de las moléculas individuales de ADN acelular, especialmente las moléculas de ADN plasmático. Notablemente, los presentes datos revelan que los lugares específicos donde se cortan las moléculas de ADN acelular no son aleatorios. El ADN genómico tisular de alta masa molecular que se corta o trata con ultrasonidos *in vitro* muestra moléculas de ADN con posiciones terminales dispersas aleatoriamente por todo el genoma. Sin embargo, hay ciertas posiciones finales de las moléculas de ADN acelular que están altamente representadas dentro de una muestra, tal como plasma. El número de apariciones o representaciones de estas posiciones finales de manera estadísticamente significativa más alto de lo esperado por mera casualidad. Estos datos hacen que nuestra comprensión de la fragmentación del ADN acelular se encuentre un paso más allá de la variación regional de la identidad (Snyder *et al.*, Cell 2016, 164: 57-68). En el presente documento los presentes inventores demuestran que el proceso de fragmentación del ADN acelular está orquestado incluso hasta la posición específica del nucleótido de corte o escisión. Los presentes inventores denominan estas posiciones no aleatorias de los extremos del ADN acelular como posiciones finales preferidas o extremos preferidos.

55 En la presente divulgación, los presentes inventores demuestran que hay posiciones finales del ADN acelular que se dan comúnmente en individuos con diferentes estados fisiológicos o patológicos. Por ejemplo, hay fines preferidos comunes compartidos por mujeres embarazadas y no embarazadas, compartidos por una embarazada y un paciente con cáncer, compartidos con individuos con y sin cáncer. Por otro lado, hay extremos preferidos que se dan sobre todo en las mujeres embarazadas, solo en pacientes con cáncer, o solo en mujeres embarazadas sin cáncer. Curiosamente, estos extremos específicos del embarazo o específicos del cáncer o específicos de una enfermedad también están altamente representados en otros individuos con un estado fisiológico o patológico comparable. Por ejemplo, los extremos preferidos identificados en el plasma de una mujer embarazada son detectables en el plasma de otras mujeres embarazadas. Además, la cantidad de una proporción de dichos extremos preferidos se correlaciona con la fracción de ADN fetal en el plasma de otras mujeres embarazadas. Dichas terminaciones preferidas están de hecho asociadas al embarazo o al feto porque sus cantidades se reducen sustancialmente en las muestras de plasma

materno posteriores al parto. De manera similar, en el cáncer, los extremos preferidos identificados en el plasma de un paciente con cáncer son detectables en el plasma de otro paciente con cáncer. Además, la cantidad de una proporción de dichos extremos preferidos se correlaciona con la fracción de ADN tumoral en el plasma de otros pacientes con cáncer. Estas terminaciones preferidas se asocian al cáncer porque sus cantidades se reducen tras el tratamiento del mismo, por ejemplo, la resección quirúrgica.

Existen varias aplicaciones o utilidades para el análisis de los extremos preferidos del ADN acelular. Estas podrían proporcionar información sobre la fracción de ADN fetal en el embarazo y, por tanto, sobre la salud del feto. Por ejemplo, se ha informado de que una serie de trastornos asociados al embarazo, como la preeclampsia, el parto prematuro, la restricción del crecimiento intrauterino (RCIU), las aneuploidías cromosómicas fetales y otros, se asocian a alteraciones en la concentración fraccional de ADN fetal, a saber, la fracción de ADN fetal, o fracción fetal, en comparación con embarazos de control emparejados según la edad gestacional. Los extremos preferidos del ADN acelular plasmático asociados al cáncer revelan la fracción de ADN tumoral o la concentración fraccional en una muestra de plasma. Conocer la fracción de ADN del tumor proporciona información sobre el estadio del cáncer, su pronóstico y ayuda a controlar la eficacia del tratamiento o la recurrencia del cáncer. El perfil de los extremos preferidos del ADN acelular también revelaría la composición de los tejidos que aportan ADN a la muestra biológica que contiene ADN acelular, por ejemplo, el plasma. Por tanto, se podría identificar el origen tisular del cáncer u otras patologías, por ejemplo, accidentes cerebrovasculares (es decir, apoplejía), manifestaciones orgánicas del lupus eritematoso sistémico.

Se puede identificar un catálogo de extremos preferidos relevantes para determinados estados fisiológicos o patológicos comparando los perfiles de ADN acelular de los extremos preferidos entre individuos con diferentes estados fisiológicos o patológicos, por ejemplo, las muestras de no embarazadas comparadas con las de embarazadas, muestras de cáncer en comparación con no cancerosas, o el perfil de la mujer embarazada sin cáncer en comparación con el perfil de las pacientes con cáncer no embarazadas. Otra estrategia consiste en comparar los perfiles de ADN acelular de extremos preferidos en diferentes momentos de un proceso fisiológico (por ejemplo, el embarazo) o patológico (por ejemplo, el cáncer). Algunos ejemplos de estos momentos son antes y después del embarazo, antes y después del parto de un feto, muestras recogidas en diferentes edades gestacionales durante el embarazo, antes y después del tratamiento del cáncer (por ejemplo, terapia dirigida, inmunoterapia, quimioterapia, cirugía), diferentes momentos tras el diagnóstico de cáncer, antes y después de la progresión del cáncer, antes y después del desarrollo de la metástasis, antes y después del aumento de la gravedad de la enfermedad, o antes y después del desarrollo de complicaciones.

Además, los extremos preferidos podrían identificarse utilizando marcadores genéticos que son relevantes para un tejido particular. Por ejemplo, las moléculas de ADN acelular que contienen un alelo SNP específico del feto serían útiles para identificar los extremos preferidos específicos del feto en una muestra como el plasma materno. Por el contrario, las moléculas de ADN plasmático que contienen un alelo SNP específico de la madre serían útiles para identificar los extremos preferidos por la madre en el plasma materno. Las moléculas de ADN plasmático que contienen una mutación específica del tumor podrían utilizarse para identificar los extremos preferidos asociados al cáncer. Las moléculas de ADN plasmático que contienen un alelo SNP específico del donante o del receptor en el contexto del trasplante de órganos son útiles para identificar los extremos preferidos del órgano trasplantado o no trasplantado. Por ejemplo, los alelos SNP específicos del donante serían útiles para identificar las terminaciones preferidas representativas del órgano trasplantado.

Un extremo preferido puede considerarse relevante para un estado fisiológico o patológico cuando tiene una alta probabilidad o posibilidad de ser detectado en ese estado fisiológico o patológico. En otras realizaciones, con una determinada probabilidad es más probable detectar un extremo preferido en el estado fisiológico o patológico relevante que en otros estados. Dado que la probabilidad de detectar un extremo preferido en un estado fisiológico o patológico relevante es mayor, dichos extremos preferidos o recurrentes (o posiciones finales) se verían en más de un individuo con ese mismo estado fisiológico o patológico. La alta probabilidad también haría que esos extremos preferidos o recurrentes fueran detectables muchas veces en la misma muestra de ADN acelular o en la alícuota del mismo individuo. En algunas realizaciones, se puede establecer un umbral cuantitativo para limitar la inclusión de los extremos que se detecten al menos un número determinado de veces (por ejemplo, 5, 10, 15, 20, etc.) dentro de la misma muestra o de la misma alícuota de muestra para ser considerado como un extremo preferente.

Tras establecer un catálogo de extremos preferidos de ADN acelular para cualquier estado fisiológico o patológico, podrían utilizarse métodos dirigidos o no dirigidos para detectar su presencia en muestras de ADN acelular, por ejemplo, el plasma, u otros individuos para determinar una clasificación de los otros individuos analizados que tienen un estado de salud, fisiológico o patológico similar. Los extremos preferidos del ADN acelular podrían detectarse mediante secuenciación no aleatoria no dirigida. Debería tenerse en cuenta la profundidad de la secuenciación, de modo que se logre una probabilidad razonable de identificar la totalidad o una porción de los extremos preferidos relevantes. Como alternativa, la captura por hibridación de locus con alta densidad de extremos preferidos podría llevarse a cabo en las muestras de ADN acelular para enriquecer la muestra con moléculas de ADN acelular con dichos extremos preferidos tras, pero sin limitación, la detección por secuenciación, micromatriz, o PCR. No obstante, como alternativa, podrían utilizarse estrategias basadas en la amplificación para amplificar y enriquecer específicamente las moléculas de ADN acelular con los extremos preferidos, por ejemplo, PCR inversa, amplificación

en círculo rodante. Los productos de amplificación podrían identificarse mediante secuenciación, micromatriz, sondas fluorescentes, electroforesis en gel y otras estrategias convencionales conocidas por los expertos en la materia.

5 En la práctica, una posición final puede ser la coordenada genómica o la identidad de los nucleótidos de la base más externa de un extremo de una molécula de ADN acelular que se detecta o determina mediante un método analítico, como, por ejemplo, secuenciación paralela masiva o secuenciación de última generación, secuenciación de molécula única, protocolos de preparación de bibliotecas de secuenciación de ADN monocatenario o bicatenario, PCR, otros métodos enzimáticos para la amplificación del ADN (por ejemplo, la amplificación isotérmica) o micromatriz. Estas técnicas *in vitro* pueden alterar los extremos auténticos *in vivo* de las moléculas de ADN acelulares. Por tanto, cada extremo detectable puede representar el extremo biológicamente verdadero o el extremo es uno o más nucleótidos hacia el interior o uno o más nucleótidos extendidos desde el extremo original de la molécula. Por ejemplo, el fragmento Klenow se utiliza para crear moléculas de ADN bicatenario con extremos romos durante la construcción de bibliotecas de secuenciación de ADN mediante el despunte de los salientes 5' y el relleno de los salientes 3'. Aunque estos procedimientos pueden revelar una posición final del ADN acelular que no es idéntica al final biológico, todavía podría determinarse la relevancia clínica. Esto se debe a que la identificación de los preferidos que son relevantes o están asociados a un estado fisiológico o patológico concreto podría basarse en los mismos protocolos de laboratorio o principios metodológicos que darían lugar a alteraciones consistentes y reproducibles de los extremos del ADN acelular tanto en las muestras de calibración como en las de prueba. Varios protocolos de secuenciación de ADN utilizan bibliotecas de ADN monocatenario (Snyder *et al.*, Cell 2016, 164: 57-68). Los extremos de las lecturas de secuencia de bibliotecas monocatenarias pueden encontrarse más hacia el interior o extenderse más allá que los extremos de las bibliotecas de ADN bicatenario.

25 La identidad genómica o la coordenada genómica de la posición final podría derivarse de los resultados de la alineación de las lecturas de la secuencia con un genoma humano de referencia, por ejemplo, hg19. Podría derivarse de un catálogo de índices o códigos que representan las coordenadas originales del genoma humano. Mientras que un extremo es el nucleótido en uno o ambos extremos de una molécula de ADN acelular, la detección del extremo podría hacerse mediante el reconocimiento de otro nucleótido u otros tramos de nucleótidos en la molécula de ADN plasmático. Por ejemplo, la amplificación positiva de una molécula de ADN plasmática con un extremo preferido se detecta mediante una sonda fluorescente que se une a las bases intermedias del amplicón. Por ejemplo, un extremo podría identificarse mediante la hibridación positiva de una sonda fluorescente que se une a algunas bases en una sección intermedia de una molécula de ADN plasmático, donde se conoce el tamaño del fragmento. De este modo, se podría determinar la identidad genómica o la coordenada genómica de un extremo calculando cuántas bases son externas a la sonda fluorescente con secuencia e identidad genómica conocidas. En otras palabras, podría identificarse o detectarse un extremo mediante la detección de otras bases en la misma molécula de ADN plasmático. Un extremo podría ser una posición o identidad de nucleótidos en una molécula de ADN acelular que se lee mediante, pero sin limitación, sondas específicas para la diana, minisequenciación y amplificación del ADN.

II. PATRONES DE FRAGMENTACIÓN DEL ADN PLASMÁTICO

40 Para el análisis del patrón de fragmentación del ADN plasmático materno, los presentes inventores secuenciaron el ADN plasmático de una mujer embarazada reclutada en el Departamento de Obstetricia y Ginecología a una edad gestacional de 12 semanas (Lo *et al.* Sci Transl Med 2010; 2(61):61ra91). El ADN plasmático obtenido de la madre se sometió a una secuenciación masiva en paralelo utilizando la plataforma Illumina Genome Analyzer. Podrían utilizarse otros secuenciadores para secuenciación masiva en paralelo o de una sola molécula. Se realizó la secuenciación *paired-end* de las moléculas de ADN plasmático. Se secuenciaron 50 pb de cada uno de los extremos de la molécula, obteniéndose de este modo un total de 100 pb por molécula. Los dos extremos de cada secuencia se alinearon con el genoma humano de referencia (Hg18 NCBI.36) utilizando el programa SOAP2 (Li R *et al.*, Bioinformatics 2009, 25:1966-7). También se extrajo el ADN de las muestras de la capa leucocitaria del padre y de la madre, así como de la muestra del CVS. Estas muestras de ADN se genotiparon utilizando el sistema Affymetrix Genome-Wide Human SNP Array 6.0.

A. Ejemplo de cuantificación de la fragmentación

55 A fin de reflejar los patrones de fragmentación, la probabilidad de intacto (P_I) puede determinarse para cada nucleótido del genoma basándose en los resultados de la secuenciación del ADN plasmático materno.

$$P_I = \frac{N_z}{N_T}$$

60 donde N_z es el número de lecturas secuenciadas de longitud completa que abarcan al menos z nucleótidos (nt) a ambos extremos (5' y 3') del nucleótido diana; y N_T es el número total de lecturas secuenciadas que abarcan el nucleótido diana.

65 El valor de P_I puede reflejar la probabilidad de tener una molécula de ADN intacta centrada en una posición concreta con una longitud del doble del valor de z más 1 ($2z + 1$). Cuanto mayor sea el valor de la probabilidad de intacto (P_I), menos probable es que el ADN plasmático se fragmente en la posición del nucleótido concreto. A fin de ilustrar esto,

la definición de la probabilidad de intacto se ilustra en la FIG. 1.

La FIG. 1 muestra un ejemplo ilustrativo para la definición de la probabilidad de intacto (P_i). T es la posición del nucleótido diana para el que se calcula la P_i . A y B son dos posiciones a z nucleótidos (nt) cadena arriba (5') y z nt cadena abajo (3') de T, respectivamente. Las líneas negras marcadas de la a a la j representan fragmentos de ADN plasmático secuenciado a partir del plasma materno. Los fragmentos a a d abarcan las tres posiciones A, B y T. Por tanto, el número de fragmentos que cubren al menos z nt a ambos lados (5' y 3') del nucleótido diana (N_z) es 4. Además, los fragmentos e, f y g también abarcan la posición T, pero no cubren las dos posiciones A y B. Por tanto, hay un total de 7 fragmentos que abarcan la posición T ($N_T=7$). Los fragmentos h y j abarcan A o B, pero no T. Estos fragmentos no se cuentan en N_z o N_T . Por consiguiente, la P_i en este ejemplo concreto es de 4/7 (57 %).

En una realización, La P_i puede calcularse utilizando 25 como valor de z. Por tanto, los fragmentos de ADN plasmático intactos se definirían como fragmentos que cubren al menos 25 nt cadena arriba de la posición diana hasta 25 nt cadena abajo de la posición diana. En otras realizaciones, pueden utilizarse otros valores de z, por ejemplo, pero sin limitación, 10, 15, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 y 80.

La P_i es un ejemplo de abundancia relativa de moléculas de ADN acelular que terminan en una ventana de posiciones genómicas. Pueden utilizarse otras métricas, por ejemplo, el valor recíproco de la P_i , que tendría una relación opuesta con la probabilidad de tener una molécula de ADN intacta. Un valor más alto del valor recíproco de la P_i indicaría una mayor probabilidad de ser una posición final o una ventana final. Otros ejemplos son un valor de p para un número medido de fragmentos de ADN finales frente a un número esperado de fragmentos de ADN finales, una proporción de fragmentos de ADN terminados de entre todos los fragmentos de ADN alineados, o una proporción de terminación final preferida (PETR), todos los cuales se describen con más detalle a continuación. Todas estas métricas de una abundancia relativa miden una tasa en la que los fragmentos de ADN acelular terminan dentro de una ventana, por ejemplo, con una amplitud de $2z+1$, donde z puede ser cero, por lo que la ventana es equivalente a una posición genómica.

B. Periodicidad del patrón de fragmentación

Ciertas regiones del genoma son propensas a una mayor tasa (frecuencia) de rotura de una región cromosómica en un tejido concreto y, por tanto, tienen una mayor tasa de fragmentos de ADN acelular que terminan dentro de una ventana en la región. Un gráfico de la abundancia relativa muestra un patrón de fragmentación, que puede tener una estructura periódica. La estructura periódica muestra las posiciones de finalización máxima (escisión alta) y las posiciones de finalización mínima (escisión baja). Cuando se utiliza la P_i , un valor máximo corresponde a una ventana de baja escisión, dado que la P_i mide una probabilidad de intacta en contraposición a una probabilidad de escisión (probabilidad de posición final), que tienen una relación inversa entre sí.

Las FIG. 2A y 2B muestran la variación de la P_i a lo largo de un segmento del cromosoma 6 utilizando 25 como valor de z, de acuerdo con las realizaciones de la presente invención. En la figura 2A, la variación de la P_i se presenta en diferentes intensidades de gris como se muestra en la clave a la izquierda. En la FIG. 2B, la variación de la P_i se visualiza en un segmento más corto. El eje x es la coordenada genómica en nucleótidos (nt) y el eje y es la P_i . La variación de la P_i tiene una periodicidad aparente de unos 180 pb.

C. Variación sincrónica de la P_i para el ADN materno y fetal en el plasma materno

Mientras que la P_i varía a lo largo del genoma con una periodicidad de aproximadamente 180 pb, los presentes inventores investigaron además si la variación en la P_i sería sincrónica para las moléculas de ADN plasmáticas derivadas del feto y de la madre. La variación sincrónica significa que los picos (máximos) y los valles (mínimos) de la P_i se producen en las mismas posiciones relativas de los nucleótidos en todo el genoma o en una proporción suficientemente alta del mismo. El umbral para definir la proporción suficientemente alta puede ajustarse para aplicaciones específicas, por ejemplo, pero sin limitación, >20 %, >25 %, >30 %, >35 %, >40 %, >45 %, >50 %, >55 %, >60 %, >65 %, >70 %, >75 %, >80 %, >85 %, >90 % y >95 %. Las dos figuras que aparecen a continuación (FIG. 3 y FIG. 4) muestran dos posibles relaciones entre las variaciones de la P_i para el ADN de origen materno y fetal en el plasma materno.

La FIG. 3 muestra la ilustración de la variación sincrónica de la P_i para el ADN de origen materno y fetal en el plasma materno. Los picos y valles de la P_i se producen en las mismas posiciones relativas para el ADN materno y fetal en todo el genoma o en la mayor parte del mismo. Si hubiera una variación sincrónica en una región, el ADN procedente del feto y el procedente de la madre tendrían el mismo patrón de fragmentación, dificultando por tanto el uso de la periodicidad de un patrón de fragmentación en la región como firma de uno de los tipos de tejido.

La FIG. 4 muestra una ilustración de la variación asincrónica de la P_i para el ADN de origen materno y fetal en el plasma materno. Los picos y valles de la P_i para el ADN materno y fetal no tienen una relación relativa constante en todo el genoma. En la región I, los picos de la P_i del ADN materno coinciden con el pico del ADN fetal. En la región II, los picos de la P_i para el ADN materno coinciden con el valle para el ADN fetal. En las regiones III y IV, los picos de la P_i del ADN materno se encuentran entre los picos y las depresiones del ADN fetal. Si la variación no fue sincrónica,

dicha diferencia en los patrones de fragmentación fetal y maternos podría utilizarse como firma para identificar el ADN que probablemente proceda del feto o de la madre. Además, dicha diferencia puede utilizarse para determinar una contribución proporcional del tejido fetal o materno, como se describe con más detalle a continuación. Por ejemplo, es más probable que los fragmentos de ADN que terminan en uno de los picos de la región II sean ADN fetal, y la abundancia relativa de fragmentos de ADN que terminan en dicho pico en comparación con otras posiciones genómicas aumentaría con el incremento de la fracción de ADN fetal.

La FIG. 5 es un diagrama de flujo 500 que muestra un análisis sobre si las moléculas de ADN materno y fetal son sincrónicas en la variación de la P_i . El análisis 500 investiga si la variación de la P_i es sincrónica entre el ADN materno y el fetal en el plasma materno. El análisis 500 puede utilizar un sistema informático. Aunque el análisis 500 se realizó mediante secuenciación, como se ha descrito anteriormente, pueden utilizarse otras técnicas, por ejemplo, como se describe en el presente documento.

En el bloque 510, el análisis 500 identifica los SNP en los que la mujer embarazada es homocigota (AA) y el feto es heterocigoto (AB). Estos SNP se denominan SNP informativos. El alelo B es el alelo específico del feto. Estos SNP informativos pueden identificarse analizando una muestra que sea única o predominantemente de origen materno. Por ejemplo, se puede utilizar la capa leucocitaria de una muestra de sangre, ya que los glóbulos blancos serían predominantemente de la madre. Las posiciones genómicas en las que solo aparece un nucleótido (o un alto porcentaje de un nucleótido, por ejemplo, por encima del 80 %, que puede depender de la fracción de ADN fetal) pueden identificarse como homocigóticas en la madre. El plasma puede analizarse para identificar posiciones homocigotas en la madre cuando se identifique un porcentaje suficiente de fragmentos de ADN que tengan otro alelo identificado.

En el bloque 520, se identificaron las moléculas de ADN plasmático que tenían el alelo B específico del feto. Estas moléculas de ADN pueden identificarse como correspondientes a tejido fetal como resultado de la identificación del alelo B.

En el bloque 530, se determinó el valor de la P_i para el ADN acelular en el plasma materno. Estos valores de P_i incluyen el ADN fetal y el materno. El valor de la P_i para una posición genómica determinada se obtuvo analizando las lecturas de la secuencia alineadas a esa posición genómica de un genoma de referencia.

En el bloque 540, los picos de P_i se determinaron analizando la salida del bloque 530. Los picos pueden identificarse de varias maneras, y cada pico puede restringirse a una sola posición genómica o permitirse que corresponda a más de una posición genómica. Los presentes inventores observaron que la P_i varía a lo largo de todo el genoma para el ADN mayoritariamente procedente de la madre en el plasma materno con un patrón de tipo senoide con una periodicidad de aproximadamente 180 pb.

En el bloque 550, se determinó la distancia entre los SNP informativos y la P_i más cercana (bloque 540) para el total del plasma materno. Los presentes inventores identificaron la posición del SNP en relación con el pico más cercano de la variación de la P_i para el ADN total del plasma procedente predominantemente de la propia mujer embarazada.

En el bloque 560, todos los fragmentos de ADN derivados del feto se agregaron. Todos los fragmentos de ADN plasmático detectados que portaban un alelo específico del feto se agregaron para el cálculo de la P_i del ADN procedente del feto. A continuación, se calculó la P_i para los fragmentos de ADN agregados procedentes del feto con referencia a la posición del pico de la P_i más cercana para el ADN total del plasma materno. El cálculo de la P_i para el ADN procedente del feto se realizó de forma similar al cálculo de la P_i para el ADN total del plasma materno.

En el bloque 570, se determinó una variación de la P_i para los fragmentos de ADN procedentes del feto en relación con los picos de la P_i para el ADN total del plasma materno. La variación se muestra en la FIG. 6.

La FIG. 6 muestra un análisis de dos muestras de plasma materno (S24 y S26) para la variación de la P_i para los fragmentos de ADN procedentes del feto (rojo/gris) y total (azul/negro) en las muestras de plasma materno. El eje vertical muestra la P_i como porcentaje. El eje horizontal muestra la distancia en pares de bases (pb) entre el SNP informativo y el pico más cercano en la P_i .

Los valores totales incluyen las contribuciones del ADN fetal y materno. Los valores totales se agregan a lo largo de todos los picos de P_i . Como puede observarse, cuanto más cerca esté el SNP del pico de la P_i , mayor será el valor de la P_i . De hecho, para los fragmentos de ADN procedentes del feto, el pico de la P_i se situó aproximadamente en la posición 0. Por tanto, la P_i alcanzó su punto máximo aproximadamente en la misma posición para los fragmentos de ADN procedentes de la madre y del feto. A partir de estos datos, los presentes inventores llegaron a la conclusión de que las variaciones de la P_i para el ADN de origen materno y fetal son sincrónicas.

Aunque los patrones de fragmentación parecen ser sincrónicos, la descripción a continuación muestra que se pueden utilizar otras propiedades además de la periodicidad para distinguir los patrones de fragmentación, permitiendo así determinar una firma para un tipo de tejido concreto. Por ejemplo, se ha observado una diferencia en la amplitud de los picos y los valles para ciertas regiones genómicas, permitiendo así que ciertas posiciones dentro de esas regiones

se utilicen para determinar un patrón de fragmentación específico del tejido.

D. Factores que afectan a la variación de los patrones de fragmentación del ADN plasmático

5 En estudios previos, se demostró que la fragmentación del ADN plasmático no era aleatoria cerca del TSS (Fan *et al.* PNAS 2008;105:16266-71). La probabilidad de que cualquier ADN plasmático termine en un nucleótido específico variaría con la distancia al TSS con una periodicidad de aproximadamente el tamaño de los nucleosomas. En general, se cree que este patrón de fragmentación es consecuencia de la degradación apoptótica del ADN. Por consiguiente, el tamaño del ADN plasmático se asemeja generalmente al tamaño del ADN asociado a un complejo de histonas.

10 En estudios previos, también se demostró que el tamaño del ADN plasmático se asemeja generalmente al tamaño del ADN asociado a un nucleosoma (Lo *et al.* Sci Transl Med 2010; 2(61):61ra91). Se cree que el ADN plasmático se genera mediante la degradación apoptótica del ADN celular (ADN nuclear y ADN mitocondrial). Este punto de vista se ve respaldado por la ausencia de este patrón nucleosómico en el ADN mitocondrial circulante, ya que el ADN mitocondrial no se asocia con las histonas en las células. Aunque se demostró que la posición de los nucleótidos en la que termina un fragmento de ADN plasmático no es aleatoria cerca de los sitios de inicio de la transcripción (Fan *et al.* PNAS 2008;105:16266-71), el mecanismo exacto que rige los patrones de fragmentación del ADN plasmático aún no está claro.

20 Recientemente, se ha demostrado además que el tamaño del ADN plasmático sería diferente en regiones con diferentes contextos de secuencia (Chandrananda *et al.* BMC Med Genomics 2015;8:29). Estos últimos datos también apoyan la hipótesis anterior de que es más probable que los fragmentos de ADN acelular comiencen y terminen en las regiones enlazadoras de los nucleosomas, en lugar de en los centros nucleosómicos. Estos resultados son coherentes con nuestro hallazgo de la variación de nucleótido a nucleótido en la probabilidad de intacto, tal y como se ha analizado en las secciones anteriores. En este punto, planteamos además la hipótesis de que la amplitud de la variación de la probabilidad de intacto variaría en las diferentes regiones genómicas. Esta variación de región a región en la variabilidad de la fragmentación no se ha explorado ni cuantificado adecuadamente en ningún estudio anterior. Las siguientes figuras ilustran el concepto de variación local y regional de la P_i .

30 La FIG. 7 muestra una ilustración de la amplitud de la variación de la P_i . En las secciones anteriores, los presentes inventores han demostrado que existe un patrón de variación de tipo sinusoidal en la P_i en un tramo corto de ADN. Aquí analizamos además la amplitud de la variación en regiones genómicas más amplias. La amplitud de la variación se refiere a la diferencia de la P_i entre el pico más alto y la variación más baja de la P_i en una región concreta con un tamaño determinado. En una realización, el tamaño de una región concreta puede ser de 1000 pb. En otras realizaciones, pueden utilizarse otros tamaños, por ejemplo, pero sin limitación, 600 pb, 800 pb, 1500 pb, 2000 pb, 3000 pb, 5000 pb y 10000 pb.

40 Como se muestra en la FIG. 7, la amplitud de la región 1 es mayor que la de la región 2. Este comportamiento se observa en los datos a continuación. Si dichas apariciones de altas amplitudes se dan en diferentes regiones genómicas para diferentes tejidos, puede utilizarse una medición de la amplitud para determinar una contribución proporcional de un tipo de tejido cuando se analiza una región en la que la amplitud difiere entre los tipos de tejido. Por ejemplo, si la amplitud es diferente para los distintos tipos de tejidos, la contribución proporcional variaría proporcionalmente con una cantidad creciente de ADN de un tipo de tejido concreto (por ejemplo, tejido fetal o tejido tumoral). Por consiguiente, una medida de la amplitud correspondería a una contribución proporcional particular. Las realizaciones pueden utilizar datos de calibración de muestras en las que la contribución proporcional se mide mediante otra técnica (por ejemplo, mediante el análisis de alelos, firmas de metilación, grado de amplificación/delección) como se describen en las publicaciones de patentes estadounidenses n.º 2009/0087847, 2011/0276277, 2011/0105353, 2013/0237431 y 2014/0100121.

50 En nuestros datos de secuenciación, observamos que la amplitud de la variación de la P_i variaba en diferentes regiones genómicas. La hipótesis de los presentes inventores es que la amplitud de la variación de la P_i está relacionada con la accesibilidad de la cromatina a la degradación durante la apoptosis. Por tanto, los presentes inventores investigaron la posible relación entre la amplitud de la variación y los sitios de hipersensibilidad a DNAsas en el genoma. En un estudio anterior, se observó que el patrón de fragmentación del ADN plasmático se ve afectado por su posición relativa respecto al TSS. En el este análisis, los presentes inventores investigaron la importancia relativa de los sitios de hipersensibilidad al TSS y a DNAsas en el efecto de los patrones de fragmentación del ADN plasmático. Se pueden utilizar otros lugares en los que la amplitud se corresponda con el tejido analizado. Un ejemplo de este tipo de sitios es el que se identifica mediante el ensayo de cromatina accesible por transposasa con secuenciación de alto rendimiento (ATAC-Seq) (Buenrostro *et al.* Nat Methods 2013; 10: 1213-1218). Otro ejemplo de este tipo de sitios es el que se identifica con la nucleasa de micrococos (MNasa).

Los presentes inventores compararon la amplitud de la variación de la P_i en dos tipos de regiones genómicas:

- 65 i. Regiones que son TSS pero no sitios de hipersensibilidad a DNAsas; y
ii. Regiones que son sitios de hipersensibilidad a DNAsas pero no TSS.

Las coordenadas del TSS y de los sitios de hipersensibilidad a DNAsas se recuperaron de la base de datos ENCODE (genome.ucsc.edu/ENCODE/downloads.html).

Los patrones de P_i alrededor de los sitios de TSS y DNasa I se perfilaron utilizando la siguiente estrategia.

1) Se recuperaron las regiones a 2 kb cadena arriba y cadena abajo alrededor de los sitios de referencia seleccionados.

2) A continuación, se reescalaron las coordenadas genómicas absolutas en función de la distancia a un sitio de referencia. Por ejemplo, si una ventana particular con 60 pb de tamaño está a 50 pb de un sitio de referencia en dirección cadena arriba, se marcará como -50. De lo contrario, si una ventana particular con 60 pb de tamaño está a 50 pb del sitio de referencia en dirección cadena abajo, se marcará como +50.

3) El valor de P_i en una ventana concreta con las mismas coordenadas nuevas reescaladas se recalculará utilizando el recuento de fragmentos intactos y de todos los fragmentos que se solapan con dicha ventana.

La FIG. 8A muestra los patrones de variación de la P_i en regiones que son sitios de hipersensibilidad a DNasa, pero no TSS. La FIG. 8B muestra patrones de variación de la P_i en regiones que son TSS, pero no sitios de hipersensibilidad a DNasa. Como se muestra, la amplitud de la variación es mucho mayor en las regiones que son sitios de hipersensibilidad a DNAsas, pero no TSS, que las que son TSS, pero no sitios de hipersensibilidad a DNAsas. Estas observaciones sugieren que un factor que influye en el patrón de fragmentación del ADN plasmático es la posición relativa de una región sometida a fragmentación a los sitios de hipersensibilidad a DNAsas.

III. USO DE PICOS Y VALLES PARA DETERMINAR LA PROPORCIÓN DE TEJIDO

Habiendo demostrado que la posición relativa a los sitios de hipersensibilidad a DNAsas es un factor importante que rige el patrón de fragmentación del ADN plasmático, los presentes inventores investigaron si esta observación puede traducirse en aplicaciones clínicas. Se ha observado que los perfiles de los sitios de hipersensibilidad a DNAsas son diferentes en los distintos tipos de tejidos. Los perfiles corresponden a las localizaciones genómicas de los sitios; las localizaciones de los sitios de hipersensibilidad a DNAsas son diferentes para los distintos tejidos. Por tanto, razonamos que el ADN plasmático liberado de diferentes tipos de tejidos mostraría patrones de fragmentación específicos de los tejidos. De manera similar, pueden usarse otras regiones donde la amplitud de una región varía de un tejido a otro.

A. Ejemplo de sitios de hipersensibilidad a DNAsas

La FIG. 9 muestra una ilustración del principio para la medición de la proporción de ADN liberado de diferentes tejidos. El ADN plasmático derivado del tejido A tiene una menor probabilidad de fragmentarse en las posiciones de nucleótidos con alta P_i (picos, indicados por P). Por consiguiente, los extremos del ADN plasmático derivado del tejido A tiene una menor probabilidad de estar localizados en estas posiciones de nucleótidos. Por el contrario, los extremos del ADN plasmático derivado del tejido A tienen una mayor probabilidad de estar localizados en posiciones de nucleótidos con baja P_i (valles, indicados por T). Por otro lado, ya que este sitio no es un sitio de hipersensibilidad a DNAsas para el tejido B, la amplitud de la variación de P_i es baja para el ADN plasmático derivado del tejido B. Por lo tanto, la probabilidad de que el ADN plasmático del tejido B termine en las posiciones P y T sería similar, al menos en relación con la cantidad de variación observada para el tejido A.

Los presentes inventores definieron la proporción de extremos de fragmentos en las regiones que son sitios de hipersensibilidad a DNAsas del tejido A (FRA) como sigue:

$$FRA = \frac{N_T}{N_P}$$

donde N_T es el número de fragmentos de ADN plasmático que terminan en posiciones de nucleótidos de los valles de P_i y N_P es el número de fragmentos de ADN plasmático que terminan en posiciones de nucleótidos de los picos de P_i . El FRA es un ejemplo de un valor de separación, y más concretamente un ejemplo de la abundancia relativa de los fragmentos de ADN que terminan en el valle con respecto a los que terminan en el pico. En otras realizaciones, se pueden determinar las relaciones separadas de los valles (mínimo local) y los picos (máximo local) vecinos, y se puede determinar una media de las relaciones separadas.

Para el tejido A, FRA sería mayor que 1 dado que N_T sería mayor que N_P . Para el tejido B, FRA sería aproximadamente 1 dado que N_T y N_P serían similares. Por consiguiente, en una mezcla que contenga el ADN plasmático derivado de los tejidos A y B, el valor de FRA tendría una correlación positiva con la contribución proporcional del tejido A. En la práctica, FRA para el tejido B no es necesariamente 1. Siempre que FRA del tejido B sea diferente de FRA del tejido A, la contribución proporcional de los dos tipos de tejidos puede determinarse a partir de FRA.

En dichas regiones, la alta variación en la probabilidad de que los fragmentos de ADN terminen en los valles dará

lugar a un mayor número de fragmentos de ADN que terminen en dichas posiciones que en los puntos máximos (tenga en cuenta que, para diferentes valores de abundancia relativa definidos, puede darse una mayor probabilidad para los picos). Cuando más fragmentos de ADN son del tipo de tejido A, mayor será la diferencia en el número de fragmentos de ADN que terminan en los valles y en los picos. Por tanto, a medida que aumenta la contribución proporcional del tejido A, mayor será la separación entre el número de fragmentos de ADN que terminan en un valle y el número de fragmentos de ADN que terminan en un pico. Este valor de separación corresponde a la alta amplitud en la función de verosimilitud mostrada en la FIG. 9 para el tejido A.

B. Relación entre la abundancia relativa y la contribución proporcional

La FIG. 10 muestra la relación entre la FRA y la contribución proporcional del tejido A al ADN en una mezcla determinada mediante el análisis de dos o más muestras de calibración con concentraciones proporcionales conocidas de ADN del tejido A. En el ejemplo mostrado, se analizan dos muestras con contribución proporcional de tejido A de X_1 y X_2 . Los valores FRA de las dos muestras se determinaron como y_1 e y_2 , respectivamente. La relación entre FRA y la contribución proporcional de A puede determinarse a partir de los valores de x_1 , x_2 , y_1 e y_2 .

Los valores y_1 e y_2 son ejemplos de valores de calibración. Los puntos de datos (x_1, y_1) y (x_2, y_2) son ejemplos de puntos de datos de calibración. Los puntos de datos de calibración pueden ajustarse a una función para obtener una curva de calibración 1010, que puede ser lineal. Cuando se mide un nuevo FRA (u otro valor de abundancia relativa) para una nueva muestra, el nuevo FRA puede compararse con al menos uno de los valores de calibración para determinar una clasificación de la contribución proporcional de la nueva muestra. La comparación con el valor de calibración puede hacerse de varias maneras. Por ejemplo, la curva de calibración puede utilizarse para encontrar la contribución proporcional x correspondiente a la nueva FRA. Como otro ejemplo, el nuevo FRA puede compararse con el valor de calibración y_1 de un primer punto de datos de calibración para determinar si la nueva muestra como contribución proporcional es mayor o menor que x_1 .

En otras realizaciones, una mezcla que contenga más de dos tipos de tejidos puede analizarse de forma similar para la contribución proporcional de los tejidos A siempre que FRA de los demás tejidos sea relativamente constante. Estos métodos son útiles en la práctica para el análisis de diferentes escenarios clínicos, por ejemplo, pero sin limitación, para la detección del cáncer, el seguimiento de trasplantes, el seguimiento de traumatismos, el diagnóstico de infecciones y prenatal.

En una realización, se puede determinar la concentración fraccional del tejido afectado en el plasma de un paciente con cáncer. Por ejemplo, en un paciente con cáncer hepático, la contribución fraccional del ADN hepático puede determinarse mediante el análisis de las regiones de cromatina abierta específicas del hígado, por ejemplo, sitios de hipersensibilidad a DNAsas. En una realización, esto se puede hacer usando DNasa-Seq (Boyle *et al.* Cell 2008; 132: 311-322; Madrigal *et al.* Front Genet 2012; 16: 123-131). En otra realización, esto puede realizarse mediante el Aislamiento de Elementos Reguladores Asistido por Formaldehído (FAIRE)-Seq (Giresi *et al.* Genome Res 2007; 17: 877-885). En otra realización más, esto puede realizarse mediante ATAC-Seq (Buenrostro *et al.* Nat Methods 2013; 10: 1213-1218). El $FR_{\text{hígado}}$ puede determinarse en estos sitios y compararse con sujetos sanos normales. En los sitios de hipersensibilidad a DNAsas específicos del hígado, la variación en la P_i entre las regiones de pico y de valle estaría aportada principalmente por el hígado. Mediante la comparación con una curva de calibración similar a la de la FIG. 10, puede determinarse la contribución del hígado. El valor de $FR_{\text{hígado}}$ del caso analizado puede compararse con un rango de la contribución del hígado en los sujetos sanos. Se pueden utilizar otras regiones que tengan una gran variación de amplitud en la función de probabilidad de los fragmentos de ADN que terminan en una posición genómica entre varios tejidos de una mezcla. Los ejemplos de estas otras regiones se describen con más detalle en secciones posteriores.

De manera similar, la contribución del órgano trasplantado en un paciente que ha recibido un trasplante de órgano puede determinarse mediante este método. En estudios previos, se demostró que en los pacientes con rechazo se producía una mayor liberación de ADN del órgano trasplantado, lo que daba lugar a una concentración elevada del ADN del órgano trasplantado en el plasma. El análisis de la FR del órgano trasplantado sería un medio útil para la detección y el seguimiento del rechazo del órgano. Las regiones utilizadas para dicho análisis pueden variar en función del órgano trasplantado.

En otra realización, este método puede utilizarse para determinar la concentración de ADN fetal en el plasma materno. En el plasma materno, las moléculas de ADN que portan los genotipos fetales proceden en realidad de la placenta. Por tanto, si nos centramos en los sitios de hipersensibilidad a DNAsas que son específicos de la placenta pero que no están presentes en las células sanguíneas, se podría determinar la contribución proporcional de la placenta al ADN plasmático mediante el análisis de FR_{placenta} .

La FIG. 11 muestra una correlación entre la FR_{placenta} y el porcentaje de ADN fetal en el plasma materno según las realizaciones de la presente invención. El eje vertical corresponde a la FR_{placenta} determinada mediante uno o más máximos locales y mínimos locales que se encuentran en uno o más sitios de hipersensibilidad a DNAsas. El eje horizontal es la fracción de ADN fetal medida mediante una técnica de medición independiente. Como puede observarse, el valor de FR_{placenta} está correlacionado con la fracción de ADN fetal. En este ejemplo, la fracción de ADN

fetal se determinó basándose en la proporción de alelos específicos del feto en los SNP para los que la madre era homocigota y el feto era heterocigoto. Por tanto, el porcentaje de ADN fetal puede estimarse mediante $FR_{placenta}$ basándose en los resultados de la secuenciación del ADN plasmático materno.

5 Como alternativa, ya que los dos componentes clave en el plasma materno son el ADN procedente de la placenta y el ADN procedente de las células sanguíneas (un tipo de tejido diferente), los presentes inventores razonaron que FR_{sangre} estaría correlacionado negativamente con la concentración fraccional de ADN fetal en el plasma sanguíneo. Por tanto, se identificaron los sitios de hipersensibilidad a DNAsas específicos para las células sanguíneas y se determinó FR_{sangre} .

10 La FIG. 12 muestra una correlación entre FR_{sangre} y la concentración de ADN fetal en el plasma materno. El eje vertical corresponde a la FR_{sangre} determinada mediante uno o más máximos locales y mínimos locales que se encuentran en uno o más sitios de hipersensibilidad a DNAsas. El eje horizontal es la fracción de ADN fetal medida en función de la proporción de alelos específicos del feto en el plasma materno. Se pudo observar una correlación negativa entre FR_{sangre} y el porcentaje de ADN fetal. Por tanto, el porcentaje de ADN fetal puede estimarse mediante FR_{sangre} basándose en los resultados de la secuenciación del ADN plasmático materno. Por consiguiente, una región genómica puede tener un patrón de fragmentación específico para múltiples tipos de tejidos, por ejemplo, correlaciones positivas para algunos tejidos y negativas para otros.

20 C. Método que utiliza los máximos y los mínimos

La FIG. 13 es un diagrama de flujo de un método 1300 de análisis de una muestra biológica para determinar una clasificación de una contribución proporcional del primer tipo de tejido de acuerdo con las realizaciones de la presente invención. La muestra biológica incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye el primer tipo de tejido. Como en el caso de los otros métodos descritos en el presente documento, el método 1300 puede utilizar un sistema informático. El primer tipo de tejido (por ejemplo, tejido hepático o tejido fetal) puede seleccionarse en función del sujeto específico. Por ejemplo, si el sujeto ha tenido cáncer de hígado en el pasado, se puede realizar un cribado para comprobar si el cáncer de hígado ha regresado, lo que supondría un aumento de la contribución proporcional del tejido hepático. Este criterio de selección se aplica a otros métodos descritos en el presente documento.

En el bloque 1310, se identifica al menos una región genómica que tiene un patrón de fragmentación específico del primer tipo de tejido. A modo de ejemplo, la al menos una región genómica puede incluir uno o más sitios de hipersensibilidad a DNAsas. Cada una de las al menos una región genómica con un patrón de fragmentación específico del primer tipo de tejido puede incluir uno o más alelos específicos del primer tejido en al menos una muestra adicional, por ejemplo, como se describirá con detalle en la sección VI. Como otro ejemplo, la al menos una región genómica puede incluir uno o más sitios de nucleasas ATAC-seq o de micrococos. El primer tipo de tejido puede corresponder a un órgano concreto o incluso a un cáncer particular del órgano.

40 En el bloque 1320, se analiza una pluralidad de moléculas de ADN acelular de la muestra biológica. El análisis de una molécula de ADN acelular incluye la determinación de una posición genómica (posición final) en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular. Por tanto, se pueden determinar dos posiciones finales, o solo una posición final de la molécula de ADN acelular.

45 Las posiciones finales preferidas pueden determinarse de varias maneras, como se describe en el presente documento. Por ejemplo, las moléculas de ADN acelular pueden secuenciarse para obtener lecturas de secuencia, y las lecturas de secuencia pueden cartografiarse (alinearse) con el genoma de referencia. Si el organismo es un ser humano, el genoma de referencia sería un genoma humano de referencia, potencialmente de una subpoblación particular. Como otro ejemplo, las moléculas de ADN acelular pueden analizarse con diferentes sondas (por ejemplo, tras la PCR u otro tipo de amplificación), donde cada sonda corresponde a una localización genómica, que puede abarcar al menos una región genómica.

Se puede analizar un número estadísticamente significativo de moléculas de ADN acelular para poder determinar con precisión la contribución proporcional del primer tipo de tejido. En algunas realizaciones, se analizan al menos 1.000 moléculas de ADN acelular. En otras realizaciones, se pueden analizar al menos 10.000 o 50.000 o 100.000 o 500.000 o 1.000.000 o 5.000.000 de moléculas de ADN libre de células o más.

60 En el bloque 1330, se identifica un primer conjunto de primeras posiciones genómicas. Cada primera posición genómica tiene un mínimo local de extremos de moléculas de ADN acelular corresponde a la primera posición genómica. Varias posiciones genómicas vecinas pueden definirse como un extremo local (máximo o mínimo), y por tanto un máximo local no se limita a una sola posición.

En algunas realizaciones, se puede determinar una proporción para cada una de una pluralidad de posiciones genómicas. Se puede determinar una primera cantidad de moléculas de ADN acelular que terminan en la posición genómica y se extienden al menos un número especificado de nucleótidos a ambos lados de la posición genómica, por ejemplo, como se describe para la FIG. 1. Una segunda cantidad de moléculas de ADN acelular que se encuentran

en la posición genómica puede utilizarse con la primera cantidad para determinar la proporción. Se puede identificar una pluralidad de mínimos locales y una pluralidad de máximos locales en las relaciones, por ejemplo, recorriendo los valores de la proporción para identificar una o más posiciones genómicas contiguas que se encuentren en cada uno de los extremos (máximo o mínimo).

5 En el bloque 1340, se identifica un segundo conjunto de segundas posiciones genómicas. Cada segunda posición genómica que tiene un máximo local de extremos de moléculas de ADN acelular corresponde a la segunda posición genómica. El segundo conjunto puede identificarse de manera similar al primer conjunto.

10 En el bloque 1350, se determina un primer número de moléculas de ADN acelular que terminan en cualquiera de las primeras posiciones genómicas en cualquiera de la al menos una región genómica. El primer número puede determinarse de varias maneras, por ejemplo, como una suma a lo largo de todas las primeras posiciones genómicas. Como otro ejemplo, se puede determinar una cantidad independiente en cada posición genómica. Por tanto, la determinación del primer número de moléculas de ADN acelular puede incluir la determinación de una primera cantidad de moléculas de ADN acelular que terminan en cada primera posición genómica, determinando así una pluralidad de primeras cantidades.

15 En el bloque 1360, se determina un segundo número de moléculas de ADN acelular que terminan en cualquiera de las segundas posiciones genómicas en cualquiera de la al menos una región genómica. El segundo número puede determinarse de forma similar al primero. Por tanto, la determinación del segundo número de moléculas de ADN acelular puede incluir la determinación de una segunda cantidad de moléculas de ADN acelular que terminan en cada segunda posición genómica, determinando así una pluralidad de segundas cantidades.

20 En el bloque 1370, se calcula un valor de separación utilizando el primer número y el segundo. El valor de separación puede calcularse de varios modos, por ejemplo, como una proporción del primer número y el segundo número, como se describe en la sección III.A. En otra implementación que utiliza múltiples máximos y mínimos, se puede determinar una cantidad en cada una de esas posiciones genómicas. El cálculo del valor de separación puede incluir la determinación de una pluralidad de proporciones separadas, cada proporción separada de una de la pluralidad de primeras cantidades y una de la pluralidad de segundas cantidades. El valor de separación puede determinarse utilizando la pluralidad de relaciones separadas, por ejemplo, la media o la mediana de las proporciones separadas.

25 En el bloque 1380, la clasificación de la contribución proporcional del primer tipo de tejido se determina comparando el valor de separación con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido se conocen.

35 *D. Análisis sin amplificación*

El análisis de las moléculas de ADN acelular en el bloque 1310 puede ser sin amplificación. Cuando se utiliza la PCR, la profundidad de la secuenciación (es decir, el número de lecturas de la secuencia que cubren un nucleótido concreto o que terminan en ese nucleótido concreto en un genoma de referencia) no refleja directamente cuántas moléculas de ADN plasmático que cubren ese nucleótido concreto se analizan. Esto se debe a que una molécula de ADN plasmático puede generar múltiples réplicas durante el proceso de la PCR, y pueden originarse múltiples lecturas de secuencias a partir de una única molécula de ADN plasmático. Este problema de duplicación sería más importante con i) un mayor número de ciclos de PCR para amplificar la biblioteca de secuenciación; ii) una mayor profundidad de secuenciación, y iii) un menor número de moléculas de ADN en la muestra de plasma original (por ejemplo, un menor volumen de plasma).

40 Además, el paso de la PCR introduce más errores (Kinde *et al.* Proc Natl Acad Sci USA 2011; 108: 9530-9535) dado que la fidelidad de una ADN polimerasa no es del 100 %, y ocasionalmente, podría incorporarse un nucleótido erróneo en la hebra hija producida mediante la PCR. Si este error de PCR se produce durante los primeros ciclos de la PCR, se generarían clones de moléculas hijas que mostrarían el mismo error. La concentración fraccional de la base errónea puede alcanzar una proporción tan alta entre las demás moléculas de ADN del mismo locus que el error sería mal interpretado, por ejemplo, como una mutación procedente del feto o del tumor. Algunos ejemplos de protocolos sin PCR son: Berry Genomics (investor.illumina.com/mobile.view?c=121127&v=203&d=1&id=1949110); Illumina (www.illumina.com/products/truseq-dna-pcr-free-sample-prep-kits.html), y diversas técnicas de secuenciación de una sola molécula. Se pueden encontrar más detalles de un análisis sin amplificación en la solicitud PCT n.º PCT/CN2016/073753

50 Por consiguiente, algunas realizaciones pueden incluir la obtención de moléculas de ADN molde a partir de la muestra biológica a analizar; preparar una biblioteca de secuenciación de moléculas de ADN analizables utilizando las moléculas de ADN molde, la preparación de la biblioteca de secuenciación de moléculas de ADN analizables que no incluye una etapa de amplificación del ADN de las moléculas de ADN molde; secuenciar la biblioteca de secuenciación de moléculas de ADN analizables para obtener una pluralidad de lecturas de secuencia correspondientes a la primera pluralidad de moléculas de ADN acelular. El análisis de la primera pluralidad de moléculas de ADN acelular puede incluir recibir, en el sistema informático, la pluralidad de lecturas de la secuencia y alinear, mediante el sistema informático, la pluralidad de lecturas de la secuencia al genoma de referencia para determinar las posiciones

genómicas de la pluralidad de lecturas de la secuencia.

IV. ABUNDANCIA RELATIVA DE LOS NUCLEÓTIDOS A LA IZQUIERDA Y A LA DERECHA

5 La FIG. 14 muestra una ilustración del principio de una diferencia de los fragmentos de ADN circulante en relación con el ADN de origen tumoral o fetal. En estudios previos, se ha demostrado que el tamaño del ADN circulante se asemeja mucho al tamaño del ADN nucleosómico. El pico principal de 166 pb en la distribución del tamaño del ADN plasmático representa el ADN asociado al núcleo del complejo de histonas junto con el ADN enlazador que conecta dos complejos de histonas sucesivos.

10 También se ha observado que las distribuciones de tamaño de las moléculas de ADN derivadas del feto y del tumor son más cortas que las del ADN no tumoral y no derivado del feto en el plasma de pacientes con cáncer y de mujeres embarazadas (Lo *et al.* Sci Transl Med 2010; 2(61):61ra91 y Jiang *et al.* Proc Natl Acad Sci USA 2015;112:E1317-25.). Para la distribución del tamaño del ADN derivado del tumor y del feto en el plasma, el pico de 166 pb disminuye y un pico de 144 pb es más prominente. El pico de 144 pb se debe probablemente a la degradación del ADN enlazador de ~20 pb que conecta dos complejos de histonas sucesivos.

15 Para ilustrar el principio de este método, los presentes inventores utilizaron el escenario de un paciente con cáncer como ejemplo. El mismo principio puede aplicarse a otros escenarios, incluido el análisis del ADN fetal circulante en el plasma materno en el embarazo, y el análisis del plasma de pacientes que han recibido un trasplante. Las realizaciones pueden analizar los extremos de las moléculas de ADN plasmático, denotados como los extremos izquierdo y derecho en la FIG. 14.

20 Cuando el ADN de los tejidos no malignos se fragmenta y se libera en el plasma, los extremos de conexión de las dos moléculas se situarían ambos en la posición del nucleótido A. En otras palabras, para la molécula del lado derecho, el nucleótido más externo de la izquierda está justo al lado de la posición del nucleótido A. Para la molécula del lado izquierdo, el nucleótido más externo de la derecha también está justo al lado de la posición del nucleótido A. Cuando la abundancia relativa de las moléculas que terminan en un nucleótido particular se traza contra la coordenada del nucleótido, los picos de abundancia de los extremos estarían en la posición A para los nucleótidos más externos a la izquierda y a la derecha que corresponden a esta región. Para las moléculas de ADN derivadas de células tumorales, un fragmento de 20 pb se eliminaría de las moléculas tras el proceso de fragmentación.

25 Como resultado, habría un espacio de 20 pb entre el lado izquierdo de la molécula de la derecha y el lado derecho de la molécula de la izquierda. Cuando la abundancia relativa de las moléculas que terminan en un determinado nucleótido se representa frente a la coordenada del nucleótido, los picos del nucleótido más externo de la derecha (situado en B) y el pico del nucleótido más externo de la izquierda (situado en C) estarían separados por 20 pb. Por consiguiente, la relación entre la abundancia de moléculas que terminan en las posiciones de nucleótidos B y C y la abundancia de moléculas que terminan en la posición A representaría la concentración fraccional de ADN derivado del tumor en la muestra de plasma.

30 El mismo principio puede aplicarse para la cuantificación de especies de ADN que tienen una distribución de tamaño diferencial, por ejemplo, pero sin limitación, la medición del ADN fetal en el plasma de las mujeres embarazadas y la medición del ADN de un órgano trasplantado.

35 La FIG. 15 es un diagrama de flujo de un método 1500 de análisis de una muestra biológica que incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido. Se pueden utilizar partes del método 1500 para implementar el bloque 1310 y otros bloques que identifiquen las posiciones finales preferidas.

40 En el bloque 1510, las moléculas de ADN acelular se analizan para determinar las posiciones finales izquierda y derecha en un genoma de referencia. El bloque 1510 puede realizarse de un modo similar al bloque 1320. En el bloque 1510, se puede analizar una primera pluralidad de moléculas de ADN acelular procedentes de la muestra biológica de un sujeto, donde cada una de la primera pluralidad de moléculas de ADN acelular tiene un extremo izquierdo y un extremo derecho. Se puede determinar una posición final izquierda en el genoma de referencia correspondiente al extremo izquierdo de la molécula de ADN acelular, por ejemplo, mediante la alineación (cartografía) de una lectura de la secuencia del fragmento de ADN con el genoma de referencia o mediante una sonda cuya posición se conoce en el genoma de referencia. El extremo izquierdo puede referirse a cualquier extremo, dependiendo del sistema de coordenadas elegido para definir el genoma de referencia. De manera similar, se puede determinar una posición final derecha en el genoma de referencia correspondiente al extremo derecho de la molécula de ADN acelular. Las dos posiciones finales pueden determinarse en dos pasos de alineación separados, por ejemplo, si los dos extremos tienen lecturas de secuencias separadas.

45 En el bloque 1520, se identifica un conjunto de posiciones genómicas izquierdas. Cada posición genómica del conjunto izquierdo tiene un máximo local de extremos izquierdos de la primera pluralidad de moléculas de ADN acelular correspondiente a una de las posiciones genómicas del conjunto izquierdo. El conjunto izquierdo puede determinarse de forma similar a la descrita para los máximos del método 1300.

En el bloque 1530, se identifica un conjunto de posiciones genómicas correctas. Cada posición genómica del conjunto derecho tiene un máximo local de extremos derechos de la primera pluralidad de moléculas de ADN acelar correspondiente a una de las posiciones genómicas del conjunto derecho. El conjunto correcto puede determinarse de forma similar a la descrita para los máximos del método 1300.

5 En el bloque 1540, un primer conjunto de posiciones genómicas se identifica como específico del primer tipo de tejido. Todas o una parte de las posiciones genómicas izquierdas del conjunto izquierdo pueden compararse con todas o una parte de las posiciones genómicas derechas del conjunto derecho para identificar el primer conjunto de posiciones genómicas en el que una distancia desde una posición genómica izquierda a una posición genómica derecha más cercana es mayor que una primera distancia umbral de posiciones genómicas (por ejemplo, nucleótidos) en el genoma de referencia. Algunos ejemplos de la primera distancia umbral son 5, 6, 7, 8, 9, 10, 15 y 20 nucleótidos.

15 En el bloque 1550, se identifica un segundo conjunto de posiciones genómicas. Todas o una parte de las posiciones genómicas izquierdas del conjunto izquierdo pueden compararse con todas o una parte de las posiciones genómicas derechas del conjunto derecho para identificar el segundo conjunto de posiciones genómicas en el que una distancia de una posición genómica izquierda a una posición genómica derecha más cercana es menor que una segunda distancia umbral de posición genómica en el genoma de referencia. Algunos ejemplos de la segunda distancia umbral son 2, 3, 4 y 5 posiciones genómicas (por ejemplo, nucleótidos).

20 En el bloque 1560, se determina un valor de separación utilizando un primer número de la primera pluralidad de moléculas de ADN acelar que termina en una de las posiciones genómicas del conjunto izquierdo y un segundo número de la primera pluralidad de moléculas de ADN acelar que termina en una de las posiciones genómicas del conjunto derecho. Se puede determinar un valor de separación (por ejemplo, un valor de abundancia relativa) entre el primer número y el segundo.

25 En una realización, se identifican los pares del primer conjunto de posiciones genómicas y del segundo conjunto de posiciones genómicas. Los pares pueden ser de posiciones más cercanas entre sí. Para cada uno de uno o más de los pares, se puede determinar una primera cantidad de moléculas de ADN acelar que termina en la primera posición genómica, y una segunda cantidad de moléculas de ADN acelar que termina en la primera posición genómica. Las primeras cantidades de moléculas de ADN acelar corresponden al primer número de la pluralidad de moléculas de ADN acelar y las segundas cantidades de moléculas de ADN acelar corresponden al segundo número de la pluralidad de moléculas de ADN acelar. Por ejemplo, las primeras cantidades pueden sumar el primer número y las segundas cantidades pueden sumar el segundo número, y el valor de separación puede determinarse directamente a partir del primer número y del segundo número. Como otro ejemplo, el valor de separación puede determinarse a partir de una pluralidad de proporciones, cada una de las cuales incluye la primera cantidad y la segunda cantidad para uno de los pares. En diversas implementaciones, se puede utilizar un promedio o una mediana de las proporciones como valor de separación. Las respectivas cantidades primera y segunda de los pares pueden utilizarse de otras maneras para determinar los valores de separación individuales utilizados para determinar el valor de separación total.

40 En el bloque 1570, la clasificación de la contribución proporcional del primer tipo de tejido se determina comparando el valor de separación con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido se conocen. El bloque 1570 puede realizarse de un modo similar a otras determinaciones de las contribuciones proporcionales.

45 En diversas realizaciones, los conjuntos derechos e izquierdos pueden usarse como el primer conjunto de posiciones genómicas; solo se puede utilizar el conjunto izquierdo; se puede utilizar el conjunto derecho; o se pueden utilizar parte del conjunto izquierdo y parte del derecho. Para todo el conjunto de posiciones a la izquierda, hay un subconjunto de posiciones a la izquierda que tiene un conjunto correspondiente de posiciones a la derecha separadas del subconjunto de posiciones a la izquierda por un número umbral de nucleótidos. Por consiguiente, es posible utilizar el subconjunto de posiciones a la izquierda o el correspondiente subconjunto de posiciones a la derecha para realizar el cálculo.

V. USO DE POSICIONES FINALES ESPECÍFICAS DE LOS TEJIDOS

55 Los presentes inventores barajan la hipótesis de que los patrones de fragmentación del ADN circulante derivado de las células cancerosas, las células de la placenta y los tipos de células serían diferentes. Basándose en esta hipótesis, la coordenada de los nucleótidos terminales en uno o ambos extremos de un fragmento de ADN circulante puede utilizarse para predecir si el fragmento de ADN portador de una supuesta mutación procede realmente de un tumor. En los fragmentos de ADN plasmático pueden identificarse posiciones finales específicas del cáncer y del embarazo.

60 A. Ejemplo de cáncer utilizando el carcinoma hepatocelular (HCC)

A fin de ilustrar la viabilidad de esta estrategia, se analizaron los datos de secuenciación del ADN plasmático de un paciente con carcinoma hepatocelular (HCC) y de una mujer embarazada. A modo ilustrativo, el análisis se centró en el cromosoma 8. Puede aplicarse la misma estrategia al genoma completo o a cualquier otro cromosoma.

65 Se determinaron las coordenadas de los nucleótidos terminales en ambos extremos de cada fragmento de ADN

plasmático secuenciado. Posteriormente, se contó el número de fragmentos que terminaban en cada nucleótido del cromosoma 8. Se determinó el 1 millón de nucleótidos que tenían el mayor número de fragmentos de ADN que terminaban en ellos para el caso de HCC y la mujer embarazada. Se puede considerar que el millón superior está por encima de un umbral.

5 La FIG. 16 es un diagrama de Venn que muestra el número de sitios de terminación frecuentes que son específicos para el caso HCC, específicos para la mujer embarazada y compartidos por ambos casos. Las coordenadas de los 536.772 nucleótidos que fueron las posiciones finales más frecuentes específicas para el caso de HCC se muestran en el Apéndice A. Las coordenadas de los 536.772 nucleótidos que fueron las posiciones finales más frecuentes específicas para la mujer embarazada se enumeran en el Apéndice B. Se omiten las coordenadas de los 463.228 nucleótidos que fueron las posiciones finales más frecuentes compartidas por los dos casos.

15 Los presentes inventores razonaron que los fragmentos de ADN plasmático con un nucleótido terminal que termina exactamente en las posiciones finales específicas del HCC 536.772 tendrían más probabilidades de proceder del tumor. Según esta suposición, el número de fragmentos de ADN plasmático secuenciados que terminan en las posiciones finales específicas del HCC puede utilizarse para indicar la presencia o ausencia de HCC u otros cánceres que tengan el mismo patrón de fragmentación del ADN plasmático. En otra realización, este parámetro también puede utilizarse para reflejar el nivel de cáncer, por ejemplo, pero sin limitación, el tamaño del tumor, el estadio del cáncer, la carga tumoral y la presencia de metástasis.

20 En otra realización más, el número de fragmentos que terminan en las posiciones finales específicas del HCC puede correlacionarse con la concentración fraccional de ADN derivado del cáncer en el plasma para las muestras con una fracción de ADN tumoral conocida en el plasma. La fracción de ADN tumoral en el plasma puede determinarse mediante, por ejemplo, pero sin limitación, la cuantificación de las mutaciones del cáncer en el plasma o la magnitud de las aberraciones del número de copias en el ADN plasmático (Chan *et al.* Clin Chem 2013;59:211-24). Esta correlación puede utilizarse como curva de calibración (figura 1). Para pacientes con una fracción de ADN tumoral desconocida en el plasma, se puede determinar la cantidad de fragmentos de ADN que terminan en las posiciones finales específicas del HCC. Posteriormente, la fracción de ADN tumoral en el plasma puede determinarse basándose en la curva de calibración y en la cantidad de fragmentos de ADN que terminan en las posiciones finales específicas del HCC. En una implementación, la cantidad de fragmentos de ADN que terminan en las posiciones finales específicas de HCC puede normalizarse con respecto al número total de fragmentos de ADN secuenciados, el número total de lecturas alineables o el número de fragmentos de ADN alineados con determinadas regiones cromosómicas. Por tanto, la proporción de fragmentos de ADN secuenciados que terminan en posiciones específicas del cáncer puede utilizarse como parámetro.

35 La FIG. 17 muestra una curva de calibración que muestra la relación entre la proporción de fragmentos de ADN secuenciados que terminan en posiciones finales específicas para el cáncer y la fracción de ADN tumoral en plasma para pacientes con cáncer con fracciones conocidas de ADN tumoral en plasma. Este diagrama conceptual muestra una correlación de la curva de calibración entre la fracción de ADN tumoral y la proporción de fragmentos de ADN de la secuencia que terminan en las posiciones finales específicas del cáncer. Se puede determinar una curva de calibración ajustando los puntos de datos determinados a partir de las muestras de calibración, cuya fracción de ADN tumoral se determinó mediante otras técnicas.

45 En otra realización de la presente invención, se pueden determinar los patrones de fragmentación del ADN plasmático de los pacientes que padecen diferentes tipos de cáncer. Los extremos superpuestos de estos pacientes de cáncer pueden considerarse como extremos específicos del cáncer, mientras que las posiciones finales de los tipos de cáncer individuales pueden considerarse como específicas de un tipo de cáncer concreto. Para cualquier individuo sospechoso de parecer cáncer, los fragmentos de ADN plasmático secuenciados pueden compararse primero con las posiciones finales específicas del cáncer para determinar la probabilidad de que el individuo tenga un cáncer. Si el individuo es susceptible de padecer un cáncer, los fragmentos secuenciados pueden analizarse en busca de las posiciones finales específicas del tipo de cáncer para determinar el cáncer más probable que padece un individuo.

50 En otra realización de la presente invención, se pueden determinar las posiciones finales del ADN derivado de diferentes órganos y pueden utilizarse para determinar las contribuciones relativas del ADN de diferentes órganos al plasma.

B. Ejemplo fetal

60 En otra realización, esta estrategia puede utilizarse para determinar la fracción de ADN fetal en una muestra de plasma materno. Se puede establecer una curva de calibración determinando en primer lugar la correlación entre la proporción de fragmentos de ADN plasmático secuenciados que terminan en las posiciones finales específicas del embarazo y las fracciones de ADN fetal en una serie de muestras de plasma materno con una fracción de ADN fetal conocida. Las fracciones de ADN fetal pueden determinarse por varios métodos, por ejemplo, pero sin limitación la determinación de los alelos específicos del feto en la muestra, la cuantificación de objetivos en el cromosoma Y para los embarazos masculinos y el análisis de marcadores de metilación específicos del feto. Para una muestra de plasma de embarazada con una fracción de ADN fetal desconocida, se puede determinar la proporción de fragmentos de ADN plasmático

secuenciados que terminan en las posiciones finales específicas del embarazo. Utilizando esta información, la fracción de ADN fetal en la muestra de ADN plasmático analizada puede determinarse basándose en la curva de calibración.

C. Kit para el uso de posiciones finales preferidas

En algunas realizaciones, se proporciona un kit para analizar el ADN de una muestra biológica que contiene una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos. El kit puede comprender uno o más oligonucleótidos para hibridar específicamente con al menos una parte de la región genómica enumerada en los Apéndices A y B. En una realización, el kit incluye uno o más oligonucleótidos para hibridar específicamente con al menos una parte de la región genómica enumerada en el Apéndice A para su uso en la realización de pruebas para el HCC en un sujeto. En otra realización, el kit incluye uno o más oligonucleótidos para hibridar específicamente con al menos una sección de una región genómica enumerada en el Apéndice B para su uso en la realización de pruebas en una mujer embarazada, por ejemplo, para determinar una fracción de ADN fetal en una muestra biológica materna procedente de la mujer embarazada.

VI. ANÁLISIS DE LA POSICIÓN FINAL MEDIANTE POLIMORFISMOS

En algunas realizaciones, las regiones que tienen un patrón de fragmentación específico para cada tejido pueden identificarse mediante alelos específicos para cada tejido. Por ejemplo, se puede identificar un alelo específico del feto analizando una muestra de plasma materno y comparando los alelos detectados con los detectados en una muestra solo materna, como se describe en el presente documento. Puede identificarse que las posiciones genómicas que tienen una tasa elevada de moléculas de ADN fetal que terminan en ellas en relación con la tasa de los tejidos que presentan un alelo compartido (es decir, que se comparte con el feto y la madre) tienen un patrón de fragmentación específico del tejido fetal. Estas posiciones finales con preferencia fetal pueden ser o no sitios de hipersensibilidad a DNAsas, por lo que se demuestra que varias regiones genómicas pueden tener amplitudes específicas de tejido para los patrones de fragmentación, y las realizaciones no se limitan a los sitios de hipersensibilidad a DNAsas. Se puede hacer un análisis similar para una muestra de un sujeto que se somete a una prueba de detección de un tumor.

A. Ejemplo fetal

Las posiciones finales preferidas pueden obtenerse analizando el ADN plasmático de una mujer embarazada. Los fragmentos de ADN plasmático derivados del feto y de la madre pueden diferenciarse mediante métodos basados en polimorfismos. Los fragmentos que portan alelos específicos del feto y de la madre pueden utilizarse para determinar las posiciones finales preferidas del ADN procedente del feto y de la madre.

Para este estudio se reclutó a una mujer con un embarazo único masculino a las 38 semanas de gestación en el Departamento de Obstetricia y Ginecología, Prince of Wales Hospital, Hong Kong, habiéndose obtenido el consentimiento informado. Las muestras de sangre se centrifugaron a 1.600 g durante 10 min a 4 °C. La porción de plasma se recolectó y se volvió a centrifugar a 16.000 g durante 10 min a 4 °C para eliminar las células sanguíneas. La porción de células sanguíneas se volvió a centrifugar a 2.500 g y se eliminó el plasma residual. El ADN de las células sanguíneas y el del plasma materno se extrajo con el protocolo de sangre y fluidos corporales del QIAamp DNA Blood Mini Kit y el QIAamp DSP DNA Blood Mini Kit (Qiagen), respectivamente. El ADN de la placenta se extrajo con el QIAamp DNA Mini Kit (Qiagen) según el protocolo de tejidos del fabricante. Las bibliotecas de secuenciación se secuenciaron utilizando el protocolo de preparación de bibliotecas sin PCR de Illumina TruSeq. Los datos de la secuenciación *paired-end* se analizaron utilizando Short Oligonucleotide Alignment Program 2 (SOAP2) en el modo *paired-end* (Li *et al.* Bioinformatics 2009;25:1966-1967). Las lecturas *paired-end* se alinearon con el genoma humano de referencia sin máscara de repetición (Hg19). Se permitieron hasta 2 nucleótidos desparejados para la alineación de cada extremo. A continuación, se analizaron las coordenadas genómicas de estos alineamientos potenciales para los 2 extremos con el fin de determinar si alguna combinación permitiría alinear los 2 extremos en el mismo cromosoma con la orientación correcta, que abarcan un tamaño de inserción ≤ 600 pb, y que se cartografía en una única localización en el genoma humano de referencia. La muestra de plasma materno se secuenció con una profundidad de cobertura de 270x de un genoma humano haploide. Las células de la sangre materna, las células de la sangre paterna y del cordón umbilical fueron secuenciadas con una cobertura del genoma humano haploide de 40x, 45x y 50x, respectivamente, utilizando el mismo protocolo de secuenciación.

Para este fin, se analizaron las secuencias finales recurrentes en el ADN plasmático materno.

1. Identificación de las posiciones finales específicas del feto

Al realizar una secuenciación de muy alta profundidad de la muestra de ADN plasmático materno utilizando una biblioteca no amplificada mediante la PCR, los presentes inventores investigaron si pudiese haber sitios en los genomas materno y fetal que se escindirían preferencialmente en la generación del ADN plasmático. Para demostrar este efecto, se identificaron los locus de SNP informativos para los que la madre era homocigota (genotipo indicado como AA) y el feto era heterocigoto (genotipo indicado como AB). En este ejemplo ilustrativo, el alelo B sería específico del feto y el alelo A sería compartido por la madre y el feto. En la FIG. 18 se muestra un ejemplo representativo. Como control, se muestran los resultados de la secuenciación de una muestra de ADN obtenida de células sanguíneas y

fragmentada artificialmente mediante tratamiento con ultrasonidos.

Se observó un patrón de fragmentación no aleatorio en el ADN plasmático. Para el gráfico de la probabilidad de ser un extremo de los fragmentos de ADN, se observaron tres picos para cada uno de los dos grupos de fragmentos que llevan el alelo específico del feto y el compartido por la madre. Estos picos representan los puntos calientes de las posiciones finales del ADN procedente del feto y de la madre en el plasma materno, respectivamente. Las posiciones de los picos se superponen en gran medida entre estos dos grupos. Por el contrario, el patrón de fragmentación del ADN tratado con ultrasonidos parece ser aleatorio y la probabilidad de finalización del fragmento es similar en toda la región.

La FIG. 18 muestra un ejemplo ilustrativo de los patrones de fragmentación no aleatorios del ADN plasmático que portan un alelo específico del feto y un alelo compartido por la madre y el feto. En la parte superior de la figura, cada línea horizontal representa un fragmento de ADN secuenciado. Los extremos de los fragmentos de ADN representan la posición final de la lectura secuenciada. Los fragmentos se clasifican según la coordenada del nucleótido más externo de la izquierda (la coordenada genómica más pequeña). En la parte inferior de la figura, se muestra el porcentaje de fragmentos que terminan en una posición determinada. El eje X representa las coordenadas genómicas y el SNP se encuentra en el centro indicado por la línea de puntos.

Además, los presentes inventores buscaron coordenadas que tuvieran una mayor probabilidad de ser una posición final para los fragmentos de ADN plasmático. Centraron su búsqueda en los fragmentos que cubrían los SNP informativos para poder evaluar por separado los fragmentos portadores de alelos específicos del feto y los alelos compartidos por la madre y el feto. Los presentes inventores determinaron si ciertas ubicaciones dentro del genoma humano tenían una probabilidad significativamente mayor de ser una posición final de fragmentos de ADN plasmático utilizando una función de probabilidad de Poisson. Para el análisis de los SNP para los que la madre era homocigota (genotipo AA) y el feto era heterocigoto (genotipo AB), el alelo A sería el "alelo compartido" y el alelo B sería el alelo específico del feto. Se contaría el número de lecturas secuenciadas que portan el alelo compartido y el alelo específico del feto. En la distribución por tamaño molecular del ADN plasmático, se observaría un pico en 166 pb tanto para el ADN procedente del feto como para el derivado de la madre. Si la fragmentación del ADN plasmático es aleatoria, los dos extremos se distribuirían uniformemente en una región de 166 pb cadena arriba y 166 pb cadena abajo del SNP informativo.

Se puede calcular un valor de p para determinar si una posición concreta tiene una probabilidad significativamente mayor de ser un extremo para las lecturas que portan el alelo compartido o el alelo específico del feto, basándose en la función de probabilidad de Poisson.

$$\text{valor de } p = \text{Poisson}(N_{\text{real}}, N_{\text{predicha}})$$

donde Poisson() es la función de probabilidad de Poisson; N_{real} es el número real de lecturas que terminan en el nucleótido concreto; y N_{predicha} es el número total de lecturas dividido entre 166. Se utilizó un valor de $p < 0,01$ como punto de corte para definir las posiciones finales preferidas para las lecturas portadoras del alelo específico del feto o del alelo compartido. Se determinaron las posiciones finales estadísticamente significativas para los fragmentos de ADN que llevan el alelo compartido y el alelo específico del feto de forma independiente (FIG. 19). Pueden utilizarse otras distribuciones de probabilidad, por ejemplo, la distribución binomial, la distribución binomial negativa y la distribución normal.

La FIG. 19 muestra una gráfica de la probabilidad de que una coordenada genómica sea una posición final de los fragmentos de ADN plasmático materno en una región con un SNP informativo. Los resultados de las posiciones de los nucleótidos con una probabilidad significativamente mayor de ser un extremo de los fragmentos de ADN plasmático que llevan un alelo compartido y un alelo específico del feto se muestran en rojo y azul, respectivamente. El eje X representa las coordenadas genómicas y la mutación se sitúa en el centro indicado por la línea de puntos. Como se muestra, hay coordenadas que tienen una alta tasa de aparición de posiciones finales solo para el alelo específico del feto, solo para el alelo compartido, y algunos son comunes a ambos.

Los presentes inventores identificaron un total de 4.131 (conjunto A) y 10.021 (conjunto B) posiciones de nucleótidos con una probabilidad significativamente mayor de ser un extremo de fragmentos de ADN plasmático portadores de alelos específicos del feto y de alelos compartidos, respectivamente. El conjunto C fue el conjunto superpuesto y contenía 4.258 posiciones de nucleótidos (Fig. 3). Estas posiciones finales se obtuvieron a partir de regiones que abarcan un total de 1,42 Mb y cubren 4.303 SNP. Por tanto, las posiciones finales preferidas para los fragmentos específicos del feto representaron el 0,29 % de las regiones analizadas. Había 24.500, 22.942 y 31.925 fragmentos de ADN plasmático portadores de alelos específicos del feto que terminaban en las posiciones del conjunto A, del conjunto B y del conjunto C, respectivamente. Había 27.295, 158.632 y 87.804 fragmentos de ADN plasmático portadores de alelos compartidos que terminaban en las posiciones del conjunto A, del conjunto B y del conjunto C, respectivamente. Se espera que el número o la prevalencia de las posiciones finales preferidas sea mucho mayor y se produzca en otras coordenadas genómicas.

La estrategia basada en el polimorfismo, tal y como se describe en el presente documento, solo identifica las

posiciones finales preferidas que están asociadas a un SNP informativo para este par fetal-materno. Por tanto, los extremos preferidos identificados representarían un subconjunto de dichos extremos en el genoma. Los presente inventores han desarrollado estrategias que no se basan en el polimorfismo para identificar los extremos preferidos. De hecho, se identificaron muchas más estrategias de finalización preferidas utilizando las estrategias no basadas en polimorfismos. Consulte otros experimentos descritos a continuación.

La FIG. 20 muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático a lo largo de los SNP que eran homocigóticos en la madre y heterocigóticos en el feto. El conjunto A incluía las posiciones finales preferidas para los fragmentos portadores de alelos específicos del feto. El conjunto B incluía las posiciones finales preferidas para los fragmentos portadores de alelos compartidos. El conjunto C incluía las posiciones finales preferidas para ambos tipos de fragmentos de ADN plasmático.

Usando el mismo principio, los presentes inventores analizaron además las posiciones terminales de los fragmentos de ADN de origen materno a lo largo de los SNP que eran heterocigotos para la madre (genotipo AB) y homocigotos en el feto (genotipo AA). Los presentes inventores identificaron un total de 7.527 (conjunto X) y 18.829 (conjunto Y) posiciones de nucleótidos con una probabilidad significativamente mayor de ser una posición final de fragmentos de ADN plasmático portadores de alelos específicos del feto y de alelos compartidos, respectivamente. El conjunto Z es el conjunto superpuesto y contenía 10.534 posiciones (fig. 4). Estas posiciones finales se obtuvieron a partir de regiones que abarcan un total de 3,1 Mb y cubren 9.489 SNP. Por tanto, las posiciones finales preferidas para los fragmentos específicos de la madre representaron el 0,24 % de las regiones analizadas para esta pareja de madre y feto. Había 69.136, 82.413 y 121.607 fragmentos de ADN plasmático portadores de alelos específicos maternos que terminaban en las posiciones del conjunto X, del conjunto Y y del conjunto Z, respectivamente. Había 46.554, 245.037 y 181.709 fragmentos de ADN plasmático portadores de alelos compartidos que terminaban en las posiciones del conjunto X, del conjunto Y y del conjunto Z, respectivamente. De nuevo, este análisis se centra en las moléculas de ADN plasmático que cubren al menos un SNP informativo, los extremos preferidos identificados solo representan un subconjunto de dichos extremos no aleatorios en todo el genoma.

La FIG. 21 muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático a lo largo de los SNP que eran homocigóticos en el feto y heterocigóticos en la madre. El conjunto X incluía las posiciones finales preferidas para los fragmentos portadores de alelos específicos de la madre. El conjunto Y incluía las posiciones finales preferidas para los fragmentos que llevan alelos compartidos. El conjunto Z incluía las posiciones finales preferidas para ambos tipos de fragmentos de ADN plasmático.

2. Uso de posiciones finales recurrentes para deducir la fracción de ADN fetal

Tras la identificación de las posiciones finales recurrentes de los fragmentos de ADN plasmático derivados de la madre y del feto, los presentes inventores razonaron que la abundancia relativa de ADN plasmático que termina en estos conjuntos de posiciones de nucleótidos reflejaría la fracción de ADN fetal. Para confirmar esto, los presentes inventores secuenciaron el ADN plasmático de 26 mujeres en el primer trimestre del embarazo (10~13 semanas), cada una de ellas portadora de un feto masculino. La mediana del recuento de lecturas cartografiadas fue de 16 millones (rango: 12-22 millones). La proporción de lecturas secuenciadas que se alinean con el cromosoma Y se utilizó para calcular la fracción real de ADN fetal en cada muestra de plasma. Se pudo observar una correlación positiva entre la abundancia relativa (indicada como relación F/M) del ADN plasmático con extremos fetales (conjunto A) y maternos (conjunto X) recurrentes y la fracción de ADN fetal ($R = 0,63$, $P = 0,0004$, correlación de Pearson, FIG. 22). Es interesante que, aunque las posiciones finales preferidas se identificaron a partir de SNP informativos para una pareja de feto y madre y solo representaron un subconjunto de dichos finales en el genoma, los extremos identificados también eran relevantes para otros embarazos y la correlación con la fracción fetal se logró incluso con solo este subconjunto de extremos preferidos.

La FIG. 22 muestra una correlación entre la abundancia relativa (Ratio (F/M)) de las moléculas de ADN plasmático con extremos fetales (Conjunto A) y maternos (Conjunto X) recurrentes y la fracción de ADN fetal. Cada uno de los puntos de datos puede corresponder a una muestra de calibración respectiva, por lo que se consideran puntos de datos de calibración. La línea que ajusta los puntos de datos de calibración es un ejemplo de función de calibración.

Se pueden utilizar otros conjuntos además del conjunto A y el conjunto X. Por ejemplo, se puede tomar una proporción (u otra abundancia relativa o una función de una proporción) del conjunto A con respecto al conjunto C y del conjunto A con respecto al conjunto B. Como otro ejemplo, se puede tomar una proporción del conjunto X y del conjunto Z o una proporción entre el conjunto X y el conjunto Y, lo que proporcionaría una fracción de ADN materno, que se puede suponer que es la inversa de la fracción de ADN fetal. En dicho ejemplo, el tejido materno puede ser un primer tipo de tejido cuya contribución proporcional está determinada, aunque sea de forma implícita.

3. Uso del tamaño

La distribución del tamaño molecular de los fragmentos de ADN plasmático que terminan en las posiciones finales específicas del feto proporciona más pruebas de que las posiciones son específicas del feto. Para justificar adicionalmente que las posiciones del conjunto A y del conjunto X fueron los sitios finales preferidos para los

fragmentos de ADN derivados del feto y de la madre, respectivamente, los presentes inventores compararon las distribuciones de tamaño molecular del ADN plasmático que termina en estos dos conjuntos de posiciones. Para la muestra de la que proceden estas posiciones, la distribución de tamaños fue más corta para los fragmentos que terminan en las posiciones del conjunto A que para los que terminan en las posiciones del conjunto X (FIG. 23A).

5 La FIG. 23A muestra las distribuciones del tamaño del ADN plasmático para los fragmentos que terminan en las posiciones finales con preferencia fetal (conjunto A) (en azul) y los fragmentos que terminan en las posiciones finales con preferencia materna (conjunto X) (en rojo). Se observó una distribución de tamaños más corta para los fragmentos que terminan en las posiciones del conjunto A en comparación con los que terminan en las posiciones del conjunto X. La FIG. 23B muestra la gráfica acumulativa de las distribuciones de tamaño para los dos conjuntos de fragmentos. La FIG. 23C muestra la diferencia de las frecuencias acumuladas de los dos conjuntos de fragmentos (ΔS) frente al tamaño del fragmento. La FIG. 23D muestra el ΔS frente al tamaño con el desplazamiento de las posiciones finales del conjunto A y del conjunto X a posiciones con coordenadas genómicas mayores de cero a 5 pb. La FIG. 23E muestra el ΔS frente al tamaño con el desplazamiento de las posiciones finales del conjunto A y del conjunto X de cero a 5 pb en sentido inverso (posiciones con coordenadas genómicas más menores).

Para cuantificar adicionalmente la diferencia en la distribución de tamaños, se representan las frecuencias acumuladas de las dos curvas (FIG. 23B). La diferencia en las dos curvas, representada por ΔS , se representa en la FIG. 23C. Observamos que la máxima diferencia se observó en 166 pb. Esto concuerda con los informes anteriores de que la diferencia máxima entre el ADN procedente del feto y el de la madre se podía observar en 166 pb (Yu *et al.* Proc Natl Acad Sci USA. 2014;111:8583-8). Los presentes hallazgos sugieren que existe un enriquecimiento del ADN de origen fetal para los fragmentos que terminan en las posiciones finales con preferencia fetal (conjunto A) en comparación con los que terminan en las posiciones finales con preferencia materna (conjunto X).

25 Además, los presentes inventores investigaron la especificidad de estas posiciones finales desplazando las posiciones finales del conjunto A y del conjunto X de 1 a 5 pb cadena arriba o cadena abajo del genoma. Los valores de ΔS se representan frente al tamaño molecular con el desplazamiento de las posiciones finales del conjunto A y del conjunto X en ambas direcciones (FIG. 23D y 23E). Los números positivos del desplazamiento representan el desplazamiento a una posición con una coordenada genómica mayor (FIG. 23D) y los números negativos del desplazamiento representan el desplazamiento a una posición con una coordenada genómica menor (FIG. 23E). El desplazamiento de las posiciones con preferencia fetal y materna, incluso en 1 pb, reduciría significativamente la diferencia de tamaño entre los fragmentos de ADN que terminan en estos dos conjuntos de posiciones (ΔS). El desplazamiento de 5 pb eliminó casi por completo la diferencia de tamaño. Estos resultados sugieren que las lecturas que terminan en esas posiciones alternativas no son tan específicas del feto o de la madre como las lecturas que terminan en esas posiciones finales preferidas identificadas por nuestro algoritmo. Estos datos apoyan aún más nuestra interpretación de que las moléculas de ADN plasmáticas o libres de células se fragmentan o se escinden de forma muy precisa en esas posiciones finales preferidas. En otras palabras, el proceso de fragmentación del ADN no aleatorio y libre de células es preciso hasta el nivel de nucleótidos específicos.

40 Posteriormente, los presentes inventores analizaron las lecturas secuenciadas agrupadas de las 26 muestras de plasma del primer trimestre utilizadas para el análisis de la fracción de ADN fetal. Se observó una distribución de tamaño más corta para los fragmentos que terminan en las posiciones del conjunto A en comparación con los que terminan en las posiciones del conjunto X (FIG. 24A).

45 La FIG. 24A muestra las distribuciones del tamaño del ADN plasmático en una muestra conjunta de ADN plasmático de 26 mujeres en el primer trimestre del embarazo para los fragmentos que terminan en las posiciones finales con preferencia fetal (conjunto A) (en azul) y los fragmentos que terminan en las posiciones finales con preferencia materna (conjunto X) (en rojo). Se observó una distribución de tamaños más corta para los fragmentos que terminan en las posiciones del conjunto A en comparación con los que terminan en las posiciones del conjunto X. La FIG. 24B muestra la gráfica acumulativa de las distribuciones de tamaño para los dos conjuntos de fragmentos. La FIG. 24C muestra la diferencia de las frecuencias acumuladas de los dos conjuntos de fragmentos (ΔS) frente al tamaño del fragmento. La FIG. 24D muestra el ΔS frente al tamaño con el desplazamiento de las posiciones del conjunto A y del conjunto X de cero a 5 pb (coordenadas genómicas mayores). La FIG. 24E muestra el ΔS frente al tamaño con el desplazamiento de las posiciones del conjunto A y del conjunto X de cero a 5 pb en sentido inverso (coordenadas genómicas más pequeñas). La diferencia de tamaño entre los fragmentos de ADN plasmático que terminan en los dos conjuntos de posiciones (ΔS) se reduciría con el desplazamiento de estas posiciones, indicando que estas posiciones serían precisas a nivel de un solo nucleótido.

60 B. Ejemplo de cáncer

La misma estrategia puede aplicarse también para el análisis de las posiciones finales preferidas para los fragmentos derivados del cáncer. En este ejemplo, los presentes inventores secuenciaron el plasma (cobertura 220x), la capa leucocitaria (48x) y el tejido tumoral (45x) de un paciente que padece carcinoma hepatocelular (HCC). El perfil mutacional del paciente se obtuvo comparando los genotipos del tejido tumoral y de la capa leucocitaria. Para determinar las posiciones finales preferidas para los fragmentos de ADN plasmático derivados del cáncer, los presentes inventores analizaron los fragmentos de ADN plasmático portadores de las mutaciones cancerígenas. Como

se muestra en las FIG. 24A-24E, el patrón de fragmentación del ADN plasmático en el paciente con HCC no es aleatorio. Ciertas posiciones de nucleótidos tienen una mayor probabilidad de ser un extremo de un fragmento de ADN plasmático.

5 **1. Identificación de posiciones finales específicas para el cáncer**

10 La FIG. 25 muestra un ejemplo ilustrativo de los patrones de fragmentación no aleatorios del ADN plasmático del paciente con HCC. En la parte superior de la figura, cada línea horizontal representa un fragmento de ADN secuenciado. Las líneas rojas y azules representan fragmentos de ADN portadores de los alelos de tipo silvestre y mutante, respectivamente. Los extremos de los fragmentos de ADN representan la posición final de la lectura secuenciada. Los fragmentos se clasifican según la coordenada del nucleótido más externo de la izquierda (la coordenada genómica más pequeña). En la parte inferior de la figura, se muestra el porcentaje de fragmentos que terminan en una posición determinada. El eje X representa las coordenadas genómicas y la mutación se sitúa en el centro indicado por la línea de puntos.

15 Los presentes inventores identificaron las posiciones genómicas que tienen una mayor probabilidad de ser un extremo de los fragmentos de ADN plasmático que portan alelos mutantes y alelos de tipo silvestre utilizando la función de distribución de probabilidad de Poisson, tal y como se ha descrito anteriormente. Se utilizó un valor de p de 0,01 como umbral. Lo contrario también es cierto, como se describe en la Solicitud PCT n.º PCT/CN2016/073753, es decir, cuando se identifica una molécula de ADN plasmática con un extremo específico, el alelo o la mutación del SNP en la molécula tiene más probabilidades de proceder del cáncer, estar asociada a la enfermedad o al embarazo, dependiendo del conjunto de extremos que se haya utilizado en la interpretación de los datos del ADN plasmático.

20 La FIG. 26 es una gráfica de la probabilidad de que una coordenada genómica sea una posición final de los fragmentos de ADN plasmático en una región con un sitio de mutación. Los resultados de las posiciones de los nucleótidos con una probabilidad significativamente mayor de ser un extremo de los fragmentos de ADN plasmático que llevan un alelo de tipo silvestre y un alelo mutante se muestran en rojo y azul, respectivamente. El eje X representa las coordenadas genómicas y la mutación se sitúa en el centro indicado por la línea de puntos. Como se muestra, hay coordenadas que tienen una alta tasa de aparición de posiciones finales solo para el alelo específico del mutante, solo para el alelo de tipo silvestre, y algunos son comunes a ambos.

25 La FIG. 27A muestra un análisis de las posiciones finales de los fragmentos de ADN plasmático en las posiciones genómicas en las que había mutaciones en el tejido tumoral. El conjunto E incluía las posiciones finales preferidas para los fragmentos portadores de alelos mutantes. El conjunto F incluía las posiciones finales preferidas para los fragmentos portadores de alelos de tipo silvestre. El conjunto G incluía las posiciones finales preferidas para ambos tipos de fragmentos de ADN plasmático.

30 **2. Uso de las posiciones finales recurrentes para deducir la fracción de ADN del tumor**

40 Como las posiciones del conjunto E eran los sitios finales preferidos para el ADN derivado del cáncer y las posiciones del conjunto F eran los sitios finales preferidos para el ADN de fondo derivado predominantemente de tejidos no tumorales, los presentes inventores barajaron la hipótesis de que la proporción entre los fragmentos que terminan en estos dos conjuntos de posiciones se correlacionaría con el ADN derivado del tumor. Por tanto, los presentes inventores analizaron el plasma de 71 pacientes con HCC cuyo plasma contenía al menos un 1 % de ADN derivado del tumor. En estos pacientes se analizó previamente la presencia de aberraciones en el número de copias del ADN plasmático y se estimaron las fracciones de ADN tumoral en función de la magnitud de las aberraciones de número de copias. (Jiang *et al.* Proc Natl Acad Sci USA. 2015;112:E1317-25). La relación entre los fragmentos que terminan en estos dos conjuntos de posiciones ($Relación_{M/T}$) se define como:

50
$$Relación_{M/T} = \frac{N.º \text{ de fragmentos de ADN plasmático que terminan en las posiciones del Conj.E}}{N.º \text{ de fragmentos de ADN plasmático que terminan en las posiciones del Conj.F}}$$

55 La FIG. 27B muestra una correlación entre la $Relación_{M/T}$ y la fracción de ADN tumoral en el plasma de 71 pacientes con HCC. Se observó una correlación positiva entre la $Relación_{M/T}$ y la fracción de ADN tumoral en el plasma ($r = 0,53$, $p < 0,001$, correlación de Pearson). Estos resultados sugieren que el número de fragmentos que terminan en estas posiciones finales preferidas por el cáncer sería útil para predecir la cantidad de ADN derivado del tumor en el plasma de los pacientes con cáncer.

60 Algunas realizaciones pueden aumentar el número de fragmentos de ADN de cáncer informativos accesibles mediante la detección combinada de una variedad de cambios específicos del cáncer o asociados al mismo, por ejemplo, mutaciones de un único nucleótido, en combinación con firmas de metilación del ADN específicas para el cáncer o asociadas a él (por ejemplo, la localización de la 5-metilcitosina y la hidroximetilación), moléculas cortas de ADN plasmático específicas del cáncer o asociadas al cáncer, marcas de modificación de las histonas específicas del cáncer o asociadas al cáncer, y ubicaciones de los extremos del ADN plasmático específicas del cáncer o asociadas al cáncer. Ciertos cambios específicos o asociados al cáncer pueden utilizarse como criterios de filtrado en la identificación de mutaciones.

VII. ANÁLISIS DE POSICIÓN FINAL INDEPENDIENTE DEL POLIMORFISMO

En otras realizaciones, las posiciones finales preferidas pueden obtenerse (A) comparando las posiciones finales de los fragmentos de ADN plasmático de diferentes individuos o (B) comparando las posiciones finales de los fragmentos de ADN plasmático de las muestras de un individuo tomadas en diferentes momentos.

A. Comparación entre las posiciones finales preferidas en sujetos que sufren diferentes condiciones patológicas y fisiológicas

1. Uso de conjuntos exclusivos por encima del umbral

Basándose en la función de probabilidad de la distribución de Poisson, los presentes inventores han identificado las posiciones genómicas que tenían una mayor probabilidad de ser posiciones finales de fragmentos plasmáticos para la mujer embarazada y el paciente con HCC descritos en las secciones anteriores. En este análisis, la hipótesis nula es que todos los fragmentos de ADN plasmático se fragmentarían al azar, de modo que cada posición genómica tendría la misma probabilidad de ser el final de los fragmentos de ADN plasmático. Se supuso que los fragmentos de ADN plasmático tenían un tamaño medio de 166 pb. El valor de p se calcula como

$$\text{valor de } p = \text{Poisson}(N_{\text{real}}, N_{\text{predicha}})$$

donde Poisson() es la función de probabilidad de Poisson; N_{real} es el número real de lecturas que terminan en el nucleótido concreto; y $N_{\text{predicha}} = \frac{\text{Número total de lecturas}}{3 \times 10^9 \times 166}$, el 3×10^9 en el denominador representa el número de nucleótidos en un genoma.

Se ajustó el valor de p mediante la corrección de Benjamini y Hochberg (Benjamini *et al.* Journal of the Royal Statistical Society, 1995;57:289-300) para conseguir una tasa de falsos descubrimientos (FDR) esperada <1 %.

La FIG. 28A muestra el número de posiciones finales preferidas para el ADN plasmático de la mujer embarazada y del paciente con HCC. El conjunto P contenía 29 millones de posiciones finales que eran las preferidas en la mujer embarazada. El conjunto Q contenía 6 millones de posiciones finales que se preferían en el paciente con HCC. El conjunto S es el conjunto superpuesto y contiene 15 millones de posiciones finales.

La hipótesis de los presentes inventores es que los fragmentos que terminan en las posiciones finales preferidas del HCC (conjunto Q) estarían enriquecidos para el ADN derivado del cáncer en comparación con los fragmentos que terminan en las posiciones finales preferidas del embarazo (conjunto P).

Por tanto, los presentes inventores calcularon la Relación_{HCC/Emb} como

$$\text{Relación}_{\text{HCC/Emb}} = \frac{N.^{\circ} \text{ de fragmentos de ADN plasmático que terminan en las posiciones del Conj. Q}}{N.^{\circ} \text{ de fragmentos de ADN plasmático que terminan en las posiciones del Conj. P}}$$

y correlacionaron esta proporción con la fracción de ADN tumoral en los 71 pacientes con HCC mencionados anteriormente.

La FIG. 28B muestra que se observó una correlación positiva entre la Relación_{HCC/Emb} y la fracción de ADN tumoral en plasma para los 71 pacientes con HCC. Estos resultados sugieren que el número o la proporción de fragmentos que terminan en los sitios de terminación preferidos de una afección específica podría ser útil para detectar la afección o para cuantificar la cantidad de ADN liberado del órgano enfermo.

2. Utilización del conjunto de posiciones genómicas con mayor índice de finalización

En otra realización, los sitios de terminación preferidos pueden identificarse determinando la relación entre el número de fragmentos que terminan en dicha posición y el número de fragmentos que cubren la posición, pero no terminan en ella. La FIG. 29A ilustra el cálculo de la relación de terminación preferida (PETR).

$$\text{PETR} = \frac{\text{Número de fragmentos de ADN que terminan en el nucleótido}}{\text{Número de fragmentos de ADN que abarcan el nucleótido pero no terminan en él}}$$

La FIG. 29A muestra una ilustración del concepto de la PETR. Cada línea representa un fragmento de ADN plasmático. Estos fragmentos están marcados de a a g. Los fragmentos a, b, c y d terminaron en el nucleótido de interés. Los fragmentos e, f y g cubren el nucleótido de interés, pero no terminan en dicha posición. En este ejemplo ilustrativo, la PETR es igual a 4/3, es decir, 1,33. En otras realizaciones, el denominador puede ser el número de fragmentos de ADN que abarcan el nucleótido, independientemente de si el fragmento de ADN termina en la posición.

El cálculo de la PETR puede utilizarse para identificar las posiciones de los nucleótidos que son extremos preferidos en los individuos que padecen diferentes enfermedades. El siguiente ejemplo demuestra la utilidad de la PETR. Se

compararon las muestras de plasma del paciente con HCC mencionado anteriormente y de un sujeto con infección crónica por el virus de la hepatitis B (VHB) pero sin cáncer (portador del VHB). Las muestras de ADN plasmático del portador del VHB se secuenciaron con una cobertura del genoma haploide de 215x. La PETR se calculó para cada posición genómica de cada sujeto. se identificaron 7.350.067 posiciones genómicas (conjunto H) con una PETR al menos 4 veces mayor en el paciente con HCC en comparación con el portador del VHB. Estas posiciones tenían una probabilidad al menos 4 veces mayor de ser un extremo de fragmentos de ADN plasmático en el paciente con HCC en comparación con el portador del VHB. Pueden utilizarse otros múltiplos de diferencia, por ejemplo, 1,5 veces, 2 veces y 3 veces.

Las muestras de plasma de 11 pacientes de HCC independientes se secuenciaron además con una profundidad de secuenciación mucho menor. Se obtuvo una media de 28 millones de lecturas secuenciadas de estas 11 muestras de plasma. Se calculó la PETR media en las posiciones 7.350.067 del conjunto H para cada uno de estos 11 pacientes con HCC y se correlacionó con la fracción de ADN tumoral en el plasma. La fracción de ADN tumoral en el plasma se calculó a partir de la magnitud de las aberraciones del número de copias en el plasma, como se describió anteriormente (Chan *et al.* Proc Natl Acad Sci USA. 2015;112:E1317-25).

La FIG. 29B muestra una correlación entre la fracción de ADN tumoral en plasma con la PETR en las posiciones del conjunto H en 11 pacientes con HCC. Se pudo observar una correlación positiva entre los dos parámetros, lo que sugiere que el promedio de la PETR en las posiciones preferidas del HCC (conjunto H) sería útil para indicar la cantidad de ADN tumoral en el plasma.

3. Confirmación de que la posición final está relacionada con el hígado

Para demostrar que las posiciones finales preferidas presentes en la muestra de ADN plasmático del HCC o en la muestra de ADN plasmático del VHB estaban relacionadas con el hígado, buscamos su presencia en muestras de plasma recogidas de pacientes antes y después de la extirpación quirúrgica del HCC. Los datos se muestran en la tabla 1. Las muestras antes y después de la cirugía se secuenciaron con coberturas genómicas haploides de 17x y 20x, respectivamente.

La tabla 1 muestra las posiciones finales preferidas por el HCC y las posiciones finales preferidas por el VHB en la muestra de plasma recogida antes y después de la cirugía para extirpar el tumor del hígado en el paciente con HCC.

	Sitios de finalización preferidos por el HCC	Sitios de finalización preferidos por el VHB
Sitios de finalización preferidos antes de la cirugía en HCC 1	92	16
Sitios de finalización preferidos después de la cirugía en HCC 1	5	4

Como puede observarse en la tabla 1, hay reducciones en el número de posiciones finales preferidas tanto por el HCC como por el VHB. Los datos del VHB sugieren que la mayoría de las posiciones finales preferidas son de origen hepático y su reducción se debe a la disminución de la masa de células hepáticas tras la cirugía. Por lo tanto, se reduce la liberación de moléculas de ADN acelular derivadas del hígado en el plasma. Es interesante observar que en la muestra prequirúrgica hay más de 5 veces más posiciones finales preferidas por el HCC que desaparecieron después de la cirugía. Algunos de los extremos preferidos que mostraron desaparición después de la cirugía proceden del hígado. La observación de que en la misma muestra prequirúrgica se detectaron muchas más terminaciones preferidas por el HCC que por el VHB sugiere que la mayoría de esas terminaciones son específicas del HCC y no solo están genéricamente asociadas al hígado.

Hay una serie de aplicaciones que podrían derivarse de estos datos. Los datos indican que la detección de los extremos preferidos del ADN acelular o del ADN plasmático podría utilizarse para el seguimiento del tratamiento del cáncer. Por ejemplo, la reducción posquirúrgica de los extremos preferidos indica el éxito de la extirpación quirúrgica del HCC. Si el tumor no fue extirpado completamente o con éxito, la cantidad de extremos preferidos de ADN plasmático no mostraría una reducción sustancial después de la cirugía. Esto se debe a que los focos tumorales o metastásicos restantes serían una fuente para la liberación continuada de ADN acelular o de ADN plasmático con las posiciones finales preferidas por el HCC. Los datos muestran que el seguimiento del tratamiento basado en el análisis de los extremos preferidos del ADN acelular podría lograrse con una profundidad de secuenciación relativamente baja.

Los datos también muestran que las posiciones finales preferidas del ADN plasmático asociado a los tejidos o al cáncer podrían utilizarse para identificar el tejido de la patología, incluido el tejido que alberga el cáncer. Por ejemplo, se podrían utilizar múltiples conjuntos de extremos preferidos de ADN acelular que se derivan de diferentes órganos. Así se podrían determinar las cantidades relativas de ADN acelular procedentes de diversos tejidos. Por tanto, esto podría servir como estrategia para la deconvolución de tejidos de ADN acelular. El tejido que, según esta estrategia, presenta la mayor desviación (aumento significativo o reducción significativa) respecto a los valores de referencia establecidos

a partir de las muestras de control, sería el órgano o tejido con la patología (por ejemplo, inflamación o infección viral como en el caso del portador del virus de la hepatitis B crónica) o el cáncer.

5 Otra prueba para demostrar que los extremos preferidos del ADN plasmático del HCC son específicos del cáncer o del HCC, los presentes inventores estudiaron el perfil de tamaño molecular de las moléculas de ADN plasmático que muestran los extremos preferidos por el HCC o el VHB (FIG. 30).

10 La FIG. 30 muestra la proporción de ADN corto (<150 pb) detectado entre las moléculas de ADN plasmático que terminan con extremos preferidos por el HCC, extremos preferidos por el VHB o los extremos compartidos. La FIG. 30 muestra que las moléculas de ADN plasmático que muestran los extremos preferidos por el HCC son generalmente mucho más cortas (alta proporción de ADN corto) que las que muestran los extremos preferidos por el VHB. Jiang et al (Jiang *et al.* Proc Natl Acad Sci USA. 2015;112:E1317-25) utilizaron previamente otro enfoque para demostrar que las moléculas de ADN plasmático derivadas del tumor son más cortas que el ADN no tumoral de fondo. Dado que las moléculas de ADN plasmático con los extremos preferidos por el HCC son mucho más cortas, es muy probable que sean de origen tumoral. Por tanto, se podría mejorar la posibilidad de detectar las moléculas de ADN plasmático con los extremos preferidos por el HCC con una profundidad de secuenciación aún menor, se puede enriquecer la muestra con ADN corto.

20 4. Tasa de finalización basada en la ventana

En otra realización, las posiciones preferentes de HCC pueden ampliarse para incluir los nucleótidos vecinos. La FIG. 31A ilustra este método. Se determinaría la relación PETR basada en la ventana (w-PETR) entre el número de fragmentos que terminan en la ventana A y los que terminan en la ventana B. El tamaño de la ventana A y de la ventana B puede ajustarse para conseguir el rendimiento deseado. El rendimiento de los diferentes tamaños de las ventanas puede obtenerse experimentalmente. Se puede ajustar el tamaño de la ventana A, por ejemplo, pero sin limitación, a 5 pb, 6 pb, 7 pb, 8 pb, 9 pb, 10 pb, 15 pb, 20 pb, 25 pb y 30 pb. El tamaño de la ventana B será mayor que el de la ventana A y puede ajustarse, por ejemplo, pero sin limitación, a 20 pb, 25 pb, 30 pb, 40 pb, 50 pb, 60 pb, 70 pb, 80 pb, 100 pb, 120 pb, 140 pb, 160 pb, 180 pb y 200 pb. En el siguiente ejemplo ilustrativo, los tamaños de la Ventana A y la Ventana B se fijaron en 20 pb y 150 pb, respectivamente.

30 La FIG. 31A muestra una ilustración del principio de la w-PETR. El valor de w-PETR se calcula como la relación entre el número de fragmentos de ADN que terminan dentro de la ventana A y la ventana B. La ventana A es de mayor tamaño y puede tener una amplitud de uno cuando se implementa PETR estándar. La ventana B se muestra más grande. Ambas ventanas se muestran centradas en la posición final preferida, pero se puede utilizar otro posicionamiento de las ventanas. En algunas realizaciones, la ventana A puede corresponder a una ventana final preferida.

35 La FIG. 31B muestra una correlación entre la fracción de ADN tumoral y el valor de la w-PETR en los 11 pacientes con HCC. Estos resultados sugieren que la w-PETR sería útil para determinar la cantidad de ADN derivado del tumor en el plasma de los pacientes con cáncer.

40 5. Uso de las posiciones finales más altas por muestra

Los presentes inventores compararon el millón de posiciones finales de ADN acelular más frecuentemente representadas entre los datos de una mujer embarazada, un portador crónico del virus de la hepatitis B (VHB), un paciente con cáncer de pulmón y dos con HCC. Para los pacientes con HCC, la biblioteca de secuenciación de un caso (HCC) se preparó mediante un protocolo sin PCR y la otra muestra (HCC PCR) se preparó mediante un protocolo basado en la PCR. Todas las demás muestras se preparan mediante un protocolo sin PCR. La FIG. 32 muestra la proporción de posiciones finales preferidas comúnmente compartidas detectadas en muestras de plasma de cada una de las muestras estudiadas cuando se comparan con una muestra de plasma sanguíneo del cordón umbilical (cobertura del genoma haploide 210x).

50 La FIG. 32 muestra la proporción de posiciones finales preferidas comúnmente compartidas detectadas en muestras de plasma de cada una de las muestras estudiadas cuando se comparan con una muestra de plasma sanguíneo del cordón umbilical (cobertura del genoma haploide 210x). Se muestran los porcentajes de los autosomas para cada uno de embarazo, HCC, VHB, cáncer de pulmón y HCC detectados utilizando la PCR.

55 El alto nivel de coincidencia apoya de nuevo el concepto de que la fragmentación del ADN plasmático no es un proceso aleatorio. Los datos del HCC y del HCC(PCR) muestran que el análisis de la posición final preferida podría realizarse utilizando cualquiera de los protocolos de preparación de bibliotecas con o sin PCR. Es interesante observar que todavía hay una proporción de moléculas de ADN plasmáticas que no muestran extremos comunes. Los extremos no comunes son los extremos preferidos representativos del estado fisiológico, por ejemplo, el embarazo, el feto o la placenta para la muestra; o el estado patológico, por ejemplo, cáncer. En la FIG. 33 se muestra una comparación más detallada de los extremos preferidos del ADN plasmático.

60 La FIG. 33 muestra un diagrama de Venn en el que se indica el número de posiciones finales preferidas que se

observan comúnmente en dos o más muestras, así como las que solo se observaron en una muestra. Se secuenció el ADN plasmático de un paciente con cáncer de pulmón con una cobertura del genoma haploide de 175x.

5 Cabe destacar de la FIG. 33 que 115.305 extremos preferidos son comunes en las tres muestras. Es probable que procedan de la principal fuente de ADN plasmático de fondo, por ejemplo, células sanguíneas. El análisis también muestra que había 61.035 posiciones finales preferidas observadas en las muestras de plasma del paciente con HCC y del paciente con cáncer de pulmón. Estos extremos preferidos pueden ser comunes a varios tipos de cáncer. Por tanto, son derivados del cáncer. Sin embargo, hay extremos que solo se detectaron en las moléculas de ADN plasmático del paciente con HCC (479.766 extremos) o del paciente con cáncer de pulmón (749.237 extremos) pero no en ambos. Por lo tanto, estos extremos preferidos muestran un mayor nivel de especificidad. Son específicos para un tipo de tejido canceroso concreto. Basándose en el mismo razonamiento, se pueden utilizar estrategias de extracción similares para identificar extremos específicos de los cánceres de un órgano concreto y de un tipo de histología determinado. Las moléculas de ADN plasmático que presentan las diferentes clases de extremos podrían utilizarse para diversas aplicaciones. Por ejemplo, se puede aspirar a detectar los extremos específicos del HCC o del cáncer de pulmón para la detección directa o el cribado del tipo de cáncer específico. Se pueden utilizar los extremos comunes a las muestras de HCC y de cáncer de pulmón para detectar o cribar el cáncer en general. Se pueden utilizar los extremos comunes más genéricos como denominador para la normalización de la cantidad de extremos preferidos asociados a la enfermedad detectados. Los extremos comunes genéricos también podrían detectarse con el fin de detectar el signo de cualquier enfermedad (como un examen de salud general). Los resultados positivos de una prueba de este tipo podrían servir de alerta para visitar a un médico para una investigación más detallada.

B. Comparación entre las posiciones finales preferidas entre las muestras recogidas del individuo de la muestra pero en diferentes momentos

25 Las posiciones finales preferidas de una afección particular también pueden obtenerse comparando los extremos de los fragmentos de las muestras recogidas en diferentes puntos temporales. Por ejemplo, en un paciente oncológico, una muestra de plasma puede recogerse en el momento del diagnóstico y la otra muestra puede recogerse después del tratamiento (por ejemplo, tras la resección quirúrgica del tumor). La diferencia en las posiciones finales puede reflejar potencialmente la ausencia de la contribución del ADN derivado del cáncer en este último o la respuesta corporal al cáncer. En otro ejemplo, se puede hacer una comparación entre las muestras de plasma recogidas de una mujer embarazada tomadas antes y después del parto del feto.

35 En el siguiente ejemplo, se analizaron las muestras de plasma recogidas de 8 mujeres embarazadas. Para cada mujer embarazada, se recogió una muestra de plasma antes del parto. En 6 de las 8 mujeres, se recogió una muestra de plasma adicional en el momento del parto. Se recogieron múltiples muestras de las ocho mujeres embarazadas a partir de las 6 horas después del parto y se recogieron un total de 28 muestras de plasma después del parto. Las muestras de ADN plasmático se secuenciaron con una profundidad media de 6,49x la cobertura del genoma haploide. Las lecturas secuenciadas de las muestras recogidas antes del parto y en el momento del mismo se agruparon para el análisis de la PETR y estas lecturas se denominarán "lecturas previas al parto". Las lecturas secuenciadas de las muestras recogidas a las 6 horas del parto o más tarde se agruparon para el análisis de la PETR y estas lecturas se denominarían lecturas "postparto". Para identificar las posiciones de los nucleótidos que eran extremos preferidos para el embarazo, se recuperaron las posiciones con una PETR al menos 4 veces mayor en las lecturas "antes del parto" en comparación con las lecturas "después del parto". Se identificó un total de 45.281 sitios.

45 Se reclutó una cohorte independiente de 8 mujeres en el primer trimestre del embarazo, cada una de ellas portadora de un feto masculino, y se secuenció su ADN plasmático. Se obtuvo una media de 20 millones de lecturas secuenciadas de estas muestras de ADN plasmático. Se determinó la media de los valores de PETR para los 45.281 sitios para cada una de las 8 mujeres embarazadas y estos valores se correlacionaron con la fracción de ADN fetal en el plasma que se estimó a partir de la proporción de lecturas que se alineaban con el cromosoma Y (Chiu *et al.* BMJ 2011;342:c7401).

55 La FIG. 34A muestra una correlación entre la fracción de ADN fetal en el plasma y la PETR media en el conjunto de posiciones identificadas mediante la comparación entre las muestras de ADN plasmático "antes del parto" y "después del parto". Estos resultados sugieren que el conjunto de posiciones identificadas sería el preferido para el ADN procedente del feto y que el análisis de la PETR sería útil para cuantificar el ADN fetal en el plasma materno.

60 De manera similar a la estrategia descrita anteriormente, los presentes inventores han aplicado el análisis de w-PETR a este conjunto de posiciones preferidas para el embarazo. El tamaño de la ventana A y de la ventana B se fijó en 20 pb y 150 pb, respectivamente. En otras realizaciones, pueden utilizarse otros tamaños de la ventana.

65 La FIG. 34B muestra una correlación entre la fracción de ADN fetal en el plasma y el promedio de w-PETR en el conjunto de posiciones identificadas a través de la comparación entre las muestras de ADN plasmático "antes del parto" y "después del parto". Estos resultados sugieren que el análisis de w-PETR en estas posiciones preferentes del embarazo sería útil para cuantificar el ADN fetal en el plasma materno.

C. Puntos finales comunes entre la misma afección

Comparamos el millón de posiciones finales preferidas más frecuentemente observadas en el plasma de dos mujeres embarazadas (FIG. 35A).

5 La FIG. 35A muestra el millón de posiciones finales preferidas del ADN plasmático observadas con mayor frecuencia entre dos mujeres embarazadas a las 18 semanas (mujer embarazada 1) y a las 38 semanas de gestación (mujer embarazada 2). Los datos muestran que estas mujeres compartían 217.947 extremos preferidos. Teniendo en cuenta
10 que ambas mujeres están embarazadas, estos extremos proceden del feto, la placenta o los órganos que tienen una mayor muerte celular (generación de ADN plasmático) durante el embarazo. Por lo tanto, estos marcadores son más útiles para el seguimiento del embarazo o del bienestar fetal.

15 Los presentes inventores calcularon el valor de la PETR para este conjunto de muestras. Curiosamente, se observó una correlación (r de Pearson = 0,52, valor de $p < 0,0001$) entre los valores de PETR de las moléculas de ADN del plasma en las dos muestras de plasma materno (FIG. 35B).

20 La FIG. 35B muestra una comparación de los valores de PETR del millón de posiciones finales preferidas más frecuentemente observadas en el plasma de dos mujeres embarazadas. El alto grado de correlación indica una vez más que la fragmentación del ADN en el plasma está muy orquestada. Algunos sitios de finalización son más "preferidos" que otros. Curiosamente, incluso entre el millón de sitios "más preferidos", hay un rango dinámico relativamente amplio de PETR. Si uno tuviera que elegir varios o un subconjunto de extremos preferidos para la
25 detección selectiva, por ejemplo, para hacer pruebas de detección de enfermedades, hay que elegir los que se comparten comúnmente entre el grupo de enfermedades de interés, idealmente no se observan o son menos prevalentes en el grupo de control sin enfermedad y, en particular, en las posiciones finales con una PETR muy elevada.

VIII. MÉTODOS QUE UTILIZAN POSICIONES FINALES ESPECÍFICAS DE LOS TEJIDOS

30 La FIG. 36 es un diagrama de flujo de un método 3600 de análisis de una muestra biológica para determinar una clasificación de una contribución proporcional del primer tipo de tejido en una mezcla de acuerdo con las realizaciones de la presente invención. La muestra biológica incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido.

35 En el bloque 3610, se identifica un primer conjunto de posiciones genómicas en las que los extremos de las moléculas de ADN acelular del primer tipo de tejido se producen con una frecuencia superior a un umbral. En la sección X.B se facilitan detalles adicionales acerca del bloque 3610, así como de otros bloques que realizan la identificación de las posiciones terminales preferidas. Los detalles de otros bloques de otros métodos también se pueden encontrar en la sección X.

40 En el bloque 3620, se analiza una primera pluralidad de moléculas de ADN acelular de la muestra biológica de un sujeto. El análisis de una molécula de ADN acelular incluye la determinación de una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular. El bloque 3620 puede realizarse de forma similar a otros bloques para analizar moléculas de ADN acelular, por ejemplo, el bloque 1320.

45 En el bloque 3630, se determina que un primer número de la primera pluralidad de moléculas de ADN acelular termina dentro de una de una pluralidad de ventanas. La determinación se realiza a partir del análisis de la primera pluralidad de moléculas de ADN acelular. Cada ventana incluye al menos una del primer conjunto de posiciones genómicas.

50 En el bloque 3640, se calcula una abundancia relativa de la primera pluralidad de moléculas de ADN acelular que termina dentro de una de la pluralidad de ventanas. La abundancia relativa puede determinarse normalizando el primer número de la primera pluralidad de moléculas de ADN acelular mediante un segundo número de moléculas de ADN acelular. El segundo número de moléculas de ADN acelular incluye moléculas de ADN acelular que terminan en un segundo conjunto de posiciones genómicas fuera de la pluralidad de ventanas que incluyen el primer conjunto de
55 posiciones genómicas.

60 Como se ha descrito para la FIG. 27A, el segundo conjunto de posiciones genómicas puede ser tal que los extremos de las moléculas de ADN acelular de un segundo tipo de tejido se produzcan a una tasa superior al umbral en la al menos una muestra adicional, donde el segundo tipo de tejido tiene una pluralidad de segundos alelos específicos del tejido en la al menos una muestra adicional. El segundo conjunto de posiciones genómicas puede determinarse
65 utilizando moléculas de ADN acelular de la menos una muestra adicional que incluyan al menos uno de la pluralidad de segundos alelos específicos del tejido. Dado que puede excluirse el conjunto G de ambos conjuntos utilizados para determinar la FIG. 27B, las posiciones genómicas en las que los extremos de las moléculas de ADN acelular que tienen un alelo compartido entre el primer tipo de tejido y el segundo tipo de tejido ocurren en una segunda tasa por encima del umbral pueden ser excluidas del primer conjunto de posiciones genómicas y excluidas del segundo conjunto de posiciones genómicas.

En el bloque 3650, la clasificación de la contribución proporcional del primer tipo de tejido se determina comparando la abundancia relativa con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido se conocen.

- 5 Si la contribución proporcional es alta, pueden realizarse acciones adicionales, tales como una intervención terapéutica o la obtención de imágenes del sujeto (por ejemplo, si el primer tipo de tejido corresponde a un tumor). Por ejemplo, una investigación puede utilizar modalidades de imagen, por ejemplo, tomografía computarizada (TC) o resonancia magnética (RM), del sujeto (todo el sujeto o una parte específica del cuerpo (por ejemplo, el tórax o el abdomen), o específicamente del órgano candidato) que puede realizarse para confirmar o descartar la presencia de un tumor en el sujeto. Si se confirma la presencia de un tumor, puede realizarse un tratamiento, por ejemplo, cirugía (con bisturí o con radioterapia) o quimioterapia.

- 15 El tratamiento puede proporcionarse según el nivel de cáncer determinado, las mutaciones identificadas y/o el tejido de origen. Por ejemplo, una mutación identificada (por ejemplo, en el caso de las implementaciones polimórficas) puede tratarse con un fármaco o quimioterapia particular. El tejido de origen puede utilizarse para guiar una cirugía o cualquier otra forma de tratamiento. Asimismo, el nivel de cáncer puede utilizarse para determinar la agresividad de cualquier tipo de tratamiento, que también puede determinarse en función del nivel de cáncer.

IX. DETERMINACIÓN DEL GENOTIPO

- 20 Dado que se pueden determinar las posiciones finales preferidas para un tipo de tejido concreto, las moléculas de ADN acelular que terminan en esas posiciones finales preferidas tienen una alta probabilidad de ser de ese tejido. En algunas situaciones, un tipo de tejido particular en una mezcla de ADN acelular puede tener un genotipo diferente en una posición genómica particular en relación con otros tipos de tejido. Por ejemplo, el tejido fetal o el tejido tumoral pueden tener un genotipo diferente. Dado que las moléculas de ADN acelular tienen una alta probabilidad de ser del tipo de tejido de interés, la molécula de ADN acelular que termina en dicha posición puede analizarse para determinar un genotipo del tipo de tejido en esa posición. De esta manera, la posición final preferida puede utilizarse como filtro para identificar el ADN del tipo de tejido.

30 A. Genotipo fetal

- La información relativa a las posiciones finales de los fragmentos de ADN plasmático secuenciados puede utilizarse para determinar qué alelo materno ha sido heredado por el feto de la mujer embarazada. En este punto, utilizamos un ejemplo hipotético para ilustrar el principio de este método. Se supone que los genotipos de la madre, el padre y el feto son AT, TT y TT, respectivamente. Para determinar el genotipo fetal, es necesario determinar si el feto ha heredado el alelo A o el T de la madre. Los presentes inventores han descrito con anterioridad un método denominado análisis de la dosis relativa de mutaciones (RMD) (Lun *et al.* Proc Natl Acad Sci USA 2008;105:19920-5). En este método, se compararía la dosis de los dos alelos maternos en el plasma materno. Si el feto ha heredado el alelo T materno, el feto sería homocigoto para el alelo T. En este escenario, el alelo T estaría sobrerrepresentado en el plasma materno en comparación con el alelo A. Por otro lado, si el feto ha heredado el alelo A de la madre, el genotipo del feto sería AT. En este escenario, los alelos A y T estarían presentes en aproximadamente la misma dosis en el plasma materno porque tanto la madre como el feto serían heterocigotos para AT. Por tanto, en el análisis RMD, se compararía la dosis relativa de los dos alelos maternos en el plasma materno. Las posiciones finales de las lecturas secuenciadas pueden analizarse para mejorar la precisión de la estrategia de RMD.

- 45 La FIG. 37 muestra las moléculas de ADN plasmático materno que portan diferentes alelos al alinearse con un genoma de referencia cerca de una posición final específica del feto. Las moléculas en líneas continuas proceden de la madre y las moléculas en líneas de puntos proceden del feto. Las moléculas de ADN fetal son más propensas a terminar en las posiciones finales específicas del embarazo. En una realización, las moléculas que terminan en las posiciones finales específicas del embarazo pueden tener más peso en el análisis de RMD. En otra realización, solo los fragmentos de ADN plasmático que terminan en posiciones específicas del embarazo se utilizan para el análisis posterior. Esta selección puede enriquecer potencialmente los fragmentos de ADN plasmático derivados del feto para su análisis posterior.

- 55 La FIG. 37 muestra las moléculas de ADN plasmático en una mujer embarazada cuyo genotipo es AT. Los fragmentos de ADN derivados de los tejidos maternos están en línea continua y los fragmentos de ADN derivados del feto están en línea de puntos. Es más probable que las moléculas de ADN fetal terminen en la posición final específica del embarazo.

- 60 En este ejemplo ilustrativo, las dos moléculas que terminan en la posición final específica del embarazo portan el alelo T. En una realización, solo las dos moléculas que terminan en la posición final específica del embarazo se utilizaron para el análisis posterior y el genotipo fetal se deduciría como TT. En otra realización, las dos moléculas procedentes del feto que portan el alelo T tendrían un mayor peso en el análisis RMD porque estas dos moléculas terminan en una posición final específica del embarazo. Se puede dar una ponderación diferente a las moléculas que terminan en las posiciones finales específicas del embarazo, por ejemplo, pero sin limitación, 1,1, 1,2, 1,3, 1,4, 1,5, 2, 2,5, 3 y 3,5.

65

A modo de ejemplo, el criterio para determinar si un locus es heterocigoto puede ser un umbral de dos alelos que aparezcan cada uno en al menos un porcentaje predeterminado (por ejemplo, el 30 % o el 40 %) de lecturas alineadas con el locus. Si un nucleótido aparece en un porcentaje suficiente (por ejemplo, el 70 % o más), se puede determinar que el locus es homocigoto en el GC.

5

B. Genotipo del cáncer

Se puede realizar una técnica similar para las posiciones finales específicas del cáncer. Por ejemplo, puede identificarse una posición terminal preferida del cáncer como se ha descrito anteriormente. Pueden identificarse y analizarse moléculas de ADN acelular que terminan en la posición final preferida por el cáncer. La base correspondiente (por ejemplo, alineada) a esta posición puede determinarse para cada molécula de ADN acelular de este conjunto, y los porcentajes de las bases totales pueden calcularse para cada base. Por ejemplo, se puede determinar el porcentaje de C que se observa en las moléculas de ADN acelular que terminan en esa posición. Si no se observa C en el tejido sano del sujeto, C puede identificarse como una mutación si se identifica un número suficiente de C, por ejemplo, por encima de un número umbral, que puede depender de la fracción de ADN tumoral medida en la muestra.

10

15

C. Técnicas de filtrado

Para filtrar las moléculas de ADN acelular que proceden del tejido tumoral se pueden utilizar otros criterios además de la posición final. Los otros criterios también pueden utilizarse para el escenario fetal.

20

La especificidad en la identificación de un genotipo de cáncer (por ejemplo, que incluya una mutación específica del cáncer) y cualquier prueba que utilice dichos genotipos (por ejemplo, el uso de la carga mutacional para determinar un nivel de cáncer) puede mejorarse aplicando criterios de filtrado a los locus en los que se hayan alineado una o más lecturas de la secuencia que tengan una mutación. A modo de ejemplo para el cáncer, se puede conseguir una alta especificidad puntuando una firma genética o genómica como positiva solo cuando existe una alta confianza en que esté asociada al cáncer. Esto podría lograrse minimizando el número de errores de secuenciación y alineación que pueden identificarse erróneamente como una mutación, por ejemplo, mediante la comparación con el perfil genómico de un grupo de controles sanos, o puede lograrse mediante la comparación con el propio ADN constitutivo de la persona o puede lograrse mediante la comparación con el perfil genómico de la persona en un momento anterior.

25

30

Pueden aplicarse varios criterios de filtrado para evaluar la probabilidad de que un fragmento de ADN acelular proceda del tumor y, por tanto, pueda considerarse un fragmento de ADN de cáncer informativo. Cada uno de los criterios de filtrado podría utilizarse individualmente, independientemente, colectivamente con igual ponderación o diferentes ponderaciones, o en serie en un orden especificado, o condicionalmente en función de los resultados de los pasos de filtrado anteriores. Para su uso condicional, puede utilizarse una estrategia bayesiana, así como una estrategia basada en árboles de clasificación o decisión. El uso individual de un criterio puede significar el uso de un solo criterio. Un uso independiente puede implicar más de un criterio de filtrado, pero cada criterio de filtrado no depende de la aplicación de otro criterio de filtrado (por ejemplo, se puede aplicar en paralelo), en contraste con una aplicación en serie en órdenes específicas. Como ejemplo de uso colectivo utilizando ponderaciones, pueden utilizarse técnicas de inteligencia artificial. Por ejemplo, el aprendizaje supervisado puede utilizar cargas mutacionales medidas de muestras con clasificaciones conocidas para entrenar cualquier modelo. Los datos de secuenciación de un gran número de individuos (por ejemplo, cientos, miles o millones) pueden utilizarse para entrenar los modelos. Indicado de un modo más sencillo, dichas muestras conocidas pueden utilizarse para determinar los valores umbral de una o más puntuaciones determinadas a partir de los criterios de filtrado para determinar si una mutación es válida o no.

35

40

45

Un fragmento de ADN podría recibir una mayor ponderación del carácter informativo o de la especificidad del cáncer si muestra más de un cambio específico del cáncer. Por ejemplo, muchos cánceres están globalmente hipometilados, especialmente en las regiones no promotoras. Se ha demostrado que el ADN canceroso es más corto que el ADN no canceroso en el plasma. Los fragmentos de ADN plasmático derivados del tumor tienden a fragmentarse en algunos lugares específicos. Por consiguiente, un fragmento de ADN plasmático de tamaño corto (por ejemplo, <150 pb) (Jiang *et al.* Proc Natl Acad Sci USA 2015; 112: E1317-1325), con uno o ambos extremos que caen en ubicaciones finales asociadas al cáncer, muestra una mutación de un solo nucleótido, y se localiza en una región no promotora, y tiene un sitio de CpG hipometilado se consideraría con mayores probabilidades de que esté asociado al cáncer. La detección del ADN hipometilado podría lograrse con el uso de la conversión del ADN en bisulfito o la secuenciación directa de una sola molécula que podría distinguir la metilcitosina de la no metilcitosina. En la presente solicitud, los presentes inventores describen procesos, protocolos y etapas para aumentar la especificidad en la identificación de fragmentos de ADN de cáncer informativos. Por ejemplo, se pueden utilizar uno o varios criterios de filtrado para aumentar la especificidad. Por ejemplo, se pueden utilizar uno o varios criterios de filtrado para aumentar la especificidad, tal como aproximadamente con una especificidad de al menos el 80 %, 90 %, el 95 % o el 99 %.

50

55

60

1. Uso de la localización del extremo del ADN en el plasma

Como se ha descrito anteriormente, se puede realizar el filtrado de posibles mutaciones específicas del cáncer o asociadas al cáncer o al feto, basándose en la coordenada del nucleótido terminal (posición final). Como se ha descrito

65

anteriormente, los presentes inventores han identificado localizaciones terminales de fragmentos de ADN que no son aleatorias y que varían en función del tejido de origen. Por tanto, la localización terminal puede utilizarse para determinar la probabilidad de que una secuencia leída con una supuesta mutación proceda realmente de tejido fetal o de tejido tumoral.

5 Recientemente, se ha demostrado que el patrón de fragmentación del ADN plasmático no es aleatorio (Snyder *et al.* Cell 2016; 164: 57-68 y el documento PCT WO 2016/015058 A2). El patrón de fragmentación del ADN en el plasma está influenciado por el posicionamiento nucleosómico, los sitios de unión de factores de transcripción, los sitios de corte o hipersensibles a DNAsas, los perfiles de expresión (Snyder *et al.* Cell 2016; 164: 57-68 y el documento PCT
10 WO 2016/015058; Ivanov *et al.* BMC Genomics 2015; 16 Supl. 13:S1) y los perfiles de metilación del ADN (Lun *et al.* Clin Chem 2013; 59: 1583-1594) en el genoma de las células que han aportado las moléculas de ADN plasmático. Por tanto, los patrones de fragmentación son diferentes para las células de diferentes orígenes tisulares. Aunque hay regiones genómicas que muestran fragmentos más frecuentes, los lugares de corte del ADN real dentro de la región podrían seguir siendo aleatorios.

15 Los presentes inventores barajan la hipótesis de que los diferentes tejidos están asociados a la liberación de fragmentos de ADN plasmático que tienen diferentes sitios de corte, o ubicaciones finales. En otras palabras, incluso los lugares de corte específicos no son aleatorios. De hecho, los presentes inventores han demostrado que un subconjunto de moléculas de ADN plasmático en pacientes con cáncer muestran ubicaciones finales diferentes a las
20 de los pacientes sin cáncer. Algunas realizaciones pueden utilizar moléculas de ADN plasmático con dichas localizaciones finales asociadas al cáncer como fragmentos de ADN de cáncer informativos, o utilizar dicha información de localización final como criterio de filtrado, por ejemplo, junto con uno o más criterios de filtrado. Por tanto, con la identificación de esas localizaciones finales del ADN plasmático asociadas al cáncer, se podría puntuar el fragmento de ADN plasmático como un fragmento de ADN canceroso informativo o atribuir una ponderación diferencial basada
25 en la naturaleza de la localización final de dicho fragmento. Estos criterios pueden utilizarse para evaluar la probabilidad de que los fragmentos tengan su origen en un cáncer, ciertos órganos, o el cáncer de ciertos órganos. Dicha ponderación puede utilizarse para modificar la contribución de una base concreta de un fragmento de ADN determinado al porcentaje total de una base concreta vista en la posición.

30 Por consiguiente, la probabilidad de que un fragmento de ADN plasmático sea un fragmento de ADN de cáncer informativo sería mucho mayor si muestra una supuesta mutación y/o un cambio de metilación asociado al cáncer, así como las localizaciones finales que están asociadas al cáncer. Varias realizaciones pueden también tener en cuenta el estado de dicho fragmento y su longitud, o cualquier combinación de dichos y otros parámetros. Para un fragmento de ADN plasmático que tiene dos extremos (o potencialmente hasta cuatro extremos, como se describe en una sección
35 siguiente), se puede modificar aún más la ponderación para identificarlo como un fragmento derivado del cáncer teniendo en cuenta si uno o ambos extremos están asociados al cáncer o provienen de un tipo de tejido asociado al cáncer. En una realización, puede utilizarse una estrategia similar basada en las ubicaciones finales para la detección de mutaciones asociadas con otras patologías o procesos biológicos (por ejemplo, mutaciones causadas por el proceso de envejecimiento o mutaciones causadas por factores mutagénicos ambientales).

40 También se puede utilizar una estrategia similar para identificar mutaciones *de novo* de un feto mediante la secuenciación del ADN en el plasma de una mujer embarazada portadora del feto. Por tanto, tras la identificación de las localizaciones finales que son específicas o relativamente específicas para la placenta, se puede atribuir una mayor ponderación a que una supuesta mutación fetal *de novo* sea verdadera si dicho fragmento de ADN en el plasma
45 materno también porta una localización final específica de la placenta o enriquecida por la misma. Dado que un fragmento de ADN plasmático tiene dos extremos, se puede modificar aún más la ponderación para identificarlo como un fragmento derivado del feto teniendo en cuenta si uno o ambos extremos están asociados a la placenta.

50 Como se muestra en la FIG. 16, los fragmentos de ADN plasmático con nucleótidos terminales que terminan exactamente en las 536.772 posiciones finales específicas del HCC tendrían más probabilidades de proceder del tumor. Por el contrario, los fragmentos de ADN plasmático con nucleótidos terminales que terminan exactamente en las posiciones finales específicas del embarazo o en las posiciones compartidas por los dos casos tendrían menos probabilidades de proceder del tumor, siendo las posiciones finales específicas del embarazo potencialmente menos probables y a las que se les da una menor ponderación en cualquier realización que utilice ponderaciones.

55 Por consiguiente, la lista de posiciones finales principales que son específicas para el caso del HCC puede utilizarse para seleccionar las mutaciones asociadas al cáncer, y la lista de posiciones finales principales que son específicas para el caso de la embarazada o compartidas por ambos casos puede utilizarse para filtrar las mutaciones falsas positivas. Se puede utilizar un procedimiento similar para identificar las mutaciones fetales y filtrar las mutaciones
60 falsas positivas para pruebas prenatales no invasivas.

En general, para identificar dichas localizaciones finales del ADN plasmático biológicamente relevantes, las muestras de ADN plasmático de grupos de individuos con diferentes enfermedades o antecedentes epidemiológicos o perfiles fisiológicos podrían compararse con muestras de otro grupo de individuos sin dichas enfermedades o antecedentes o perfiles. En una realización, cada una de estas muestras se pudo secuenciar en profundidad para poder identificar las
65 posiciones finales comunes de los fragmentos de ADN plasmático dentro de cada muestra. En otra realización, los

datos de la secuencia del grupo de personas con perfil complementario podrían agruparse para la identificación de localizaciones finales comunes representativas de la enfermedad o del perfil fisiológico.

5 Cada fragmento de ADN plasmático de una muestra podría interrogarse individualmente y se le asignaría una puntuación de probabilidad en función de la localización final. La puntuación de probabilidad para una determinada localización final puede depender de la separación en una cantidad de lecturas de la secuencia (por ejemplo, un porcentaje de lecturas de la secuencia u otro valor normalizado por la profundidad de la secuenciación en las muestras) que terminan en la localización final para los individuos objetivo (por ejemplo, el cáncer) en relación con la cantidad de lecturas de la secuencia que terminan para el grupo de control. Una mayor separación conllevaría una mayor especificidad y, por tanto, se puede aplicar una mayor puntuación de probabilidad. Por consiguiente, podría realizarse la clasificación de los fragmentos de ADN plasmático con localizaciones finales específicas en probablemente asociados a la enfermedad o no, fetales o maternas, etc.

15 Como alternativa, los fragmentos de ADN plasmático procedentes de la misma región podrían interpretarse colectivamente, a saber, la tasa de finalización en un nucleótido concreto puede calcularse normalizando a la profundidad de secuenciación. De esta manera, se pueden identificar ciertos nucleótidos como lugares finales comunes en relación con otros lugares del genoma, por ejemplo, solo se basa en el análisis de una muestra de un tipo concreto, aunque pueden usarse más muestras. Por consiguiente, podría realizarse la clasificación de los fragmentos de ADN plasmático con localizaciones finales específicas en probablemente asociados a la enfermedad o no, fetales o maternas, etc. Para las posiciones que muestran altas frecuencias de fragmentos de ADN plasmático con dichas ubicaciones finales de ADN plasmático biológicamente relevantes, se podría determinar que dichos locus están enriquecidos con el ADN biológicamente relevante y, por lo tanto, se incluirían como un grupo de fragmentos de ADN plasmático de alta probabilidad como asociados al cáncer o específicos del feto o asociados a otras enfermedades o procesos biológicos. El nivel de probabilidad puede basarse en lo alto que sea el índice de un determinado nucleótido en relación con otros nucleótidos, de manera similar a las comparaciones entre los distintos grupos, como se ha descrito anteriormente.

2. Resultados

30 A fin de ilustrar la eficacia de esta estrategia, las posibles mutaciones asociadas al cáncer se identificaron directamente a partir de los datos de secuenciación del ADN plasmático del paciente con HCC. Los cambios de un solo nucleótido que estaban presentes en las lecturas de la secuencia de al menos dos fragmentos de ADN plasmático se consideraron como posibles mutaciones asociadas al cáncer. También se secuenció el tejido tumoral y las mutaciones que estaban presentes en el tejido tumoral se consideraron verdaderas mutaciones asociadas al cáncer.

35 En el cromosoma 8, se identificó un total de 20.065 posibles mutaciones a partir de los datos de secuenciación del ADN plasmático del paciente con HCC sin utilizar el análisis de corte dinámico. Una variante de secuencia se consideraría una posible mutación si la variante de secuencia estuviera presente en al menos dos fragmentos de ADN secuenciados. Se identificaron 884 mutaciones somáticas verdaderas a partir del resultado de la secuenciación del tejido tumoral. Las 20.065 supuestas mutaciones incluían 802 (91 %) de las 884 mutaciones reales. Por tanto, solo el 4 % de las supuestas mutaciones eran verdaderas mutaciones somáticas en el tejido tumoral, lo que dio un PPV del 4 %.

45 Para mejorar la precisión de la detección de las mutaciones somáticas, obteniéndose de este modo un genotipo de cáncer, los presentes inventores utilizaron los siguientes algoritmos de filtrado basados en las posiciones de los nucleótidos terminales de las lecturas de la secuencia que portan las supuestas mutaciones. (1). Para cualquier supuesta mutación, si hay al menos una lectura de la secuencia que porte la mutación y que termine en posiciones finales específicas de HCC, la mutación se calificaría para el análisis mutacional posterior. (2). Se eliminaría una lectura de secuencia que portase una supuesta mutación pero que terminara en cualquier posición final específica del embarazo o en las posiciones compartidas por ambos casos. Una mutación se calificaría para el análisis mutacional posterior solo si hubiera dos o más lecturas de la secuencia que mostraran la misma mutación tras la eliminación de las lecturas basadas en este algoritmo.

55 Aplicando los algoritmos de filtrado 1 y 2 indicados anteriormente, se obtuvieron los resultados en la tabla 2. Los efectos de la aplicación de diferentes algoritmos de filtrado basándose en la posición de los nucleótidos terminales, o de los extremos, de los fragmentos de ADN portadores de las supuestas mutaciones.

Tabla 2

	Sin filtro	Inclusión de mutaciones con extremos específicos de HCC (filtro 1)	Eliminación de lecturas con extremos compartidos o específicos del embarazo (filtro 2)	Aplicando ambos algoritmos de filtrado
N.º de supuestas mutaciones identificadas	20.065	1.526	2.823	484

(continuación)

	Sin filtro	Inclusión de mutaciones con extremos específicos de HCC (filtro 1)	Eliminación de lecturas con extremos compartidos o específicos del embarazo (filtro 2)	Aplicando ambos algoritmos de filtrado
Porcentaje de mutaciones verdaderas detectadas	91 %	29 %	88 %	40 %
PPV	4 %	17 %	28 %	71 %

5 Hubo una mejora sustancial en el PPV al adoptar cualquiera de los tres algoritmos que requerían que las localizaciones finales fueran específicas del HCC o que el algoritmo filtrara las posiciones específicas del embarazo o las compartidas. Aplicando ambos algoritmos, el PPV aumentó al 71 %.

10 Se puede identificar otro número de localizaciones finales asociadas al HCC y al embarazo para cada cromosoma, o incluso para otra región genómica, o incluso para todo el genoma, por ejemplo, pero sin limitación, 0,5 millones, 2 millones, 3 millones, 4 millones, 5 millones, 6 millones, 7 millones, 8 millones, 9 millones o 10 millones. En diversas realizaciones, se pueden determinar las localizaciones finales más frecuentes en las moléculas de ADN plasmático en una o más cohortes de pacientes con cáncer, siendo cada cohorte de un tipo de cáncer. Además, las localizaciones finales más frecuentes en las moléculas de ADN plasmático pueden determinarse para los sujetos sin cáncer. En una realización, estos pacientes con cáncer y los sujetos sin cáncer pueden subdividirse en grupos con diferentes parámetros clínicos, por ejemplo, el sexo, estado de tabaquismo, la salud previa (por ejemplo, el estado de hepatitis, diabetes, peso), etc.

20 Como parte de la utilización de estos criterios de filtrado, el análisis estadístico puede utilizarse para identificar las posiciones que tienen mayor probabilidad de ser nucleótidos terminales o ubicaciones finales del ADN circulante para diferentes condiciones fisiológicas y patológicas. Los ejemplos de los análisis estadísticos incluyen, entre otros, la prueba de la t de Student, la prueba de Chi-cuadrado, y pruebas basadas en la distribución binomial o en la distribución de Poisson. Para estos análisis estadísticos, se pueden utilizar diferentes cortes de valores de p, por ejemplo, pero sin limitación, 0,05, 0,01, 0,005, 0,001 y 0,0001. Los puntos de corte del valor de p también pueden ajustarse para las comparaciones múltiples.

25 *D. Método para determinar el genotipo*

30 La FIG. 38 es un diagrama de flujo de un método 3800 de análisis de una muestra biológica para determinar un genotipo del primer tipo de tejido de acuerdo con las realizaciones de la presente invención. La muestra biológica incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye el primer tipo de tejido. El primer tipo de tejido tiene potencialmente un genotipo diferente al de otros tipos de tejido de la pluralidad de tipos de tejido. Se pueden determinar los genotipos en múltiples posiciones genómicas.

35 En el bloque 3810, se identifica una primera posición genómica en la que los extremos de las moléculas de ADN acelular del primer tipo de tejido se producen a una tasa superior a un umbral. El bloque 3810 puede realizarse de un modo similar al bloque 3610. La sección X.B proporciona ejemplos adicionales para realizar el bloque 3810.

40 En el bloque 3820, se analiza una primera pluralidad de moléculas de ADN acelular de la muestra biológica de un sujeto. El análisis de una molécula de ADN acelular incluye la determinación de una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular. El bloque 3620 puede realizarse de forma similar a otros bloques para analizar moléculas de ADN acelular.

45 En el bloque 3830, se identifica un conjunto de moléculas de ADN acelular que terminan en la primera posición genómica basándose en el análisis de la primera pluralidad de moléculas de ADN acelular. A modo de ejemplo, el conjunto puede identificarse mediante la alineación de las lecturas de la secuencia de las sondas detectadas que tienen posiciones finales conocidas. En el presente documento se proporcionan otros ejemplos.

50 En algunas realizaciones, puede realizarse un filtrado adicional, por ejemplo, como se ha descrito anteriormente. Por ejemplo, puede ser necesario que el tamaño de una molécula de ADN acelular sea menor que una cantidad especificada, por ejemplo, dado que el tejido fetal y el tejido tumoral son generalmente más cortos que los fragmentos de ADN de las células sanas. En una implementación, el conjunto de moléculas de ADN acelular puede filtrarse para excluir o modificar una ponderación de al menos una de las moléculas de ADN acelular que terminan en la primera posición genómica. El genotipo puede determinarse utilizando un conjunto filtrado de moléculas de ADN acelular.

55 En diversas realizaciones, el filtrado puede utilizar al menos uno de: el tamaño de una molécula de ADN acelular, un estado de metilación de la molécula de ADN acelular en una o más posiciones (por ejemplo, si un sitio de CpG está metilado o no), y si la molécula de ADN acelular cubre una o más posiciones genómicas en las que los extremos de las moléculas de ADN acelular del primer tipo de tejido se producen a una tasa superior a un umbral. El estado de

metilación puede proporcionar una firma del primer tipo de tejido, como se ha descrito anteriormente.

5 En el bloque 3840, para cada molécula de ADN acelular del conjunto de moléculas de ADN acelular, se determina la base correspondiente (nucleótido) que se encuentra en la primera posición genómica. Se puede determinar el número total de moléculas con cada base y calcular un porcentaje para cada base.

10 En el bloque 3850, el genotipo del primer tipo de tejido en la primera posición genómica se determina utilizando las bases correspondientes que aparecen en la primera posición genómica en el conjunto de moléculas de ADN acelular. En diversas implementaciones, un alto porcentaje de una sola base (por ejemplo, por encima del 80 %, 85 % o 90 %) puede indicar que el genotipo es homocigoto para la base, mientras que dos bases con porcentajes similares (por ejemplo, entre el 30-70 %) pueden llevar a determinar que el genotipo es heterocigoto. Por consiguiente, los porcentajes de cada base pueden compararse con los valores de corte para el genotipo. En algunas realizaciones, se puede determinar un valor de corte basado en una contribución proporcional del primer tipo de tejido a la muestra.

15 Por tanto, en algunas realizaciones, la determinación del genotipo del primer tipo de tejido en la primera posición genómica puede incluir la determinación de una contribución porcentual para cada una de una pluralidad de bases y la comparación de cada una de las contribuciones porcentuales con uno o más valores de corte. En un ejemplo, un primer valor de corte puede corresponder a un genotipo homocigoto de una primera base cuando el porcentaje de contribución de la primera base está por encima del primer valor de corte. En otros ejemplos, un primer valor de corte y un segundo valor de corte pueden corresponder a un genotipo heterocigoto para una primera base y una segunda base cuando las contribuciones porcentuales de la primera base y la segunda base están por encima del primer valor de corte y por debajo del segundo valor de corte.

25 En algunas realizaciones, se puede realizar una ponderación para cada molécula de ADN acelular en el conjunto identificado en el bloque 3830. Por ejemplo, si la probabilidad de que la molécula de ADN acelular sea del primer tipo de tejido es del 80 %, la ponderación puede ser 0,8. La contribución total de todos los coeficientes de ponderación para una base concreta puede sumarse para determinar las cantidades respectivas de cada base. Las cantidades respectivas pueden utilizarse para determinar un porcentaje de contribución para cada base, donde los porcentajes pueden utilizarse para determinar el genotipo.

30 Por consiguiente, el filtrado puede asignar una ponderación a la molécula de ADN acelular correspondiente a una probabilidad de que la molécula de ADN acelular sea del primer tipo de tejido. Se puede determinar una suma ponderada para cada una de una pluralidad de bases (por ejemplo, solo las detectadas, que pueden ser 2, 3 o 4). En caso de que solo se detecte una base, se puede determinar un genotipo homocigoto para esa única base. Se puede determinar un porcentaje de contribución para cada una de la pluralidad de bases utilizando las sumas ponderadas, donde el genotipo se determina utilizando las contribuciones porcentuales.

X. DETALLES ADICIONALES

40 Varias realizaciones descritas anteriormente identifican las posiciones finales preferidas para determinados tejidos, donde algunas de las posiciones finales preferidas pueden ser contiguas, formando así una ventana final preferente. Se pueden utilizar diferentes métricas para identificar las tasas de aparición de moléculas de ADN acelular en las ventanas genómicas (por ejemplo, una posición genómica para la ventana más pequeña). A continuación, se ofrecen más detalles sobre estas operaciones, así como detalles sobre la determinación de la posición final de una molécula de ADN acelular en un genoma de referencia. Estas técnicas específicas pueden utilizarse con las realizaciones descritas anteriormente.

A. Determinación de la posición final

50 Al secuenciar moléculas de ADN acelular, hay varias posibilidades de los patrones de terminación de los fragmentos de ADN. En general, existen cuatro configuraciones de extremos para el ADN plasmático: (A) Una molécula de ADN bicatenaria con dos extremos rasgados; (B) Una molécula de ADN bicatenario con un extremo rasgado y otro no rasgado (mostrando cada uno de los dos escenarios, ya que cualquiera de las dos cadenas puede sobresalir); (C) Una molécula de ADN bicatenario con dos extremos no rasgados, con diferentes combinaciones de extremos salientes; y 55 (D) Una molécula de ADN monocatenario.

60 Para las configuraciones con extremos no rasgados, hay diferentes patrones dependiendo de si sobresale el extremo 5' o el 3' de la molécula de ADN. Para (B), las moléculas de ADN bicatenario tienen un extremo rasgado y otro no rasgado. En un ejemplo B1, el extremo 5' es protuberante y en un ejemplo B2, el extremo 3' es protuberante. Para (C), hay tres patrones posibles cuando los dos extremos no están rasgados. En (C1), el extremo 5' es protuberante a ambos lados. En (C2), el extremo 3' es protuberante a ambos lados. En (C3), el extremo 5' es protuberante en un lado y el extremo 3' es protuberante en el otro lado.

65 Para la secuenciación, los protocolos de secuenciación de tipo *paired-end* normalmente secuencian un extremo de cada una de las hebras. Por tanto, se consideran protocolos de secuenciación de ADN bicatenario. Cuando los dos extremos no están rasgados, los protocolos pueden cortar los nucleótidos o añadir nucleótidos al final para que queden

al descubierto. El fragmento Klenow es una enzima que puede realizar estas operaciones. Otros protocolos en el campo utilizan protocolos de secuenciación de ADN monocatenario.

5 Independientemente de la técnica específica utilizada (incluido el uso de sondas), siempre que las posiciones finales sean repetibles y muestren correlación, como se muestra en el presente documento, el hecho de que se obtenga un extremo verdadero de un fragmento de ADN en la secuenciación no afecta a los resultados, ya que cualquier desplazamiento es repetible, y por lo tanto se anula. Además, se pueden utilizar ciertas técnicas para identificar una posición final, como se describe en la sección de términos.

10 *B. Identificación de las posiciones finales específicas de los tejidos*

Como se ha descrito anteriormente, en un tipo de tejido concreto, ciertas regiones genómicas presentan una mayor variación en cuanto a la probabilidad de que una molécula de ADN acelular termine en una posición determinada que otras regiones. Por ejemplo, el tejido hepático puede tener una región que es un sitio de hipersensibilidad a DNAsas, pero otros tejidos no tienen esa región como sitio de hipersensibilidad a DNAsas. Por consiguiente, ciertas posiciones dentro de dicha región tendrán un alto número de moléculas de ADN acelular que terminan en esas posiciones en relación con otras. A modo de ejemplo, estas posiciones pueden identificarse como máximas en una tasa de moléculas de ADN acelular para una región que se sabe que tiene una gran cantidad de escisión para un tejido concreto (por tanto, una gran amplitud en la función de probabilidad), por ejemplo, como se describe en la sección III. En otros ejemplos, se pueden identificar las posiciones genómicas en las que un pico izquierdo y un pico derecho están suficientemente separados, por ejemplo, como se describe en la sección IV.

En otros ejemplos más, una diferencia en los conjuntos de posiciones finales de alta tasa (por ejemplo, tasa por encima de un umbral) para las muestras que tienen y no tienen una condición (por ejemplo, embarazo o cáncer, posiblemente de un tipo concreto) puede utilizarse para identificar los lugares de finalización preferidos para un tipo de tejido concreto asociado a la enfermedad, por ejemplo, como se describe con el uso de diagramas de Venn en las secciones V, VI y VII. En otros ejemplos adicionales, una tasa significativamente mayor en una muestra con una afección que en otra que no la tenga puede proporcionar los lugares de terminación preferidos de un tipo de tejido concreto. En diversas realizaciones, algunas o todas estas técnicas de ejemplo pueden utilizarse conjuntamente. La tasa puede medirse por cualquier métrica de abundancia relativa.

En algunas realizaciones de los métodos anteriores, un primer conjunto de posiciones genómicas en las que los extremos de las moléculas de ADN acelular del primer tipo de tejido se producen a una tasa superior a un umbral puede identificarse de la siguiente manera. Una muestra de calibración puede analizarse de manera similar a la muestra de prueba, donde se sabe que las dos muestras de un mismo tipo (por ejemplo, plasma, suero, orina, etc.) y la muestra de calibración incluyen el primer tipo de tejido (por ejemplo, tejido fetal de una muestra de una mujer embarazada o tejido tumoral del hígado de un paciente con HCC). Un número de moléculas de ADN acelular que terminan en una ventana genómica (por ejemplo, de anchura uno o más) puede compararse con un valor de referencia para determinar si una tasa de posiciones finales está por encima de un umbral para esa posición. En algunas realizaciones, si el índice supera el valor de referencia, puede identificarse que cada una de las posiciones genómicas dentro de la ventana genómica tiene la tasa por encima del umbral cuando el número correspondiente supera el valor de referencia. Este proceso puede identificar las ventanas de finalización preferidas, que incluyen las posiciones finales preferidas.

45 El valor de referencia puede ser tal que solo las N ventanas genómicas superiores tengan un índice superior al umbral. Por ejemplo, el primer conjunto de posiciones genómicas puede tener los valores N más altos para los números correspondientes. A modo de ejemplo, N puede ser al menos 10.000; 50.000; 100.000; 500.000; 1.000.000; o 5.000.000.

50 Como otro ejemplo, el valor de referencia puede ser un número esperado de moléculas de ADN acelular que terminan dentro de la ventana genómica según una distribución de probabilidad y una longitud media de moléculas de ADN acelular en una muestra, por ejemplo, como se describe en la sección VI.A.1. Se puede determinar un valor de p utilizando el número correspondiente y el número esperado, en donde el umbral corresponde a un valor de p de corte (por ejemplo, 0,01). Que el valor de p sea inferior al valor de p de corte indica que la tasa está por encima del umbral. Como otro ejemplo más, el valor de referencia puede incluir un número medido de moléculas de ADN acelular que terminan dentro de la ventana genómica de una muestra identificada como con una cantidad reducida del primer tipo de tejido, por ejemplo, como se describe en las FIG. 29A y 29B.

60 Las posiciones genómicas que satisfacen el umbral de la tasa no se añaden necesariamente al primer conjunto de posiciones genómicas. Se pueden añadir otros criterios de filtrado. En las secciones VI.A.3 y IX.C se especifican ejemplos de estos criterios de filtrado. Para un criterio de filtrado de tamaño, se puede medir el tamaño (por ejemplo, la longitud o la masa) de las moléculas de ADN acelular, por ejemplo, como se describe en las Publicaciones de Patente de los Estados Unidos 2011/0276277, 2013/0040824 y 2013/0237431. Se puede determinar un primer valor estadístico de una distribución de tamaño de las moléculas de ADN acelular que terminan dentro de una primera ventana genómica (por ejemplo, en una posición genómica cuando la ventana tiene una anchura de uno) determinada para tener la tasa por encima del umbral. Las posiciones genómicas de la primera ventana genómica pueden excluirse

del primer conjunto de posiciones genómicas cuando el primer valor estadístico no supera un umbral de tamaño, por ejemplo, el tamaño medio no es lo suficientemente pequeño o no hay un número suficiente de fragmentos de ADN pequeños (por ejemplo, por debajo de un tamaño especificado) en comparación con todas las moléculas de ADN acelular o las que se encuentran en un rango mayor.

5 El primer valor estadístico puede compararse con un segundo valor estadístico de una distribución de tamaños para las moléculas de ADN acelular que se haya determinado que no tienen una tasa superior al umbral. Si los dos valores son similares (por ejemplo, lo que no se esperaría de un tejido fetal o tumoral), la primera ventana genómica puede excluirse de un conjunto de posiciones finales preferidas. La comparación del número correspondiente con el valor de referencia puede incluir el cálculo de una primera relación (por ejemplo, PETR) del número correspondiente y un número de moléculas de ADN acelular que cubren cualquier parte de la ventana genómica para una muestra, y que opcionalmente no termina en la ventana genómica, como se describe en la sección VII.A.2. El valor de referencia puede incluir una relación de referencia entre el número medido de lecturas que terminan dentro de la ventana genómica y un número de moléculas de ADN acelular que cubren la ventana genómica y no terminan dentro de la ventana genómica para la otra muestra. Se puede exigir que la primera relación sea mayor que un factor multiplicativo (por ejemplo, 4) de la relación de referencia.

20 Otro criterio de filtrado puede ser que se requiera que cada posición genómica del primer conjunto de posiciones genómicas tenga al menos un número especificado de moléculas de ADN acelular que terminen en la posición genómica. Mediante el uso de cualquiera de estas técnicas, el primer conjunto de posiciones genómicas puede comprender entre 600 y 10.000 posiciones genómicas.

25 En las realizaciones que toman una diferencia entre conjuntos (por ejemplo, el uso de diagramas de Venn), las posiciones genómicas cuya tasa (por ejemplo, determinada a partir de una ventana genómica) está por encima del umbral comprende un primer superconjunto, por ejemplo, como se muestra en la FIG. 28A como Conj. P y Conj. S. Se puede analizar una tercera pluralidad de moléculas de ADN acelular a partir de al menos una segunda muestra adicional que tenga una cantidad reducida del primer tipo de tejido (por ejemplo, menos o ningún tejido fetal o tejido HCC, como se representa en la FIG. 28A) para identificar un segundo superconjunto, por ejemplo, Conj. Q y Conj. S. El primer conjunto de posiciones genómicas puede incluir las posiciones genómicas que están en el primer superconjunto y que no están en el segundo, por ejemplo, Conj. P o Conj. S, dependiendo del tipo de tejido que se analice.

35 Como se describe en la sección VI, el primer tipo de tejido puede tener los primeros alelos específicos del tejido. Se puede hacer un recuento de la molécula de ADN acelular que termina en la posición genómica y que incluye al menos uno de la pluralidad de primeros alelos específicos del tejido. Este recuento (número) de moléculas de ADN acelular puede compararse con el valor de referencia.

C. Abundancia relativa

40 En este documento se ofrecen varios ejemplos de valores de abundancia relativa, por ejemplo, probabilidad de intacto (P_i), el valor de p descrito en la sección VI.A.1, y el valor de PETR determinado utilizando una ventana genómica o una posición genómica cuando la ventana es de anchura uno. Para la PETR de una posición genómica (ventana de anchura uno), se puede calcular un número correspondiente de la primera pluralidad de moléculas de ADN acelular que terminan en la posición genómica para cada posición genómica del primer conjunto de posiciones genómicas. Esto puede hacerse como parte de la determinación de que el primer número (por ejemplo, el numerador) de la primera pluralidad de moléculas de ADN acelular termina en cualquiera de la primera serie de posiciones genómicas. Un tercer número (por ejemplo, el denominador) de moléculas de ADN acelular que cubren la posición genómica y no terminan en la posición genómica puede calcularse como parte de la determinación del segundo número de moléculas de ADN acelular. Se puede determinar una primera relación entre el número correspondiente y el tercer número, y utilizar una media de las primeras relaciones como abundancia relativa.

55 Para w -PETR, puede calcularse un número correspondiente de moléculas de ADN acelular que terminan dentro de una primera venta (por ejemplo, la ventana A en la FIG. 31A) que incluyen la posición genómica para cada una de las posiciones genómicas del primer conjunto de posiciones genómicas. Se puede calcular un tercer número de moléculas de ADN acelular que terminan dentro de una segunda ventana (por ejemplo, de la ventana B en la FIG. 31A) que incluye la posición genómica. Como abundancia relativa, puede usarse la media de las primeras proporciones de los números correspondientes y los terceros números.

60 Otro ejemplo de valor de abundancia relativa es la proporción de moléculas de ADN acelular que terminan en una ventana genómica, por ejemplo, medida como una proporción de fragmentos de ADN secuenciados que terminan en una posición final preferida. Por tanto, el segundo conjunto de posiciones genómicas puede incluir todas las posiciones genómicas correspondientes a un extremo de al menos una de la primera pluralidad de moléculas de ADN acelular.

D. Valores de calibración

65 En diversas realizaciones, el valor o los valores de calibración pueden corresponder al valor o los valores de calibración

del punto o los puntos de datos de calibración determinados a partir de la muestra o las muestras de calibración o de cualquier valor de calibración determinado a partir de ellas, por ejemplo, de una función de calibración que se aproxima a los puntos de datos de calibración. Las una o más muestras de calibración pueden o no incluir cualquier muestra adicional utilizada para determinar los lugares de finalización preferidos.

5 Para cada una de las una o más muestras de calibración, se puede medir la correspondiente contribución proporcional del primer tipo de tejido, por ejemplo, utilizando un alelo específico de tejido. Se puede determinar una abundancia relativa correspondiente utilizando los números correspondientes de moléculas de ADN acelular que terminan dentro de la pluralidad de ventanas correspondientes al primer conjunto de posiciones genómicas. La contribución proporcional medida y la abundancia relativa pueden proporcionar un punto de datos de calibración. Los uno o más puntos de datos de calibración pueden ser una pluralidad de puntos de datos de calibración que forman una función de calibración que se aproxima a la pluralidad de puntos de datos de calibración. Pueden encontrarse más detalles sobre el uso de los valores de calibración en la Publicación de Patente de los Estados Unidos 2013/0237431.

15 E. Clasificación de la contribución proporcional

En algunas realizaciones, las posiciones finales preferidas para un tejido concreto también pueden utilizarse para medir la contribución absoluta de un tipo de tejido concreto en una muestra, por ejemplo, en número de genomas por unidad de volumen (por ejemplo, por mililitro). Por ejemplo, se podría medir la concentración del tejido de interés en relación con el volumen o el peso de las muestras de ADN acelular. En una implementación, la PCR cuantitativa podría utilizarse para medir el número de moléculas de ADN acelular que terminan en uno o más extremos preferidos en una unidad de volumen o de peso de la muestra de ADN acelular extraída. Se pueden realizar mediciones similares para las muestras de calibración y, por lo tanto, se puede determinar la contribución proporcional, ya que la contribución es una concentración por unidad de volumen o de peso.

25 En diversas realizaciones, cuando el primer tipo de tejido corresponde al tejido tumoral, la clasificación puede seleccionarse entre un grupo que consiste en: una cantidad de tejido tumoral en el sujeto, un tamaño del tumor en el sujeto, un estadio del tumor en el sujeto, una carga tumoral en el sujeto, y la presencia de metástasis tumorales en el sujeto

30 XI. REALIZACIONES ADICIONALES

La realización 1 incluye un método implementado por ordenador para analizar una muestra biológica, que incluye una mezcla de moléculas de ADN acelular de diversos tipos de tejido que incluye un primer tipo de tejido, para determinar una clasificación de una contribución proporcional del primer tipo de tejido en la mezcla, comprendiendo el método: identificar un primer conjunto de posiciones genómicas en las que los extremos de las moléculas de ADN acelular del primer tipo de tejido se producen con una frecuencia superior a un umbral; analizar, mediante un sistema informático, una primera pluralidad de moléculas de ADN acelular de la muestra biológica de un sujeto, en donde el análisis de una molécula de ADN acelular incluye: determinar una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular; basándose en el análisis de la primera pluralidad de moléculas de ADN, determinar que un primer número de la primera pluralidad de moléculas de ADN acelular termina dentro de una de una pluralidad de ventanas, incluyendo cada ventana al menos una del primer conjunto de posiciones genómicas; calcular una abundancia relativa de la primera pluralidad de moléculas de ADN acelular que terminan dentro de una de la pluralidad de ventanas mediante la normalización del primer número de la primera pluralidad de moléculas de ADN acelular utilizando un segundo número de moléculas de ADN acelular, en donde el segundo número de moléculas de ADN acelular incluye moléculas de ADN acelular que terminan en un segundo conjunto de posiciones genómicas fuera de la pluralidad de ventanas que incluyen el primer conjunto de posiciones genómicas; y determinar la clasificación de la contribución proporcional del primer tipo de tejido comparando la abundancia relativa con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido se conocen.

La realización 2 incluye el método de la realización 1, en donde la identificación del primer conjunto de posiciones genómicas incluye: analizar, mediante un sistema informático, una segunda pluralidad de moléculas de ADN acelular a partir de al menos una primera muestra adicional para identificar posiciones terminales de la segunda pluralidad de moléculas de ADN acelular, en donde se sabe que la al menos una primera muestra adicional incluye el primer tipo de tejido y es de un mismo tipo de muestra que la muestra biológica; para cada ventana genómica de una pluralidad de ventanas genómicas: calcular un número correspondiente de la segunda pluralidad de moléculas de ADN acelular que terminan en la ventana genómica; y comparar el número correspondiente con un valor de referencia para determinar si la tasa de moléculas de ADN acelular que terminan en una o más posiciones genómicas dentro de la ventana genómica está por encima del umbral.

La realización 3 incluye el método de la realización 2, en donde una primera ventana genómica de la pluralidad de ventanas genómicas tiene una anchura mayor que una posición genómica, y en donde cada una de las posiciones genómicas dentro de la primera ventana genómica se identifica por tener la tasa de moléculas de ADN acelular que terminan en la posición genómica por encima del umbral cuando el número correspondiente supera el valor de referencia. La realización 4 incluye el método de la realización 2 o 3, en donde el primer conjunto de posiciones

genómicas tiene los valores de N más altos para los números correspondientes, en donde N es al menos 10.000.

La realización 5 incluye el método de las realizaciones 2, 3 o 4, que comprende además: determinar un tamaño de cada una de la segunda pluralidad de moléculas de ADN acelular, en donde la identificación del primer conjunto de posiciones genómicas incluye además: determinar un primer valor estadístico de una distribución de tamaños de las moléculas de ADN acelular de la segunda pluralidad de moléculas de ADN acelular que terminan dentro de una primera ventana genómica determinada para tener la tasa por encima del umbral; comparar el primer valor estadístico con un umbral de tamaño; y excluir la primera ventana genómica del primer conjunto de posiciones genómicas cuando el primer valor estadístico no supera el umbral de tamaño. La realización 6 incluye el método de una cualquiera de las realizaciones 2-5, en donde las una o más muestras de calibración incluyen la al menos una primera muestra adicional. La realización 7 incluye el método de una cualquiera de las realizaciones 1 a 6, comprendiendo además: para cada una de las una o más muestras de calibración: medir una contribución proporcionar correspondiente del primer tipo de tejido; y determinar una abundancia relativa correspondiente utilizando los números correspondientes de la segunda pluralidad de moléculas de ADN acelular que terminan dentro de la pluralidad de ventanas correspondientes al primer conjunto de posiciones genómicas, obteniendo de este modo un punto de datos de calibración, en donde cada punto de datos de calibración especifica la contribución proporcional medida del primer tipo de tejido para la muestra biológica adicional y la abundancia relativa correspondiente. La realización 8 incluye el método de la realización 7, en donde uno o más puntos de datos de calibración son una pluralidad de puntos de datos de calibración que forman una función de calibración que se aproxima a la pluralidad de puntos de datos de calibración.

La realización 9 incluye el método de una cualquiera de las realizaciones 2 a 8, en donde cada posición genómica del primer conjunto de posiciones genómicas tiene al menos un número especificado de moléculas de ADN acelular de la segunda pluralidad de moléculas de ADN acelular que terminan en la posición genómica. La realización 10 incluye el método de una cualquiera de las realizaciones 2 a 9, en donde el valor de referencia es un número esperado de moléculas de ADN acelular que terminan dentro de la ventana genómica de acuerdo con una distribución de probabilidad y una longitud media de moléculas de ADN acelular en la al menos una primera muestra adicional. La realización 11 incluye el método de la realización 10, en donde la distribución de probabilidad es una distribución de Poisson, y en donde determinar si la tasa de moléculas de ADN acelular que terminan en una o más posiciones genómicas dentro de la ventana genómica está por encima del umbral incluye: determinar un valor p correspondiente usando el número correspondiente y el número esperado, en donde el umbral corresponde a un valor p de corte, siendo el valor p correspondiente menor que el valor p de corte que indica que la tasa de moléculas de ADN acelular que terminan dentro de la ventana genómica está por encima del umbral.

La realización 12 incluye el método de una cualquiera de las realizaciones 2 a 11, en donde las posiciones genómicas cuya tasa de la segunda pluralidad de moléculas de ADN acelular que terminan en la posición genómica se encuentra por encima del umbral comprende un primer superconjunto y en donde la identificación del primer conjunto de posiciones genómicas incluye además: analizar, mediante el sistema informático, una tercera pluralidad de moléculas de ADN acelular de al menos una segunda muestra adicional en la que se ha identificado que tiene una cantidad reducida del primer tipo de tejido para identificar que un segundo superconjunto de la tercera pluralidad de moléculas de ADN acelular que terminan en la posición genómica se encuentra por encima del umbral; e identificar que el primer conjunto de posiciones genómicas incluye las posiciones genómicas que se encuentran en el primer superconjunto y que no se encuentran en el segundo superconjunto.

La realización 13 incluye el método de una cualquiera de las realizaciones 2 a 12, en donde el valor de referencia incluye un número medido de moléculas de ADN acelular que terminan dentro de la ventana genómica, el número medido determinado a partir de una tercera pluralidad de moléculas de ADN acelular de al menos una segunda muestra que se ha identificado que no tiene el primer tipo de tejido. La realización 14 incluye el método de la realización 13, que comprende, además: determinar un tamaño de cada una de la segunda pluralidad de moléculas de ADN acelular, en donde la identificación del primer conjunto de posiciones genómicas incluye además: determinar un primer valor estadístico de una distribución de tamaños de las moléculas de ADN acelular de la segunda pluralidad de moléculas de ADN acelular que terminan en una primera posición genómica determinada para tener la tasa por encima del umbral; determinar un segundo valor estadístico de una segunda distribución de tamaños de las moléculas de ADN acelular de la tercera pluralidad de moléculas de ADN acelular que terminan en una o más segundas posiciones genómicas determinadas para tener la tasa por encima del umbral; comparar el primer valor estadístico con el segundo valor estadístico; y excluir la primera posición genómica del primer conjunto de posiciones genómicas cuando el primer valor estadístico no supera el segundo valor estadístico en al menos una cantidad especificada para indicar que la primera distribución de tamaños es menor que la segunda distribución de tamaños. La realización 15 incluye el método de la realización 13 o 14, en donde la comparación del número correspondiente con el valor de referencia incluye: calcular una primera proporción entre el número correspondiente y un tercer número de la segunda pluralidad de moléculas de ADN acelular que abarcan la ventana genómica; y comparar la primera proporción con el valor de referencia, incluyendo el valor de referencia una proporción de referencia del número medido de lecturas que terminan dentro de la ventana genómica y un cuarto número de la tercera pluralidad de moléculas de ADN acelular que abarcan la ventana genómica y que no terminan dentro de la ventana genómica. La realización 16 incluye el método de la realización 15, en donde el tercer número de la segunda pluralidad de moléculas de ADN acelular no terminan dentro de la ventana genómica. La realización 17 incluye el método de la realización 15 o 16, en donde determinar si la tasa de moléculas de ADN acelular que terminan dentro de la ventana genómica está por encima del umbral incluye:

determinar si la primera relación es mayor que un factor multiplicativo de la relación de referencia.

5 La realización 18 incluye el método de una cualquiera de las realizaciones 2 a 17, en donde el tipo de muestra de la muestra biológica y de la al menos una primera muestra adicional se selecciona de un grupo que consiste en plasma, suero, líquido cefalorraquídeo y orina. La realización 19 incluye el método de una cualquiera de las realizaciones 2 a 18, en donde la ventana genómica es una posición genómica y en donde el primer tipo de tejido tiene una pluralidad de primeros alelos específico de tejido y en donde el cálculo del número correspondiente de la segunda pluralidad de moléculas de ADN acelular que terminan en la posición genómica incluye: identificar si la molécula de ADN acelular que termina en la posición genómica incluye al menos uno de la pluralidad de primeros alelos específicos del tejido; incluir la molécula de ADN acelular en el número correspondiente cuando la molécula de ADN acelular incluye un primer alelo específico del tejido; y no incluir la molécula de ADN acelular en el número correspondiente cuando la molécula de ADN acelular no incluye un primer alelo específico del tejido.

15 La realización 20 incluye el método de una cualquiera de las realizaciones 1 a 19, en donde el primer tipo de tejido tiene una pluralidad de primeros alelos específicos del tejido en al menos una muestra adicional, y en donde el primer conjunto de posiciones genómicas se determina utilizando moléculas de ADN acelular de la al menos una muestra adicional que incluye al menos uno de la pluralidad de primeros alelos específicos del tejido. La realización 21 incluye el método de la reivindicación 20, en donde el segundo conjunto de posiciones genómicas es tal que los extremos de las moléculas de ADN acelular de un segundo tipo de tejido se producen a una tasa superior al umbral en la al menos una muestra adicional, en donde el segundo tipo de tejido tiene una pluralidad de segundos alelos específicos del tejido en la al menos una muestra adicional, y en donde el segundo conjunto de posiciones genómicas se determina utilizando moléculas de ADN acelular de la al menos una muestra adicional que incluye al menos uno de la pluralidad de segundos alelos específicos de tejido. La realización 22 incluye el método de la realización 21, en donde la al menos una muestra adicional es de una mujer embarazada, y en donde el primer tipo de tejido es tejido fetal y el segundo tipo de tejido es tejido materno. La realización 23 incluye el método de la realización 21 o 22, en donde las posiciones genómicas en las que los extremos de las moléculas de ADN acelular que tienen un alelo compartido entre el primer tipo de tejido y el segundo tipo de tejido se producen a una segunda tasa por encima del umbral se excluyen del primer conjunto de posiciones genómicas y se excluyen del segundo conjunto de posiciones genómicas.

30 La realización 24 incluye el método de una cualquiera de las realizaciones 1 a 23, en donde la abundancia relativa incluye una relación del primer número y el segundo número. La realización 25 incluye el método de una cualquiera de las realizaciones 1 a 24, en donde la pluralidad de ventanas tiene una amplitud de una posición genómica y en donde la abundancia relativa se calcula mediante: para cada posición genómica del primer conjunto de posiciones genómicas: calcular un número correspondiente de la primera pluralidad de moléculas de ADN acelular que terminan en la posición genómica como parte de la determinación de que el primer número de la primera pluralidad de moléculas de ADN acelular terminan en una cualquiera del primer conjunto de posiciones genómicas; calcular un tercer número de la primera pluralidad de moléculas de ADN acelular que abarcan la posición genómica y que no terminan en la posición genómica como parte de la determinación del segundo número de moléculas de ADN acelular; calcular una primera relación entre el número correspondiente y el tercer número; calcular una media de las primeras proporciones como la abundancia relativa. La realización 26 incluye el método de una cualquiera de las realizaciones 1 a 24, en donde la abundancia relativa se calcula mediante: para cada posición genómica del primer conjunto de posiciones genómicas: calcular un número correspondiente de la primera pluralidad de moléculas de ADN acelular que terminan dentro de una primera ventana que incluye la posición genómica como parte de la determinación de que el primer número de la primera pluralidad de moléculas de ADN acelular terminan dentro de una de la pluralidad de ventanas; computar un tercer número de la primera pluralidad de moléculas de ADN acelular que termina dentro de una segunda ventana que incluye la posición genómica, siendo la segunda ventana más grande que la primera; calcular una primera relación del número correspondiente y el tercer número; calcular una media de las primeras relaciones como abundancia relativa.

50 La realización 27 incluye el método de una cualquiera de las realizaciones 1 a 26, en donde el segundo conjunto de posiciones genómicas y el primer conjunto de posiciones genómicas no se solapan. La realización 28 incluye el método de una cualquiera de las realizaciones 1 a 27, en donde el segundo conjunto de posiciones genómicas incluye todas las posiciones genómicas correspondientes a un extremo de al menos una de la primera pluralidad de moléculas de ADN acelular. La realización 29 incluye el método de una cualquiera de las realizaciones 1 a 28, en donde analizar una o más de las moléculas de ADN acelular incluye determinar ambas posiciones genómicas correspondientes a ambos extremos de la molécula de ADN acelular. La realización 30 incluye el método de una cualquiera de las realizaciones 1 a 29, en donde la clasificación de la contribución proporcional corresponde a un intervalo por encima de un porcentaje especificado. La realización 31 incluye el método de una cualquiera de las realizaciones 1 a 30, en donde el primer tipo de tejido es un tumor. La realización 32 incluye el método de la realización 31, en donde la clasificación se selecciona entre un grupo que consiste en: una cantidad de tejido tumoral en el sujeto, un tamaño del tumor en el sujeto, un estadio del tumor en el sujeto, una carga tumoral en el sujeto, y la presencia de metástasis tumorales en el sujeto.

65 La realización 33 incluye el método de una cualquiera de las realizaciones 1 a 32, en donde una o más muestras biológicas adicionales son del sujeto y se obtienen en un momento diferente que la muestra biológica. La realización 34 incluye el método de una cualquiera de las realizaciones 1 a 33, que comprende además: obtener moléculas de

ADN molde de la muestra biológica a analizar; preparar una biblioteca de secuenciación de moléculas de ADN analizables usando las moléculas de ADN molde, la preparación de la biblioteca de secuenciación de moléculas de ADN analizables no incluye un paso de amplificación de ADN de las moléculas de ADN molde; secuenciar la biblioteca de secuenciación de moléculas de ADN analizables para obtener una pluralidad de lecturas de secuencia correspondientes a la primera pluralidad de moléculas de ADN acelular, en donde analizar la primera pluralidad de moléculas de ADN acelular incluye: recibir, en el sistema informático, la pluralidad de lecturas de secuencia; alinear, por el sistema informático, la pluralidad de lecturas de secuencia con el genoma de referencia para determinar posiciones genómicas para la pluralidad de lecturas de secuencia. La realización 35 incluye el método de una cualquiera de las realizaciones 1 a 34, que comprende además proporcionar una intervención terapéutica basada en la clasificación o realizar la obtención de imágenes del sujeto basada en la clasificación. La realización 36 incluye el método de una cualquiera de las realizaciones 1 a 35, en donde el primer conjunto de posiciones genómicas comprende entre 600 y 10.000 posiciones genómicas.

La realización 37 incluye un método de análisis de una muestra biológica, incluyendo una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido, para determinar una clasificación de una contribución proporcional del primer tipo de tejido en la mezcla, comprendiendo el método: identificar al menos una región genómica que tiene un patrón de fragmentación específico del primer tipo de tejido; analizar una pluralidad de moléculas de ADN acelular de la muestra biológica, en donde analizar una molécula de ADN acelular incluye: determinar una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular; identificar un primer conjunto de primeras posiciones genómicas, teniendo cada primera posición genómica un mínimo local de extremos de moléculas de ADN acelular correspondientes a la primera posición genómica; identificar un segundo conjunto de segundas posiciones genómicas, teniendo cada segunda posición genómica un máximo local de extremos de moléculas de ADN acelular correspondientes a la segunda posición genómica; determinar un primer número de moléculas de ADN acelular que terminan en una cualquiera de las primeras posiciones genómicas en una cualquiera de la al menos una región genómica; determinar un segundo número de moléculas de ADN acelular que terminan en una cualquiera de las segundas posiciones genómicas en una cualquiera de la al menos una región genómica; calcular un valor de separación utilizando el primer número y el segundo número; y determinar la clasificación de la contribución proporcional del primer tipo de tejido comparando el valor de separación con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido son conocidas.

La realización 38 incluye el método de la realización 37, en donde el primer conjunto de primeras posiciones genómicas incluye múltiples posiciones genómicas, en donde el segundo conjunto de segundas posiciones genómicas incluye múltiples posiciones genómicas, en donde determinar el primer número de moléculas de ADN acelular incluye determinar una primera cantidad de moléculas de ADN acelular que terminan en cada primera posición genómica, determinando así una pluralidad de primeras cantidades, en donde la determinación del segundo número de moléculas de ADN acelular incluye la determinación de una segunda cantidad de moléculas de ADN acelular que terminan en cada segunda posición genómica, determinando así una pluralidad de segundas cantidades, y en donde el cálculo del valor de separación incluye: determinar una pluralidad de relaciones separadas, cada relación separada de una de la pluralidad de primeras cantidades y una de la pluralidad de segundas cantidades, y determinar el valor de separación usando la pluralidad de relaciones separadas. La realización 39 incluye el método de la realización 37 o 38, en donde la al menos una región genómica incluye uno o más sitios de hipersensibilidad a DNasa. La realización 40 incluye el método de las realizaciones 37 a 38, en donde cada una de la al menos una región genómica que tiene un patrón de fragmentación específico del primer tipo de tejido incluye uno o más alelos específicos del primer tejido en al menos una muestra adicional. La realización 41 incluye el método de la realización 37 o 38, en donde la al menos una región genómica incluye uno o más sitios de nucleasa ATAC-seq o de micrococos. La realización 42 incluye el método de una cualquiera de las realizaciones 37 a 41, en donde las moléculas de ADN acelular alineadas con una posición genómica del primer conjunto de posiciones genómicas se extienden un número especificado de nucleótidos a ambos lados de la posición genómica. La realización 43 incluye el método de la realización 42, en donde el número especificado está entre 10 y 80 nucleótidos. La realización 44 incluye el método de una cualquiera de las realizaciones 37 a 43, en donde la identificación del primer conjunto de primeras posiciones genómicas incluye: para cada una de una pluralidad de posiciones genómicas: determinar una primera cantidad de moléculas de ADN acelular que se localizan en la posición genómica y se extienden un número especificado de nucleótidos a ambos lados de la posición genómica; determinar una segunda cantidad de moléculas de ADN acelular que se localizan en la posición genómica; y determinar una relación de la primera cantidad y la segunda cantidad; e identificar una pluralidad de mínimos locales y una pluralidad de máximos locales en las relaciones. La realización 45 incluye el método de una cualquiera de las realizaciones 37 a 44, en donde la mezcla es plasma o suero. La realización 46 incluye el método de una cualquiera de las realizaciones 37 a 45, en donde la pluralidad de moléculas de ADN acelular es de al menos 1.000 moléculas de ADN acelular. La realización 47 incluye el método de una cualquiera de las realizaciones 37 a 46, en donde, para una posición genómica dada de la pluralidad de posiciones genómicas, la segunda cantidad corresponde a un número total de moléculas de ADN acelular que se alinean con la posición genómica dada.

La realización 48 incluye un método de análisis de una muestra biológica, que incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido, para determinar un genotipo del primer tipo de tejido, teniendo el primer tipo de tejido potencialmente un genotipo diferente de otros tipos de tejidos de la pluralidad de tipos de tejidos, comprendiendo el método: identificar una primera posición genómica en la que se

producen extremos de moléculas de ADN acelular del primer tipo de tejido a una tasa superior a un umbral; analizar, mediante un sistema informático, una primera pluralidad de moléculas de ADN acelular de la muestra biológica de un sujeto, en donde analizar una molécula de ADN acelular incluye: determinar una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular; basándose en el análisis de la

5 primera pluralidad de moléculas de ADN acelular, identificar un conjunto de moléculas de ADN acelular que terminan en la primera posición genómica; para cada uno del conjunto de moléculas de ADN acelular: determinar una base correspondiente que se produce en la primera posición genómica, determinando así bases correspondientes en la primera posición genómica; determinar el genotipo del primer tipo de tejido en la primera posición genómica usando las bases correspondientes que se producen en la primera posición genómica en el conjunto de moléculas de ADN

10 acelular. La realización 49 incluye el método de la realización 48, que comprende, además: filtrar el conjunto de moléculas de ADN acelular para excluir o modificar una ponderación de al menos una de las moléculas de ADN acelular que terminan en la primera posición genómica, en donde el genotipo se determina utilizando un conjunto filtrado de moléculas de ADN acelular. La realización 50 incluye el método de la realización 49, en donde el filtrado utiliza al menos uno de: un tamaño de una molécula de ADN acelular, un estado de metilación de la molécula de ADN

15 acelular en una o más posiciones, y si la molécula de ADN acelular cubre una o más posiciones genómicas en las que se producen extremos de moléculas de ADN acelular del primer tipo de tejido a una tasa superior a un umbral. La realización 51 incluye el método de la realización 49 o 50, en donde el filtrado asigna una ponderación a la molécula de ADN acelular correspondiente a una probabilidad de que la molécula de ADN acelular sea del primer tipo de tejido, comprendiendo el método además: determinar una suma ponderada para cada una de una pluralidad de bases; y

20 determinar una contribución porcentual para cada una de la pluralidad de bases usando las sumas ponderadas, en donde el genotipo se determina usando las contribuciones porcentuales. La realización 52 incluye el método de una cualquiera de las realizaciones 48 a 51, en donde determinar el genotipo del primer tipo de tejido en la primera posición genómica incluye: determinar una contribución porcentual para cada una de una pluralidad de bases; y comparar cada una de las contribuciones porcentuales con uno o más valores de corte. La realización 53 incluye el método de la

25 realización 52, en donde un primer valor de corte de uno o más valores de corte corresponde a un genotipo homocigótico de una primera base cuando la contribución porcentual de la primera base está por encima del primer valor de corte. La realización 54 incluye el método de la realización 52, en donde un primer valor de corte y un segundo valor de corte de uno o más valores de corte corresponden a un genotipo heterocigoto para una primera base y una segunda base cuando las contribuciones porcentuales de la primera base y la segunda base están por encima del

30 primer valor de corte y por debajo del segundo valor de corte. La realización 55 incluye el método de una cualquiera de las realizaciones 48 a 54, en donde el primer tipo de tejido corresponde a un tumor. La realización 56 incluye el método de una cualquiera de las realizaciones 48 a 55, en donde el primer tipo de tejido corresponde a un feto, y en donde el sujeto está embarazado del feto.

35 La realización 57 incluye un método de análisis de una muestra biológica, que incluye una mezcla de moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye un primer tipo de tejido, comprendiendo el método: analizar, mediante un sistema informático, una pluralidad de moléculas de ADN acelular de la muestra biológica de un sujeto, teniendo cada una de la pluralidad de moléculas de ADN acelular un extremo izquierdo y un extremo derecho, en donde analizar una molécula de ADN acelular incluye: determinar una posición final izquierda en un genoma de

40 referencia correspondiente al extremo izquierdo de la molécula de ADN acelular; determinar una posición final derecha en el genoma de referencia correspondiente al extremo derecho de la molécula de ADN acelular; identificar un conjunto izquierdo de posiciones genómicas izquierdas, teniendo cada una un máximo local de extremos izquierdos de la pluralidad de moléculas de ADN acelular correspondiente a uno de los conjuntos izquierdos de posiciones genómicas; identificar un conjunto derecho de posiciones genómicas derechas, teniendo cada una un máximo local de extremos

45 derechos de la pluralidad de moléculas de ADN acelular correspondientes a uno de los conjuntos derechos de posiciones genómicas; identificar un primer conjunto de posiciones genómicas como específicas del primer tipo de tejido mediante: la comparación de las posiciones genómicas izquierdas del conjunto izquierdo con las posiciones genómicas derechas del conjunto derecho para identificar el primer conjunto de posiciones genómicas donde una distancia desde una posición genómica izquierda a una posición genómica derecha más cercana es mayor que una

50 primera distancia umbral, siendo la primera distancia umbral de al menos 5 posiciones genómicas en el genoma de referencia. La realización 58 incluye el método de la realización 57, que comprende además: identificar un segundo conjunto de posiciones genómicas mediante: la comparación de las posiciones genómicas izquierdas del conjunto izquierdo con las posiciones genómicas derechas del conjunto derecho para identificar el segundo conjunto de posiciones genómicas donde la distancia de una posición genómica izquierda a una posición genómica derecha más cercana es menor que una segunda distancia umbral; la determinación de un valor de separación usando un primer

55 número de la pluralidad de moléculas de ADN acelular que terminan en una del conjunto izquierdo de posiciones genómicas izquierdas y un segundo número de la pluralidad de moléculas de ADN acelular que terminan en una del conjunto derecho de posiciones genómicas derechas; y la determinación de una clasificación de una contribución proporcional del primer tipo de tejido comparando el valor de separación con uno o más valores de calibración

60 determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido son conocidas. La realización 59 incluye el método de la realización 58, en donde determinar el valor de separación incluye: identificar pares del primer conjunto de posiciones genómicas y el segundo conjunto de posiciones genómicas; para cada uno de los pares: determinar una primera cantidad de moléculas de ADN acelular que terminan en una primera posición genómica del par; y determinar una segunda cantidad de moléculas de ADN acelular que

65 terminan en una segunda posición genómica del par, en donde las primeras cantidades de moléculas de ADN acelular corresponden al primer número de la pluralidad de moléculas de ADN acelular y las segundas cantidades de moléculas

de ADN acelular corresponden al segundo número de la pluralidad de moléculas de ADN acelular. La realización 60 incluye el método de la realización 59, en donde determinar el valor de separación incluye: para cada uno de los pares: determinar una relación que incluye la primera cantidad y la segunda cantidad; y determinar el valor de separación a partir de las relaciones. La realización 61 incluye el método de la realización 59 o 60, en donde los pares del primer conjunto de posiciones genómicas y el segundo conjunto de posiciones genómicas son los más cercanos entre sí. La realización 62 incluye el método de una cualquiera de las realizaciones 57 a 61, en donde la segunda distancia umbral es inferior a 5 posiciones genómicas en el genoma de referencia. La realización 63 incluye el método de una cualquiera de las realizaciones 57 a 62, en donde el primer conjunto de posiciones genómicas incluye tanto posiciones genómicas izquierdas como posiciones genómicas derechas.

La realización 64 incluye un método para determinar una contribución proporcional de un primer tejido en una mezcla de ADN, comprendiendo el método: identificar sitios de hipersensibilidad a DNasa que son específicos del primer tejido; analizar una pluralidad de moléculas de ADN acelular de la muestra biológica, en donde analizar una molécula de ADN acelular incluye: identificar una ubicación de la molécula de ADN acelular en un genoma humano de referencia, incluyendo la ubicación ambos extremos de la molécula de ADN acelular; identificar un primer conjunto de primeras posiciones genómicas, cada una de las cuales tiene un mínimo local de moléculas de ADN acelular alineadas con la posición genómica y que se extiende un número especificado de nucleótidos a ambos lados de la posición genómica; identificar un segundo conjunto de segundas posiciones genómicas, teniendo cada una un máximo local de moléculas de ADN acelular alineadas con la posición genómica y que se extienden el número especificado de nucleótidos a ambos lados de la posición genómica; calcular un primer número de moléculas de ADN acelular que terminan en una de las primeras posiciones genómicas en uno de los sitios de hipersensibilidad a DNasa; calcular un segundo número de moléculas de ADN acelular que terminan en una de las primeras posiciones genómicas en uno de los sitios de hipersensibilidad a DNasa; determinar una relación del primer número y el segundo número; y determinar la contribución proporcional del primer tejido basándose en la relación. La realización 65 incluye el método de la realización 64, en donde la identificación de un primer conjunto de primeras posiciones genómicas incluye: para cada una de una pluralidad de posiciones genómicas: determinar una primera cantidad de moléculas de ADN acelular que se localizan en el locus y se extienden un número especificado de nucleótidos a ambos lados del locus; determinar una segunda cantidad de moléculas de ADN acelular que se localizan en el locus; y determinar una primera relación de la primera cantidad y la segunda cantidad; e identificar una pluralidad de mínimos locales en las relaciones. La realización 66 incluye el método de una cualquiera de las realizaciones 64 o 65, en donde la mezcla de ADN es plasma o suero. La realización 66 incluye el método de una cualquiera de las realizaciones 64 a 66, en donde la pluralidad de moléculas de ADN acelular es de al menos 1.000 moléculas de ADN acelular.

La realización 67 incluye un método para determinar una contribución proporcional de un primer tejido en una mezcla de ADN, comprendiendo el método: identificar ubicaciones genómicas en las que los fragmentos de ADN tienen una frecuencia por encima de un umbral para un extremo del fragmento de ADN para el primer tejido; y analizar una pluralidad de moléculas de ADN acelular de la muestra biológica, en donde analizar una molécula de ADN acelular incluye: identificar una ubicación de la molécula de ADN acelular en un genoma humano de referencia, incluyendo la ubicación ambos extremos de la molécula de ADN acelular; calcular un primer número de moléculas de ADN acelular que terminan en una de las posiciones genómicas identificadas en uno de los sitios de hipersensibilidad a DNasa; calcular una relación a partir del primer número y una cantidad de ADN secuenciado; y determinar la contribución proporcional del primer tejido basándose en la relación. La realización 68 incluye el método de la reivindicación 67, en donde el primer tejido es un tumor. La realización 69 incluye el método de la reivindicación 67, en donde el primer tejido es tejido fetal. La realización 70 incluye un método para predecir si un fragmento de ADN que porta una supuesta mutación procede realmente de un tumor, comprendiendo el método: identificar localizaciones genómicas en las que los fragmentos de ADN tienen una frecuencia por encima de un umbral para un extremo del fragmento de ADN; y determinar una probabilidad basada en el fragmento de ADN que termina en una de las localizaciones genómicas identificadas.

La realización 71 incluye un producto informático que comprende un medio legible por ordenador que almacena una pluralidad de instrucciones para controlar un sistema informático para realizar una operación de cualquiera de las realizaciones de 1 a 70. La realización 72 incluye un sistema que comprende: el producto informático de la realización 71; y uno o más procesadores para ejecutar las instrucciones almacenadas en el medio legible por ordenador. La realización 73 incluye un sistema que comprende medios para realizar cualquiera de las realizaciones 1 a 70. La realización 74 incluye un sistema configurado para realizar cualquiera de las realizaciones 1 a 70. La realización 75 incluye un sistema que comprende módulos que realizan respectivamente las etapas de cualquiera de las realizaciones 1 a 70.

XII. SISTEMA INFORMÁTICOS

Cualquiera de los sistemas informáticos mencionados en este documento puede utilizar cualquier número adecuado de subsistemas. En la FIG. 39 relacionada con el aparato informático 10 se muestran ejemplos dichos subsistemas. En algunas realizaciones, un sistema informático incluye un único aparato informático, donde los subsistemas pueden ser los componentes del aparato informático. En otras realizaciones, un sistema informático puede incluir varios aparatos informáticos, siendo cada uno de ellos un subsistema, con componentes internos. Un sistema informático puede incluir ordenadores de sobremesa y portátiles, tabletas informáticas, teléfonos móviles y otros dispositivos

móviles.

Los subsistemas mostrados en la FIG. 39 están interconectados a través de un bus de sistema 75. Se muestran subsistemas adicionales, como una impresora 74, teclado 78, dispositivos de almacenamiento 79, monitor 76 que está conectado a un adaptador de pantalla 82, y otros. Los dispositivos periféricos y de entrada/salida (E/S), que se acoplan al controlador de E/S 71, pueden conectarse al sistema informático mediante cualquier tipo de conexión conocida en la técnica, tal como el puerto de entrada/salida (E/S) 77 (por ejemplo, USB, FireWire®). Por ejemplo, el puerto de E/S 77 o la interfaz externa 81 (por ejemplo, Ethernet, Wi-Fi, etc.) pueden utilizarse para conectar el sistema informático 10 a una red de área amplia, como Internet, un dispositivo de entrada de ratón o un escáner. La interconexión mediante el bus de sistema 75 permite que el procesador central 73 se comuniquen con cada subsistema y así controlar la ejecución de diversas instrucciones procedentes de la memoria del sistema 72 o de los dispositivos de almacenamiento 79 (por ejemplo, un disco fijo, tal como un disco duro o un disco óptico), así como el intercambio de información entre subsistemas. La memoria del sistema 72 y/o los dispositivos de almacenamiento 79 pueden ser realizaciones de un medio legible por ordenador. Otro subsistema es un dispositivo de recogida de datos 85, tal como una cámara, micrófono, acelerómetro y similares. Cualquiera de los datos mencionados en el presente documento puede ser la salida de un componente a otro componente y puede ser la salida al usuario.

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

Los aspectos de las realizaciones pueden implementarse en forma de lógica de control utilizando hardware (por ejemplo, un circuito integrado de aplicación específica o una matriz de puertas programables en campo) y/o utilizando software informático con un procesador generalmente programable de forma modular o integrada. Como se usa en el presente documento, un procesador incluye un procesador de un solo núcleo, un procesador de varios núcleos en un mismo chip integrado, o múltiples unidades de procesamiento en una sola placa de circuito o en red. Basándose en la divulgación y las técnicas proporcionadas en el presente documento, un experto habitual en la materia conocerá y percibirá otras formas y/o métodos para implementar las realizaciones de la presente invención utilizando hardware y una combinación de hardware y software.

Puede implementarse cualquiera de los componentes o las funciones de software descritos en la presente solicitud como código de software para su ejecución mediante un procesador utilizando cualquier lenguaje informático adecuado, tal como, por ejemplo, Java, C, C++, C#, Objective-C, Swift, o lenguaje de scripting como Perl o Python utilizando, por ejemplo, técnicas convencionales u orientadas a objetivos. El código de software puede almacenarse como una serie de instrucciones o comandos en un medio legible por ordenador para su almacenamiento y/o transmisión. Un medio legible por ordenador no transitorio adecuado puede incluir una memoria de acceso aleatorio (RAM), una memoria de solo lectura (ROM), un medio magnético como un disco duro o disquete o un medio óptico tal como un disco compacto (CD) o un DVD (disco digital versátil), memoria flash y similares. El medio legible por ordenador puede ser cualquier combinación de tales dispositivos de almacenamiento o transmisión.

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

Cualquiera de los métodos descritos en el presente documento puede realizarse total o parcialmente con un sistema informático que incluya uno o más procesadores, que pueden configurarse para realizar las etapas. Por tanto, las realizaciones pueden dirigirse a sistemas informáticos configurados para realizar los pasos de cualquiera de los métodos descritos en el presente documento, potencialmente con diferentes componentes que realizan una etapa o un grupo de etapas respectivas. Aunque se presentan como pasos numerados, las etapas de los métodos descritos en el presente documento pueden realizarse al mismo tiempo o en un orden diferente. Además, pueden usarse partes de estas etapas con partes de otras etapas de otros métodos. Asimismo, la totalidad de una etapa o una parte de la misma puede ser opcional. Además, cualquiera de los pasos de cualquiera de los métodos puede realizarse con módulos, unidades, circuitos, u otros medios para realizar estas etapas.

La descripción anterior de las realizaciones de ejemplo de la invención se ha presentado con fines ilustrativos y descriptivos. No pretende ser exhaustiva o limitar la invención a la forma precisa descrita, y son posibles muchas modificaciones y variaciones a la luz de la enseñanza anterior.

Una cita de "un", "una" o "el/la" se entiende como "uno o más" a menos que se indique específicamente lo contrario. El uso de "o" pretende hacer referencia a un "o inclusivo", y no a un "o exclusivo", a menos que se indique específicamente lo contrario. La referencia a un "primer" componente no requiere necesariamente que se proporcione un segundo componente. Además, la referencia a un "primer" o un "segundo" componente no limita el componente referido a una ubicación concreta, a menos que se indique expresamente.

REIVINDICACIONES

1. Un método de análisis de una muestra biológica, comprendiendo el método:

5 identificar al menos una región genómica que tiene un patrón de fragmentación específico de un primer tipo de tejido;
 analizar una pluralidad de moléculas de ADN acelular de la muestra biológica, incluyendo la muestra biológica moléculas de ADN acelular de una pluralidad de tipos de tejidos que incluye el primer tipo de tejido, en donde analizar una molécula de ADN acelular incluye:
 10 determinar una posición genómica en un genoma de referencia correspondiente a al menos un extremo de la molécula de ADN acelular;
 identificar un primer conjunto de primeras posiciones genómicas, teniendo cada primera posición genómica un mínimo local de extremos de moléculas de ADN acelular correspondientes a la primera posición genómica;
 15 identificar un segundo conjunto de segundas posiciones genómicas, teniendo cada segunda posición genómica un máximo local de extremos de moléculas de ADN acelular correspondientes a la segunda posición genómica;
 determinar un primer número de moléculas de ADN acelular que terminan en una cualquiera de las primeras posiciones genómicas en una cualquiera de la al menos una región genómica;
 determinar un segundo número de moléculas de ADN acelular que terminan en una cualquiera de las segundas posiciones genómicas en una cualquiera de las al menos una región genómica;
 20 calcular un valor de separación utilizando el primer número y el segundo número; y
 determinar una clasificación de una contribución proporcional del primer tipo de tejido comparando el valor de separación con uno o más valores de calibración determinados a partir de una o más muestras de calibración cuyas contribuciones proporcionales del primer tipo de tejido son conocidas.

25 2. El método de la reivindicación 1, en donde el primer conjunto de primeras posiciones genómicas incluye múltiples posiciones genómicas, en donde el segundo conjunto de segundas posiciones genómicas incluye múltiples posiciones genómicas,

30 en donde determinar el primer número de moléculas de ADN acelular incluye determinar una primera cantidad de moléculas de ADN acelular que terminan en cada primera posición genómica, determinando de este modo una pluralidad de primeras cantidades, en donde determinar el segundo número de moléculas de ADN acelular incluye determinar una segunda cantidad de moléculas de ADN acelular que terminan en cada segunda posición genómica, determinando de este modo una pluralidad de segundas cantidades, y
 35 en donde calcular el valor de separación incluye:

determinar una pluralidad de relaciones separadas, cada relación separada de una de la pluralidad de primeras cantidades y una de la pluralidad de segundas cantidades, y
 determinar el valor de separación utilizando la pluralidad de relaciones separadas.

40 3. El método de la reivindicación 1 o 2, en donde la al menos una región genómica incluye (1) uno o más sitios de hipersensibilidad a DNasa, o (2) uno o más sitios de nucleasa ATAC-seq o de micrococos.

4. El método de una cualquiera de las reivindicaciones anteriores, en donde cada una de la al menos una región genómica que tiene el patrón de fragmentación específico del primer tipo de tejido incluye uno o más alelos específicos del primer tejido en al menos una muestra adicional.

5. El método de la reivindicación 4, en donde la una o más muestras de calibración incluyen la al menos una muestra adicional.

50 6. El método de la reivindicación 5, que comprende además:
 para cada muestra de calibración de la una o más muestras de calibración:

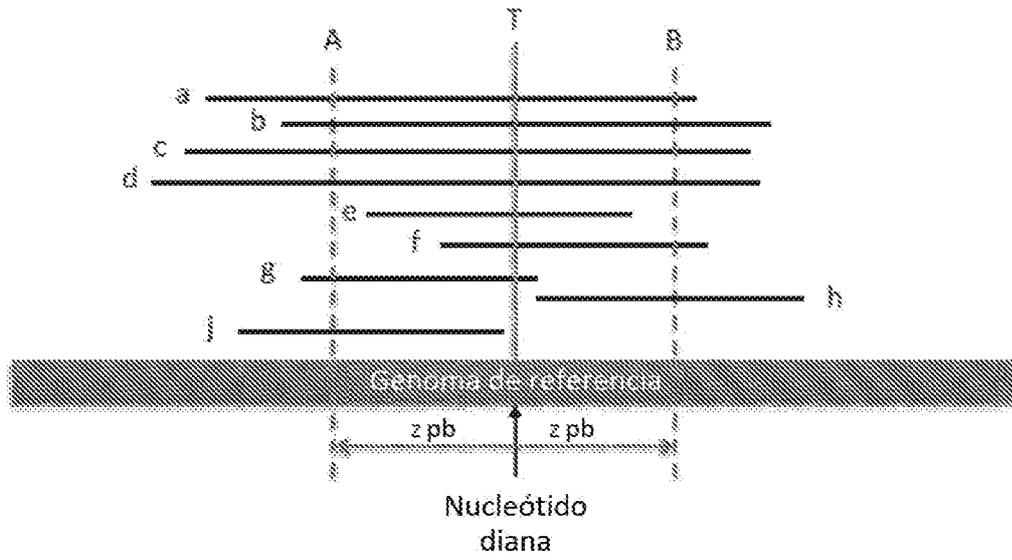
medir una contribución proporcional correspondiente del primer tipo de tejido; y
 55 determinar un valor de separación correspondiente utilizando (1) un primer número correspondiente de moléculas de ADN acelular que terminan en una cualquiera de las primeras posiciones genómicas en una cualquiera de la al menos una región genómica para la muestra de calibración y (2) un segundo número correspondiente de moléculas de ADN acelular que terminan en una cualquiera de las segundas posiciones genómicas en una cualquiera de la al menos una región genómica para la muestra de calibración, obteniendo de este modo uno o más puntos de datos de calibración, en donde cada punto de datos de calibración especifica la contribución proporcional medida del primer tipo de tejido para la al menos una muestra adicional y el valor de separación correspondiente.
 60

7. El método de la reivindicación 6, en donde el uno o más puntos de datos de calibración son una pluralidad de puntos de datos de calibración que forman una función de calibración que se aproxima a la pluralidad de puntos de datos de calibración.

65 8. El método de una cualquiera de las reivindicaciones anteriores, en donde las moléculas de ADN acelular alineadas

con una posición genómica del primer conjunto de primeras posiciones genómicas se extienden un número especificado de nucleótidos a ambos lados de la posición genómica.

- 5 9. El método de la reivindicación 8, en donde el número especificado está entre 10 y 80 nucleótidos.
- 10 10. El método de una cualquiera de las reivindicaciones anteriores, en donde identificar el primer conjunto de primeras posiciones genómicas incluye:
- 15 para cada una de una pluralidad de posiciones genómicas:
- determinar una primera cantidad de moléculas de ADN acelular que se localizan en la posición genómica y se extienden un número especificado de nucleótidos a ambos lados de la posición genómica; determinar una segunda cantidad de moléculas de ADN acelular que se localizan en la posición genómica; y determinar una relación de (1) la primera cantidad de moléculas de ADN acelular que se localizan en la posición genómica y se extienden un número especificado de nucleótidos a ambos lados de la posición genómica y (2) la segunda cantidad de moléculas de ADN acelular que se localizan en la posición genómica; e
- identificar una pluralidad de mínimos locales y una pluralidad de máximos locales en las relaciones.
- 20 11. El método de la reivindicación 8, en donde, para una posición genómica dada de la pluralidad de posiciones genómicas, la segunda cantidad de moléculas de ADN acelular que se localizan en la posición genómica corresponde a un número total de las moléculas de ADN acelular que se alinean con la posición genómica dada.
- 25 12. El método de una cualquiera de las reivindicaciones anteriores, en donde la muestra biológica es plasma o suero.
- 30 13. El método de una cualquiera de las reivindicaciones anteriores, en donde el primer tipo de tejido es un tumor, y la clasificación se selecciona de un grupo que consiste en: una cantidad de tejido tumoral en un sujeto, un tamaño del tumor en el sujeto, un estadio del tumor en el sujeto, una carga tumoral en el sujeto y presencia de metástasis tumorales en el sujeto.
- 35 14. El método de una cualquiera de las reivindicaciones 1 a 12, en donde la clasificación de la contribución proporcional corresponde a un intervalo por encima de un porcentaje especificado.
15. Un programa informático que comprende una pluralidad de instrucciones que pueden ejecutarse por un sistema informático, que cuando se ejecutan de este modo controlan el sistema informático para realizar el método de una cualquiera de las reivindicaciones anteriores.



Fragmentos que abarcan desde z nt cadena abajo hasta z nt cadena arriba del nucleótido diana : a, b, c y d

Fragmentos que abarcan el nucleótido diana: a, b, c, d, e, f y g

$$P_i = \frac{4}{7}$$

FIG. 1

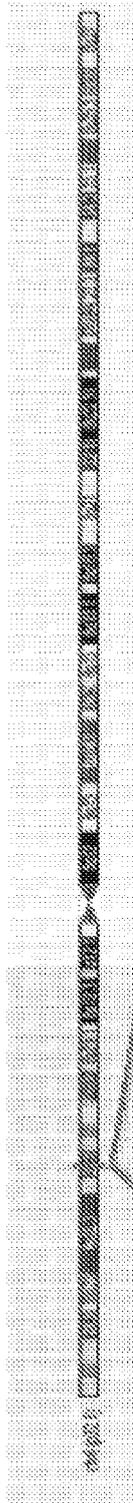


FIG. 2A

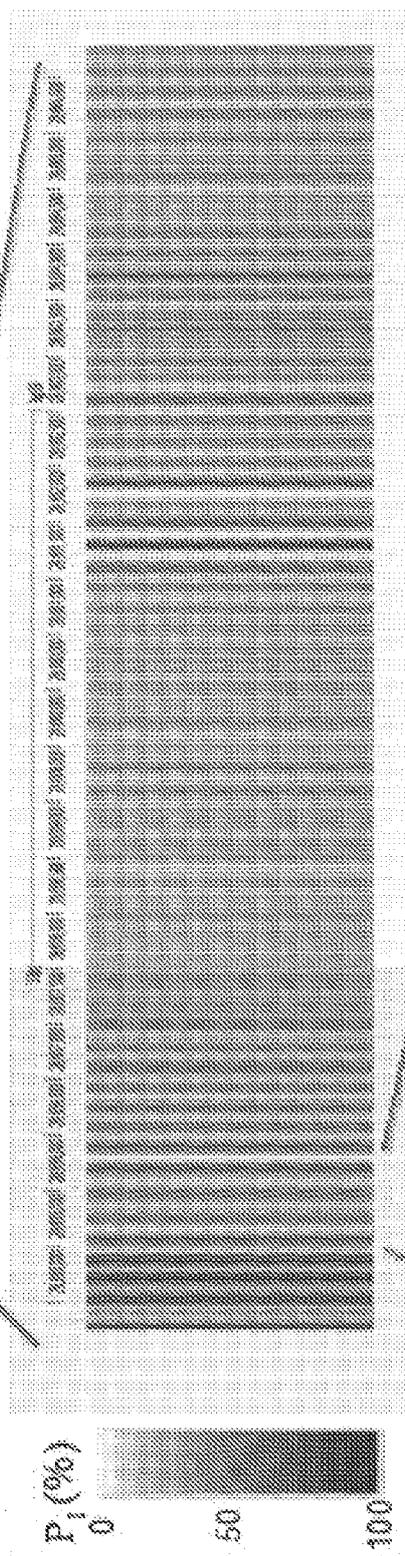
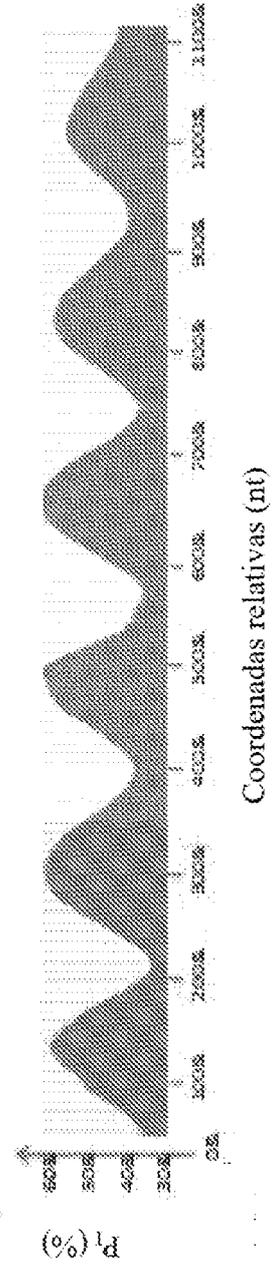


FIG. 2B



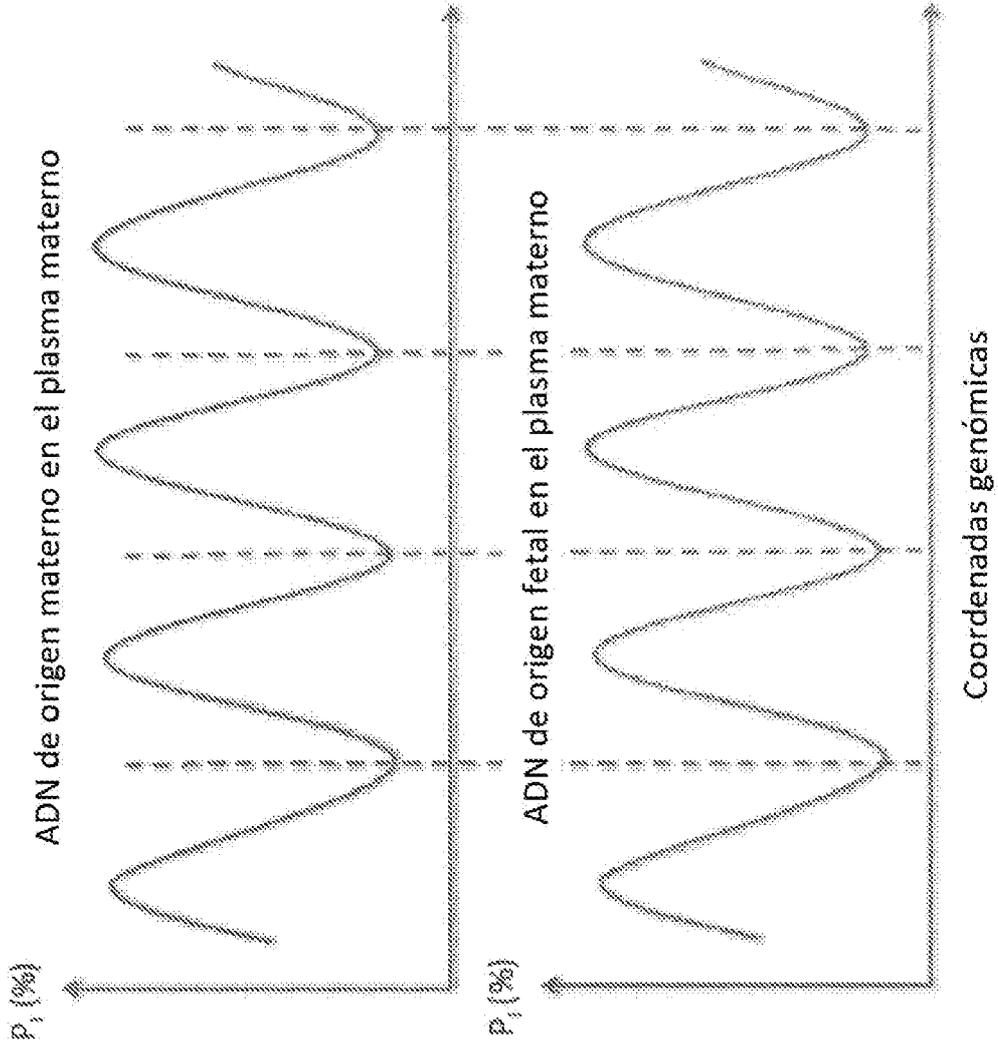


FIG. 3

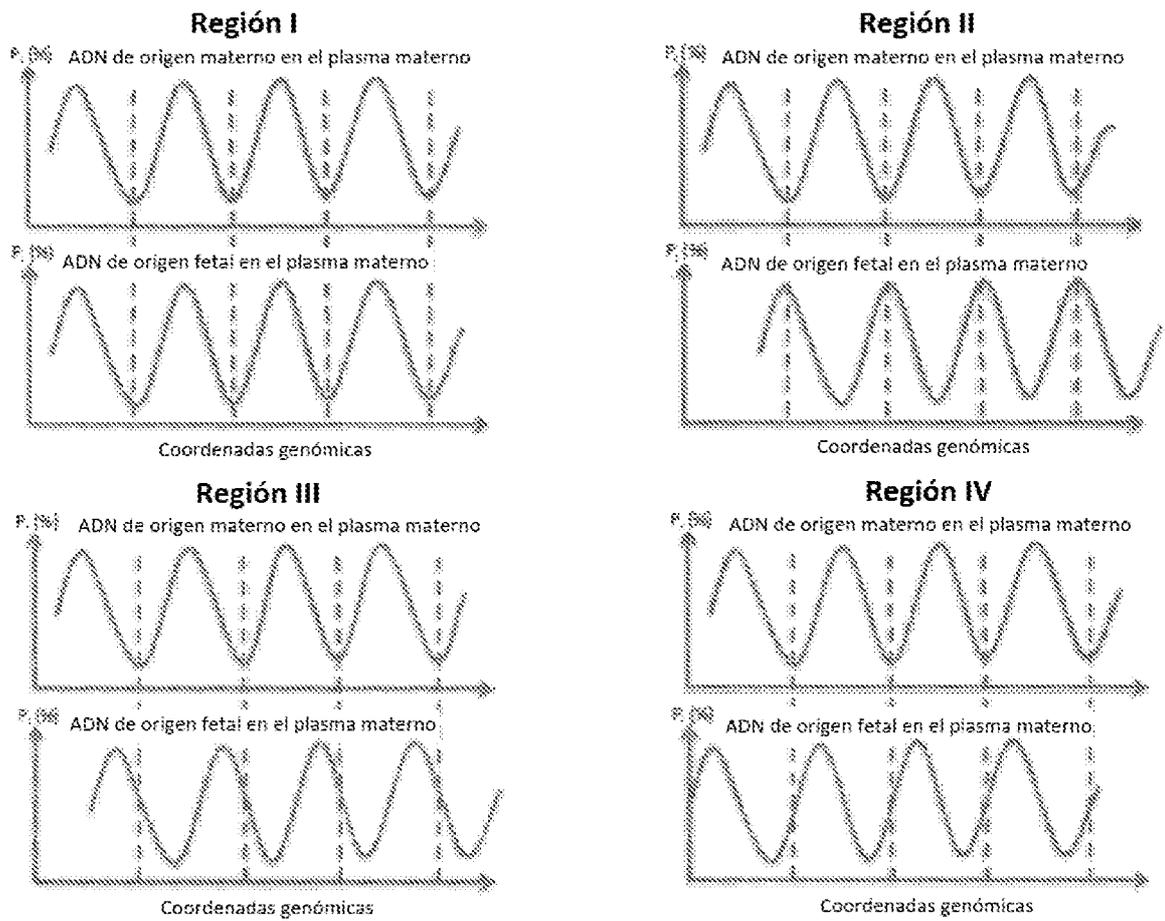


FIG. 4

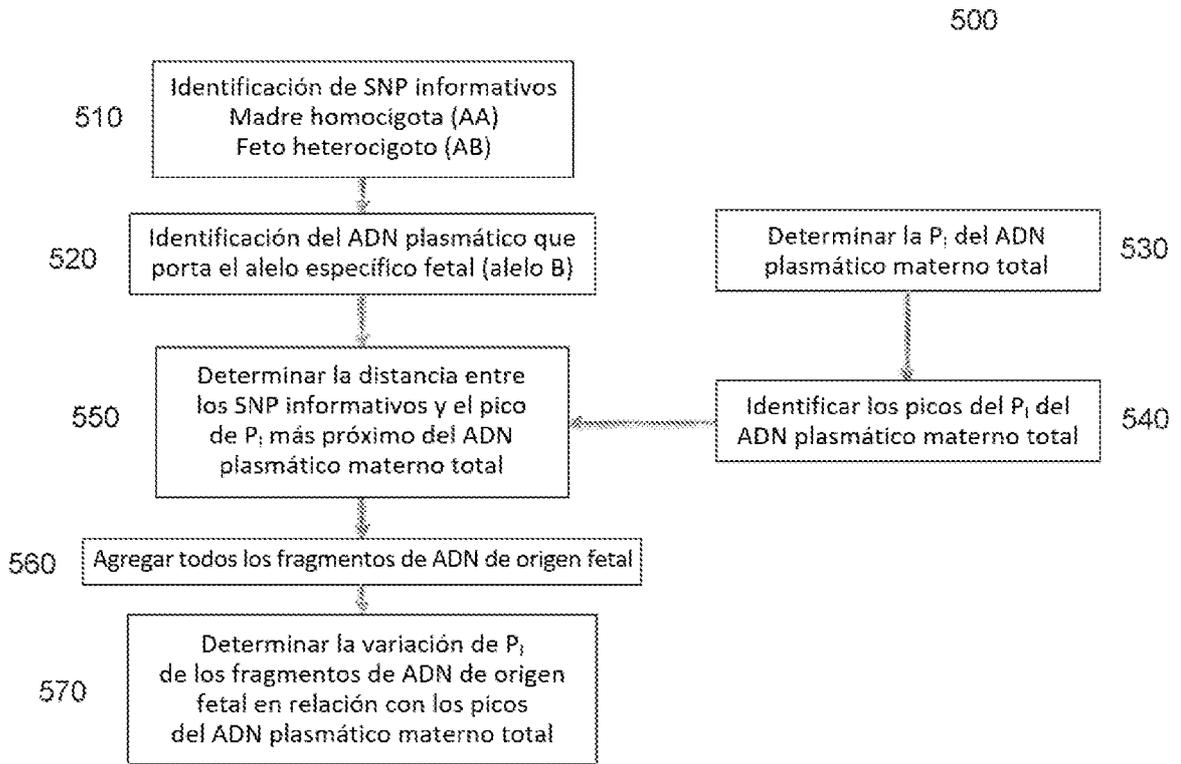


FIG. 5

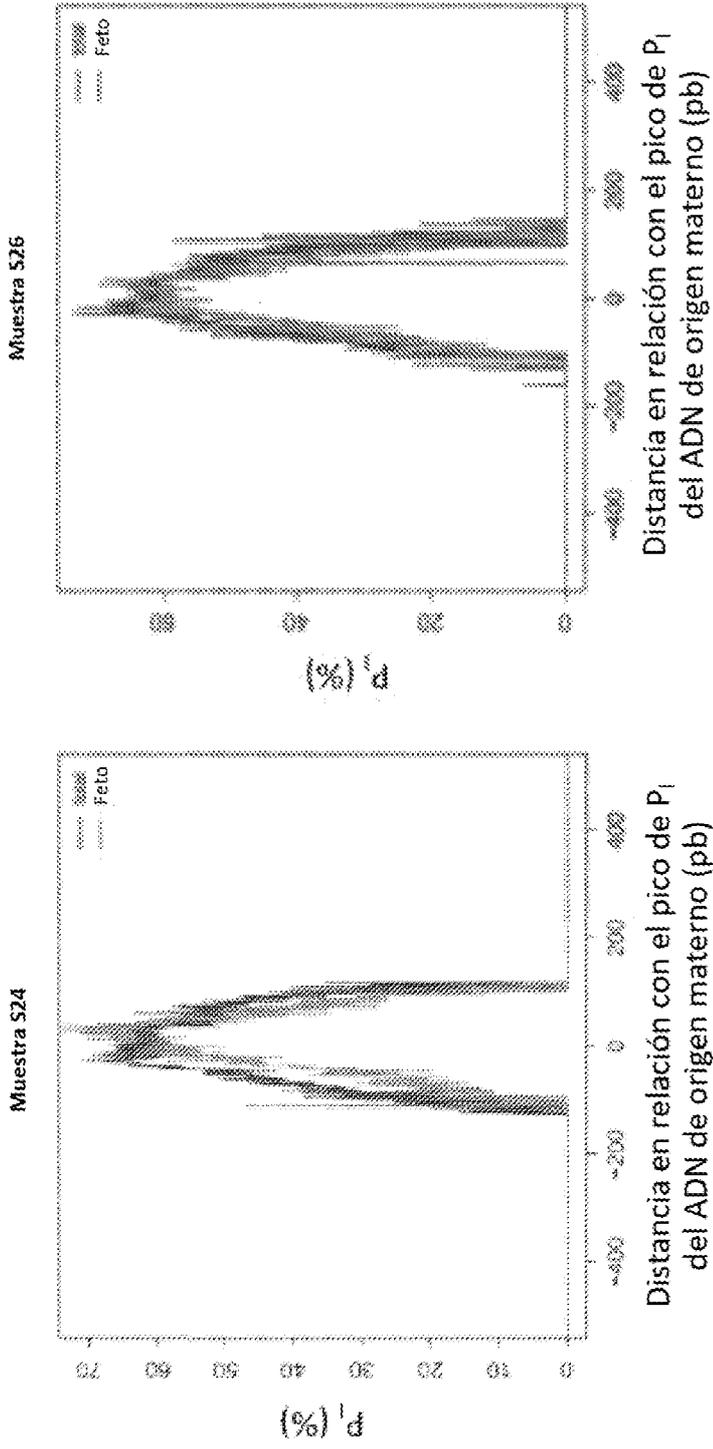


FIG. 6

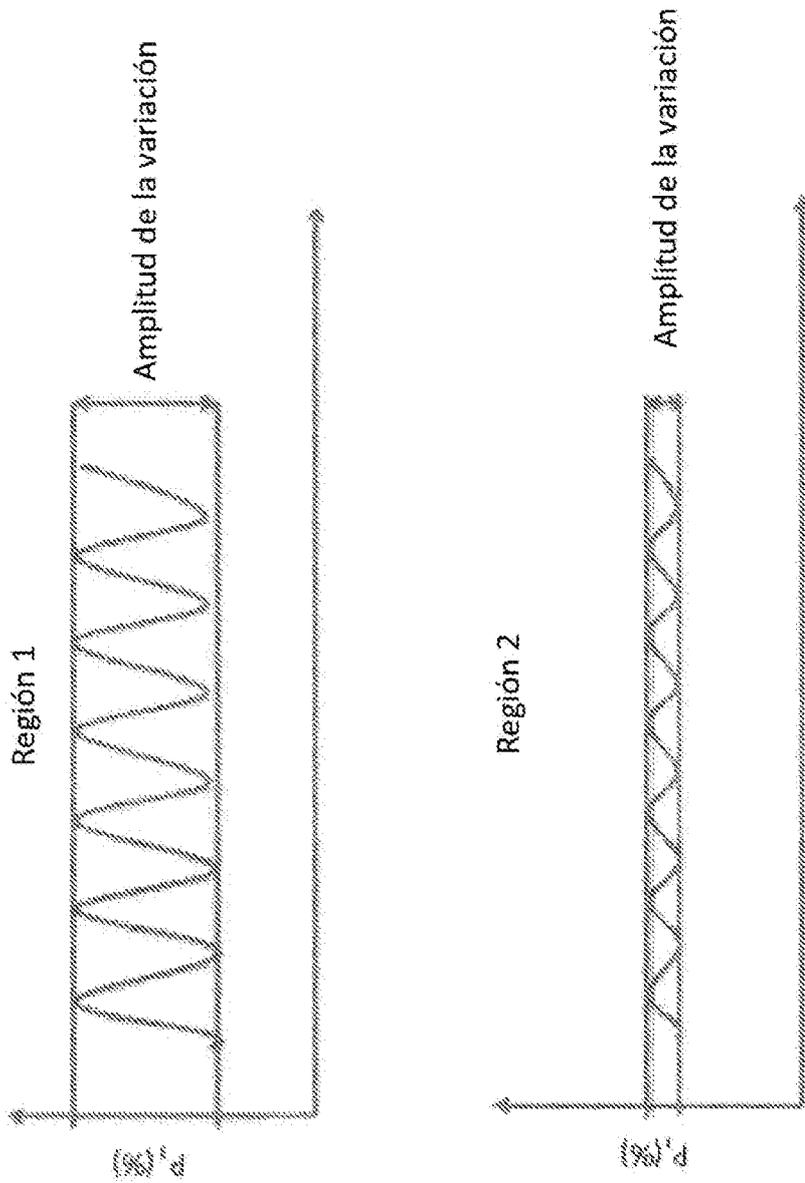


FIG. 7

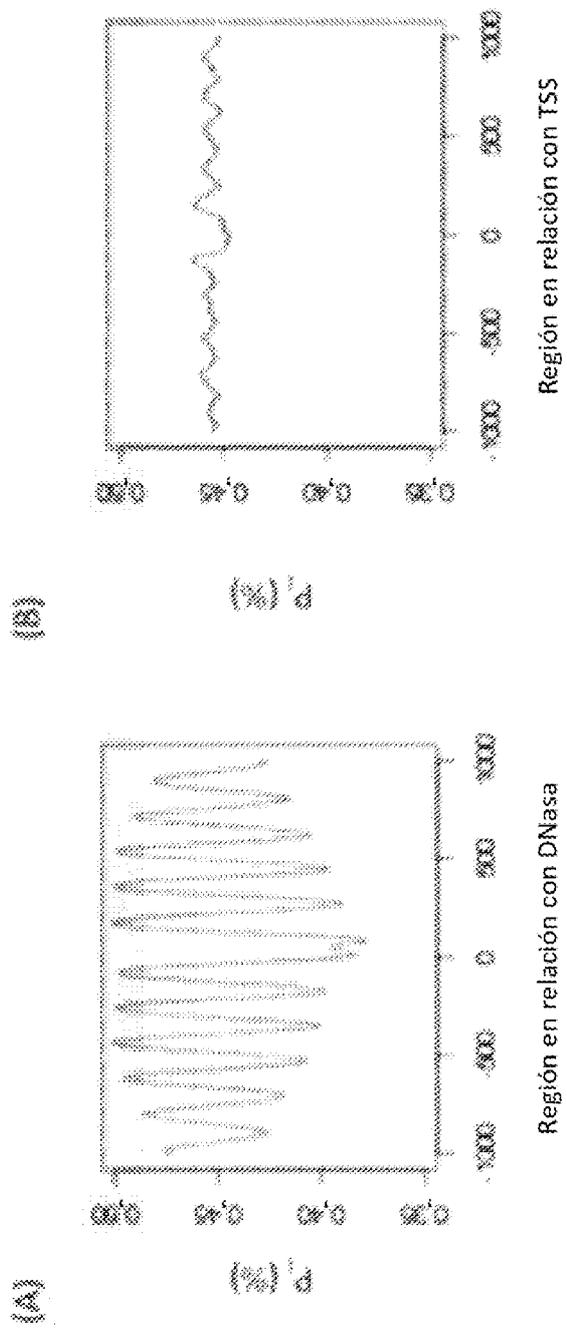


FIG. 8

Una región que es un sitio de hipersensibilidad a DNAsas para el tejido A pero no para el tejido B

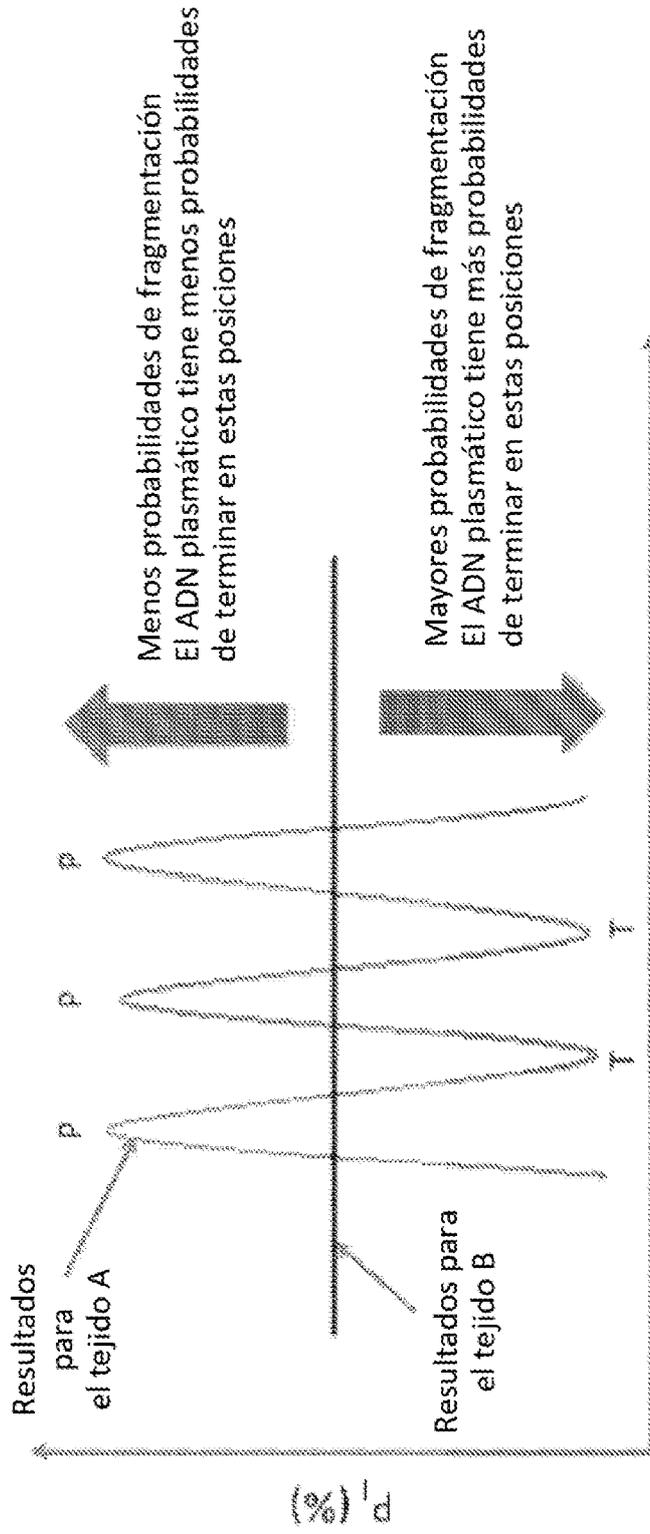
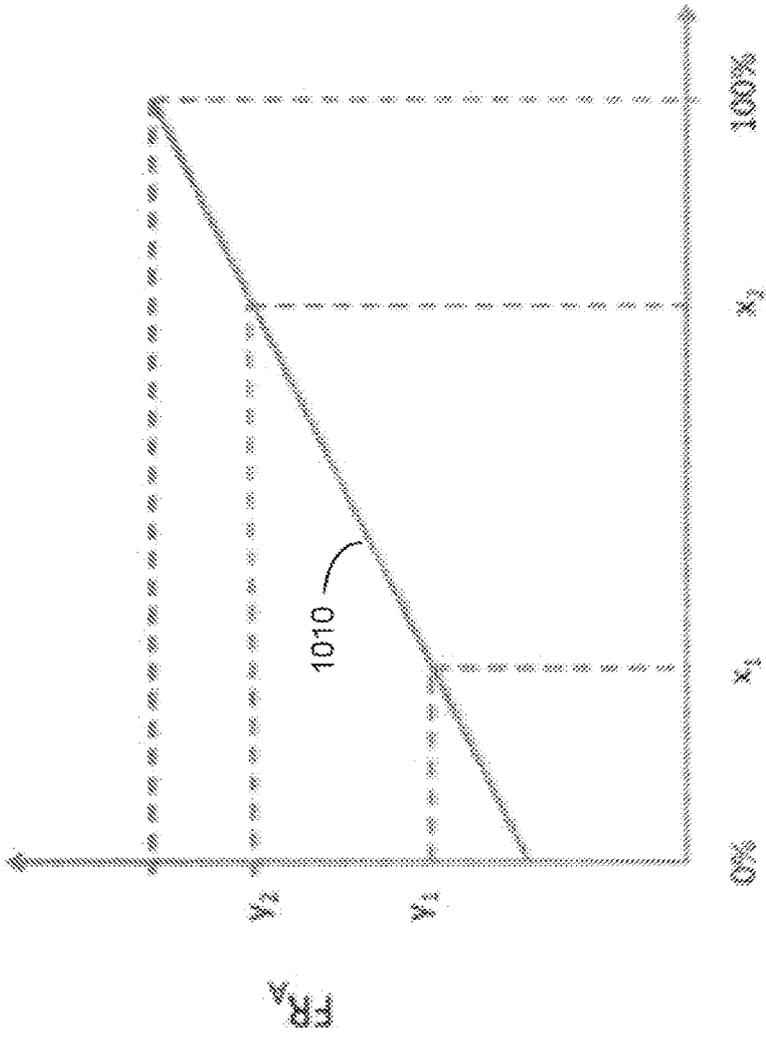


FIG. 9



Contribución proporcional del tejido A en una mezcla de ADN plasmático procedente de los tejidos A y B

FIG. 10

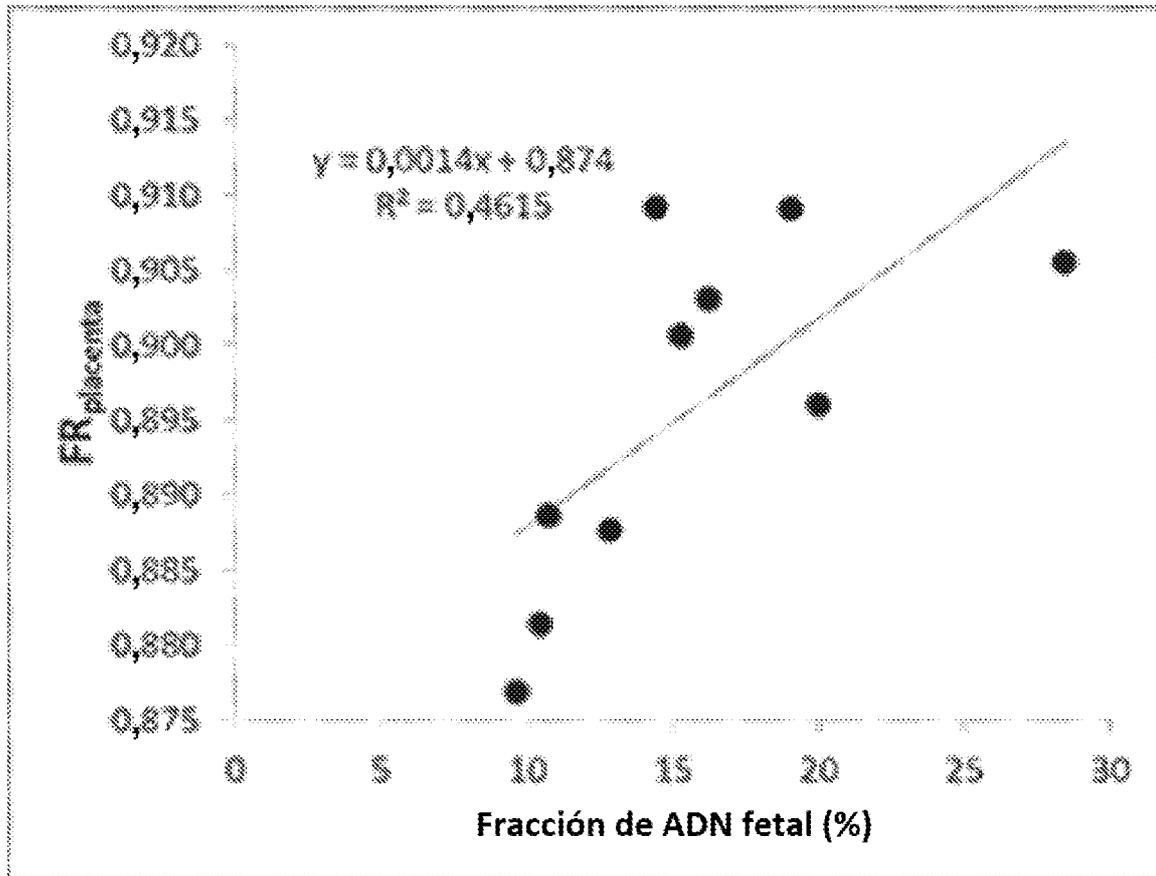


FIG. 11

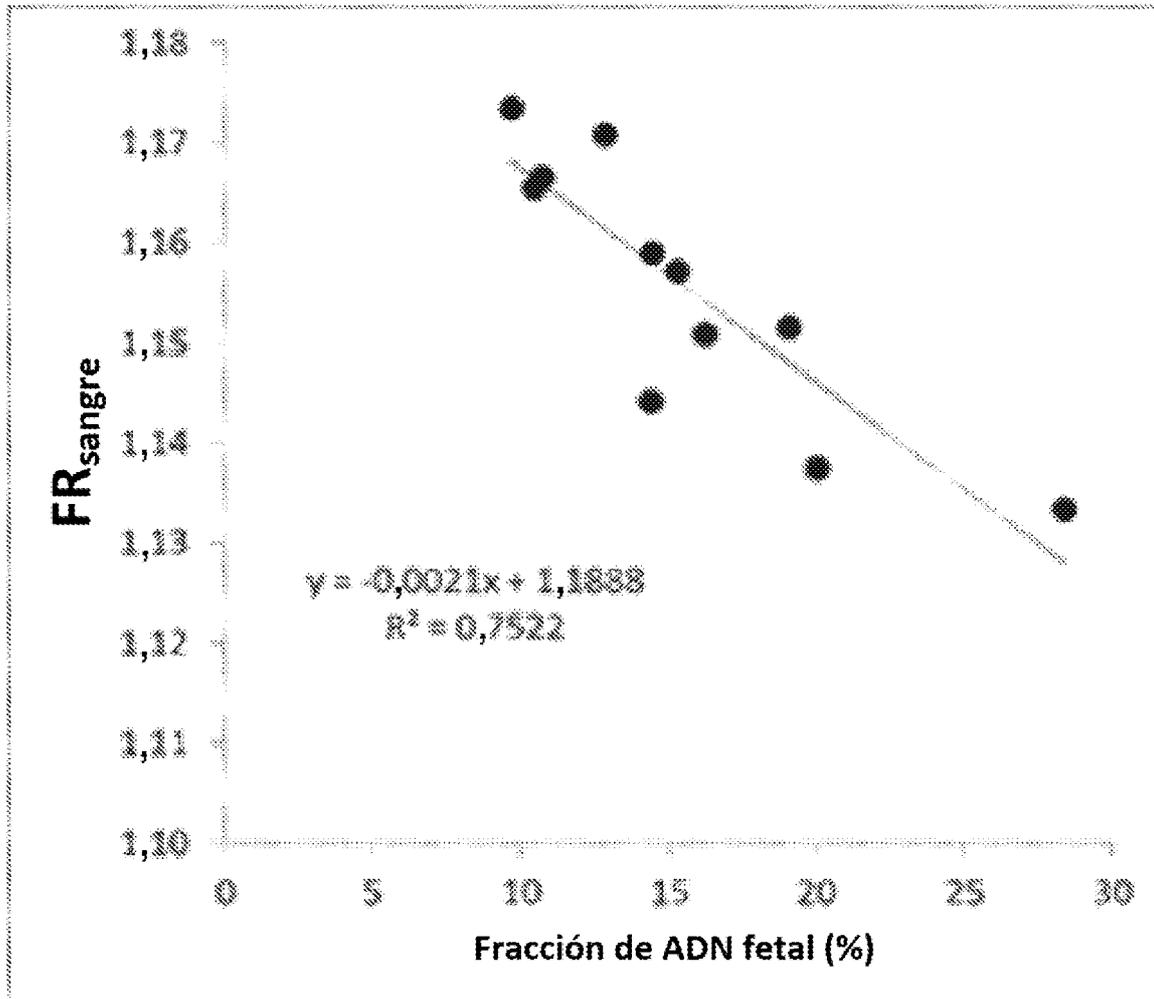


FIG. 12

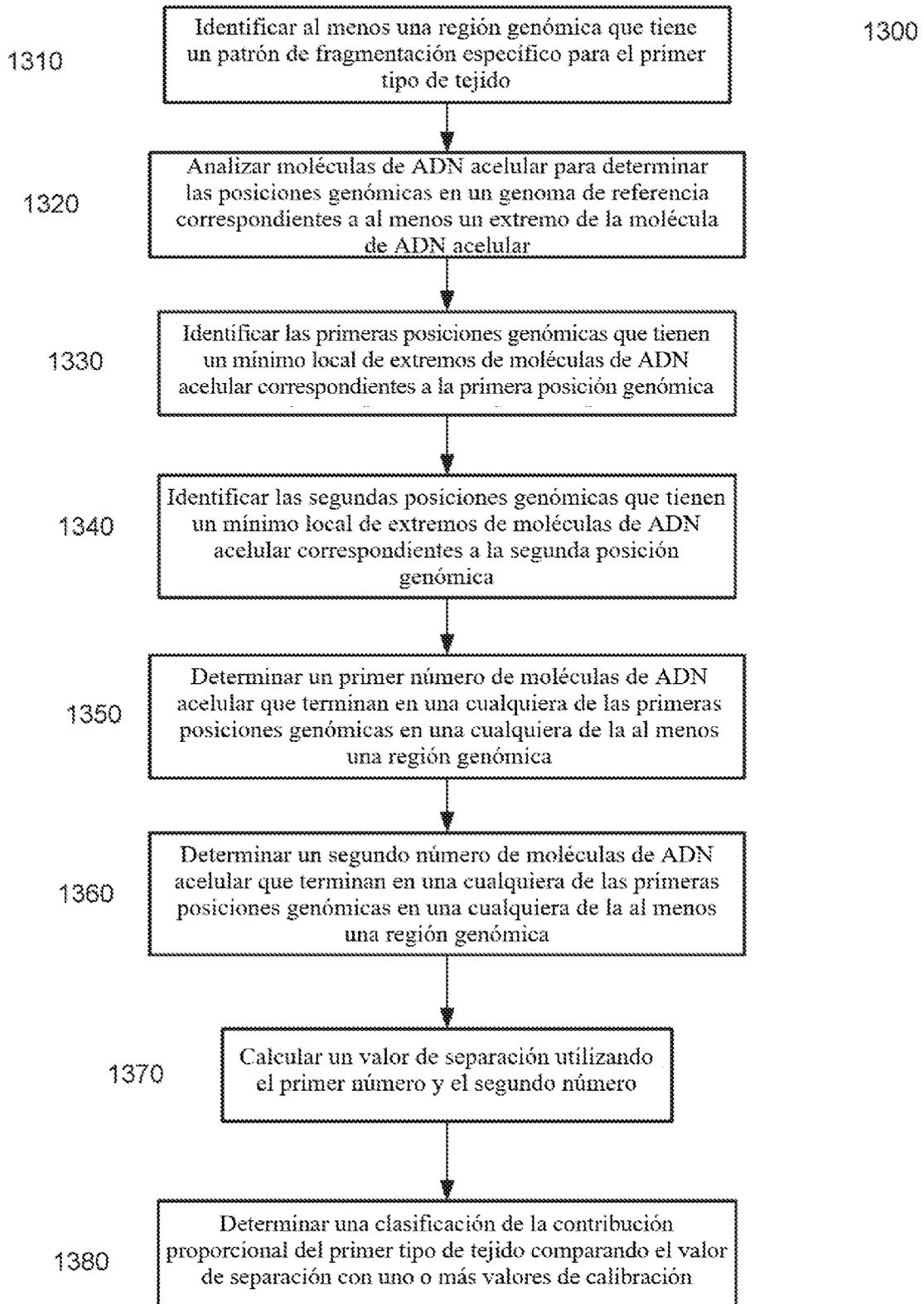


FIG. 13

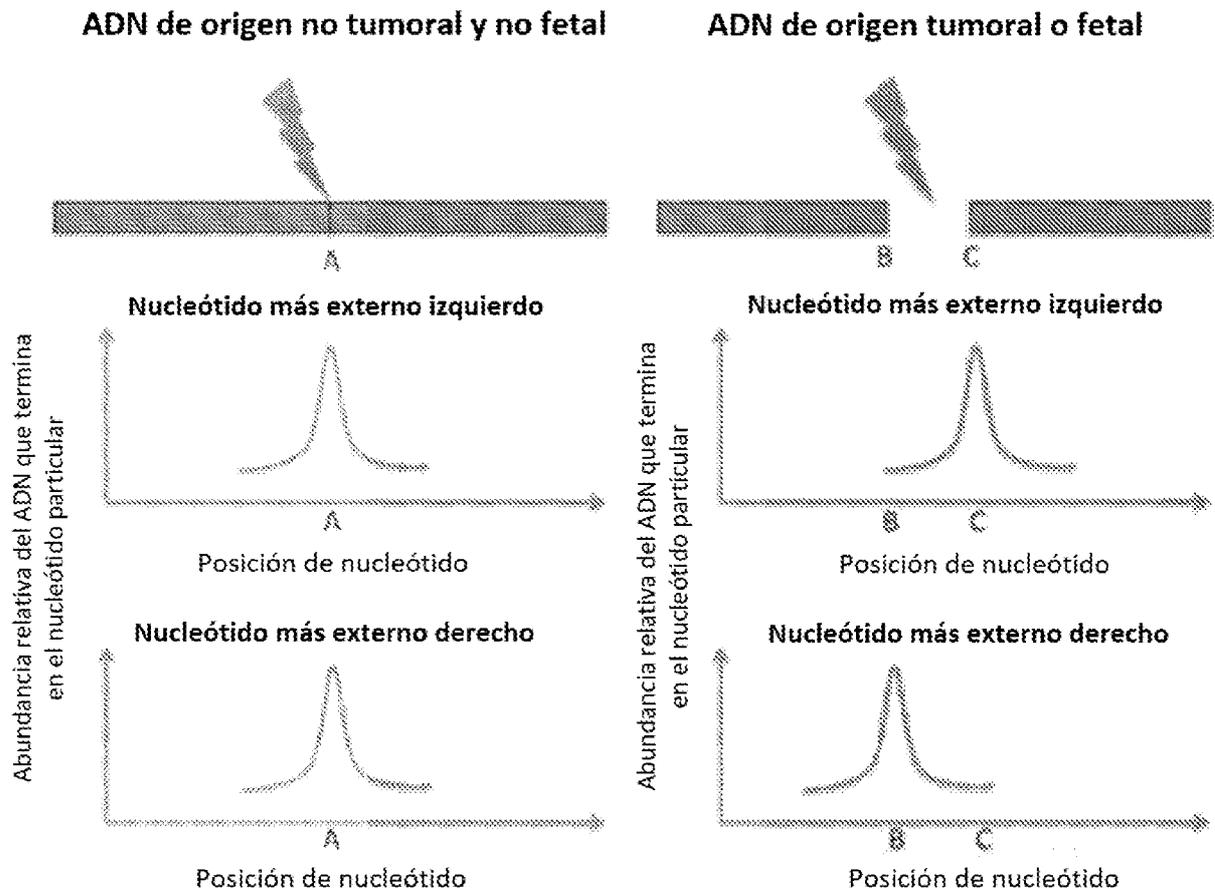


FIG. 14

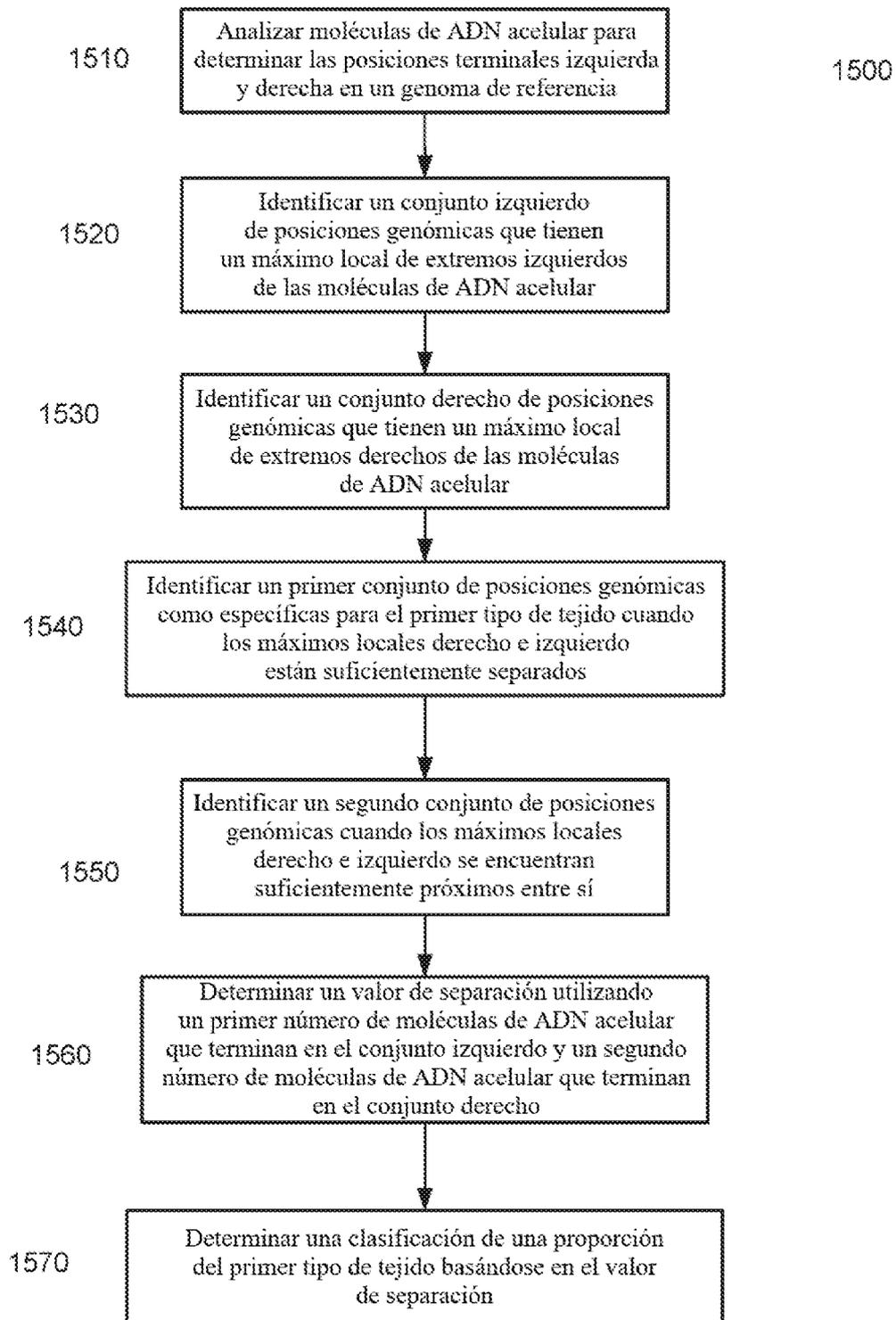


FIG. 15

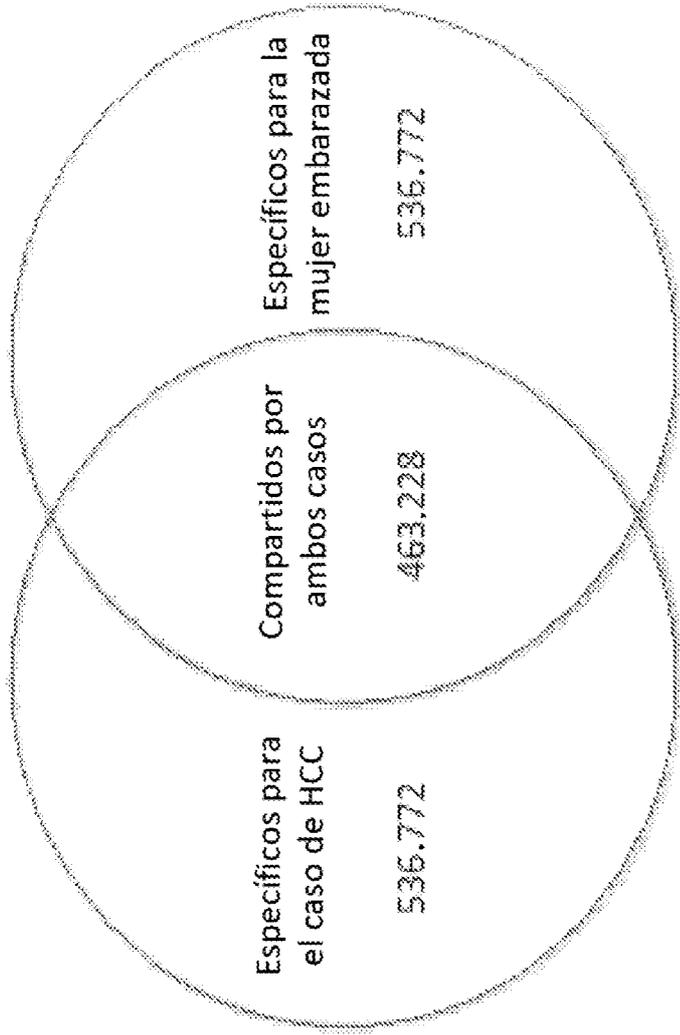


FIG. 16

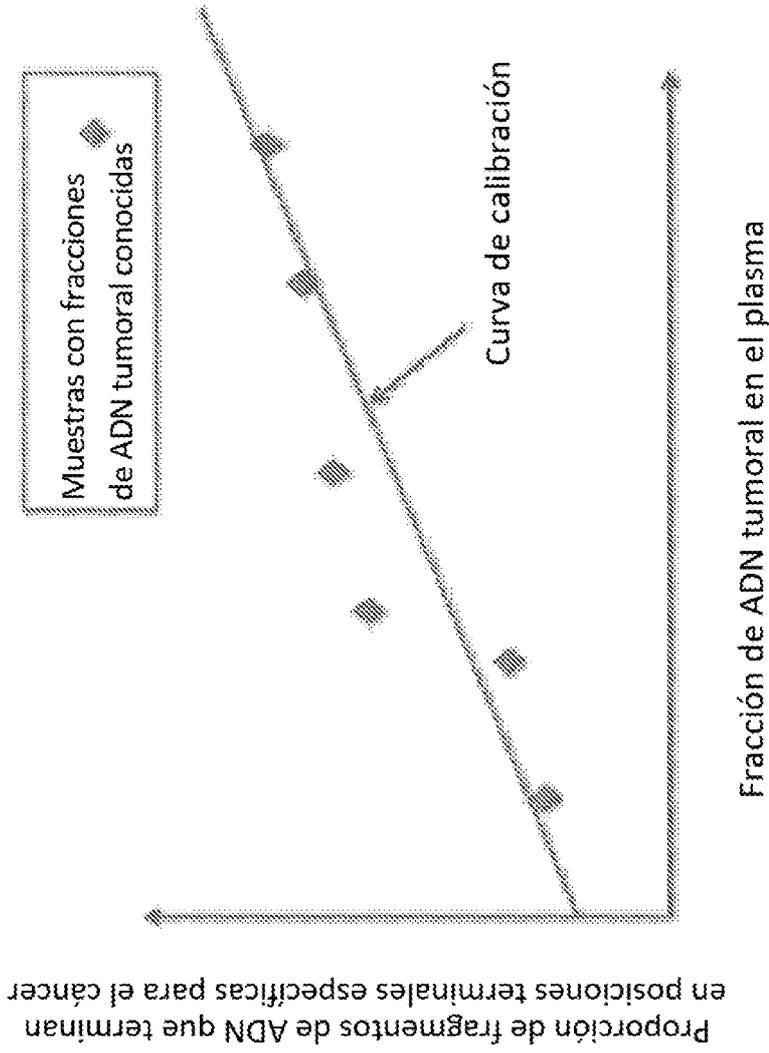
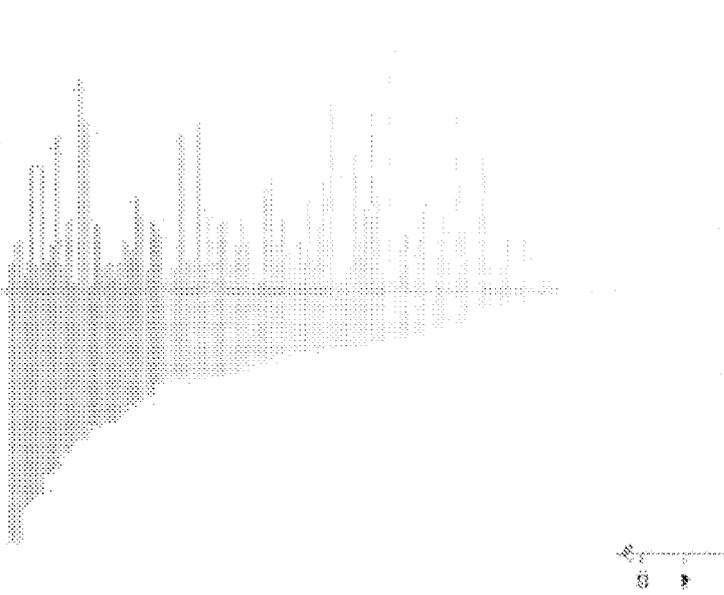


FIG. 17

ADN celular en la sangre materna tratado con ultrasonidos

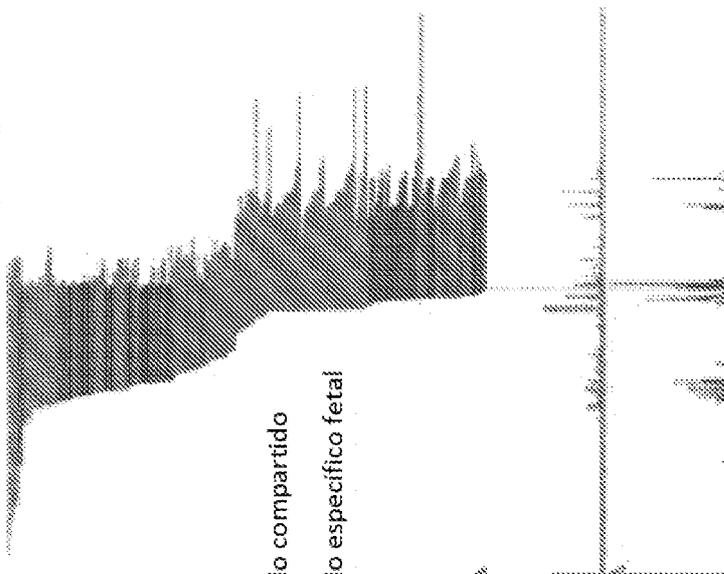
↓ SNP informativo



Coordenadas genómicas relativas

ADN plasmático materno

↓ SNP informativo



Coordenadas genómicas relativas

●●●●● Alelo compartido

●●●●● Alelo específico fetal

Proporción de fragmentos que terminan en la posición de nucleótido (%)

FIG. 18

SNP que informa que la madre era homocigota
y el feto era heterocigoto

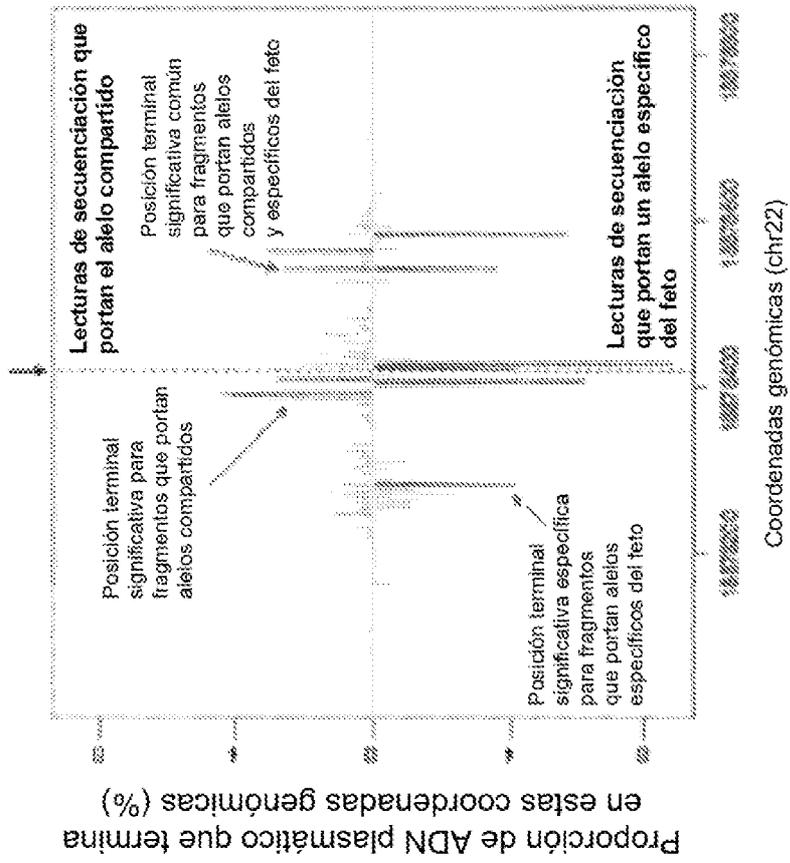


FIG. 19

**Posiciones terminales para fragmentos de ADN plasmático
a lo largo de un SNP que era homocigoto en la madre
y heterocigoto en el feto**

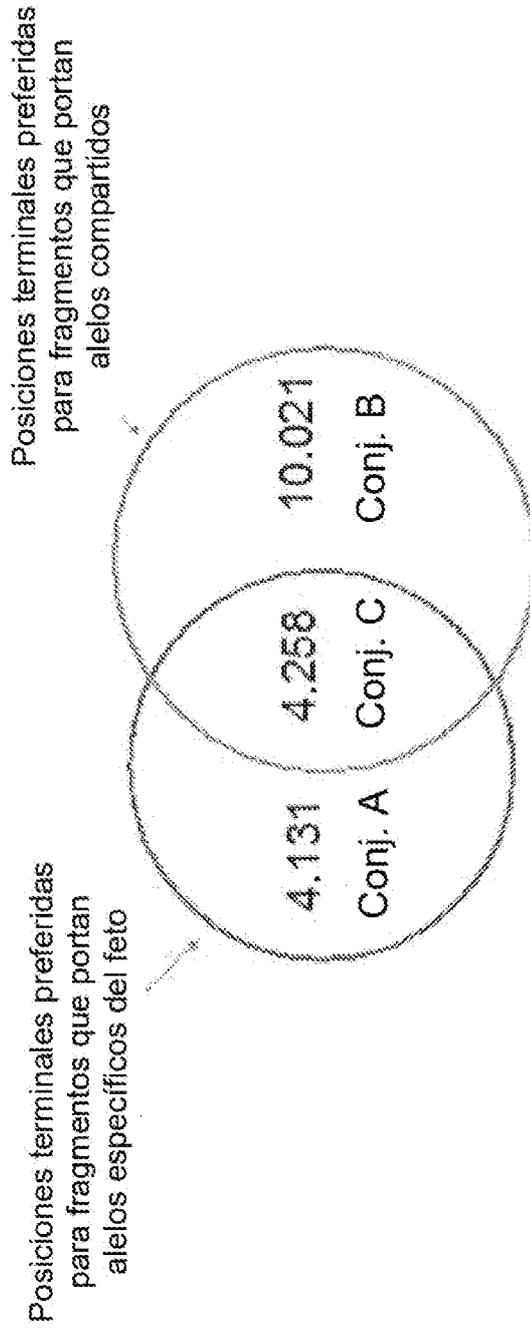


FIG. 20

**Posiciones terminales para fragmentos de ADN plasmático
a lo largo de un SNP que era heterocigoto en la madre
y homocigoto en el feto**

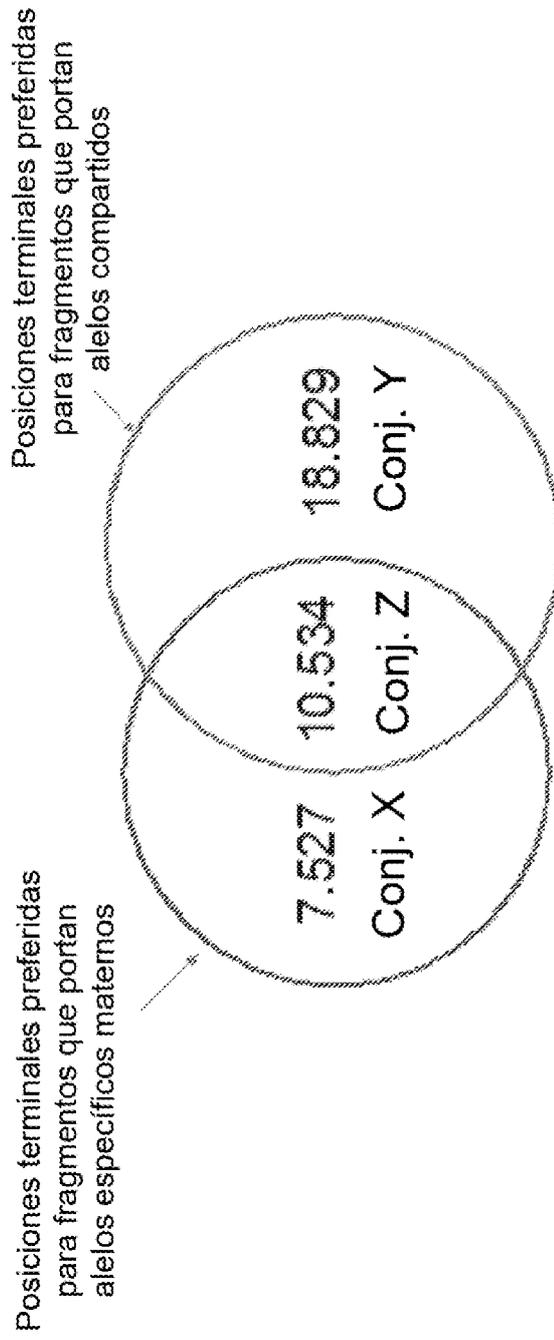


FIG. 21

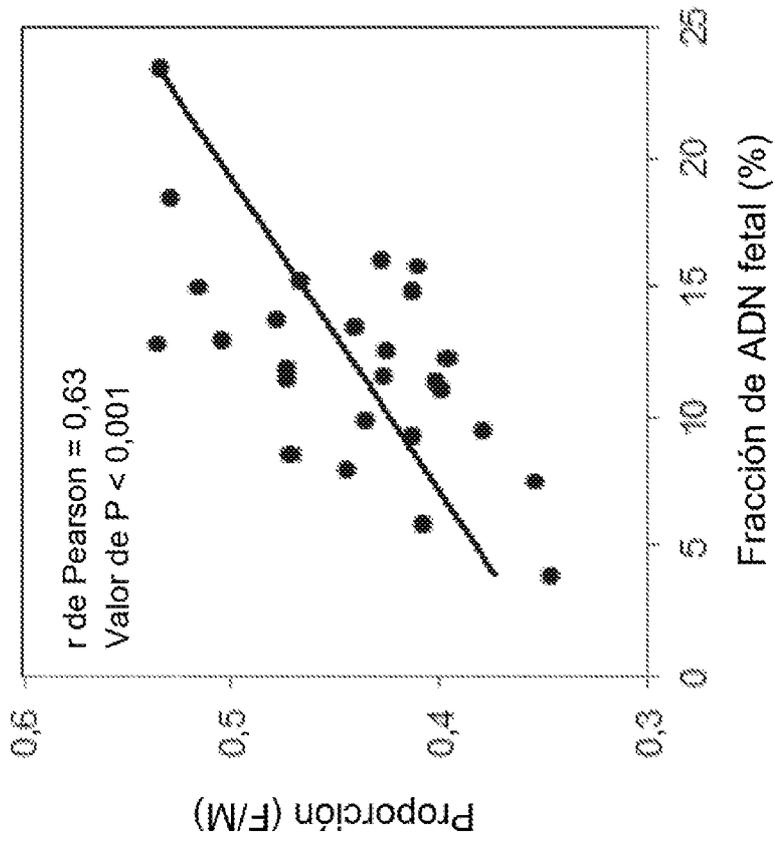


FIG. 22

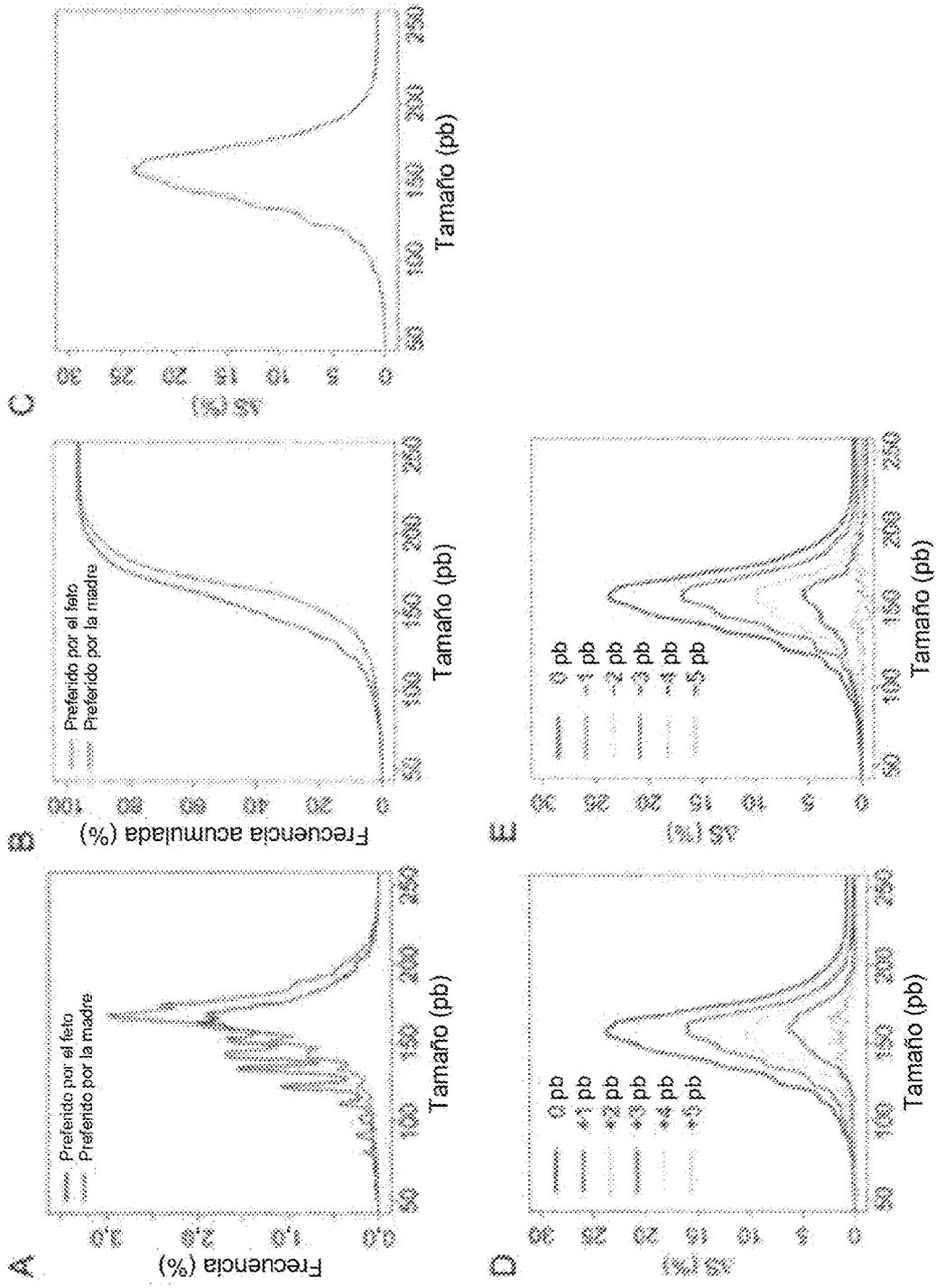


FIG. 23

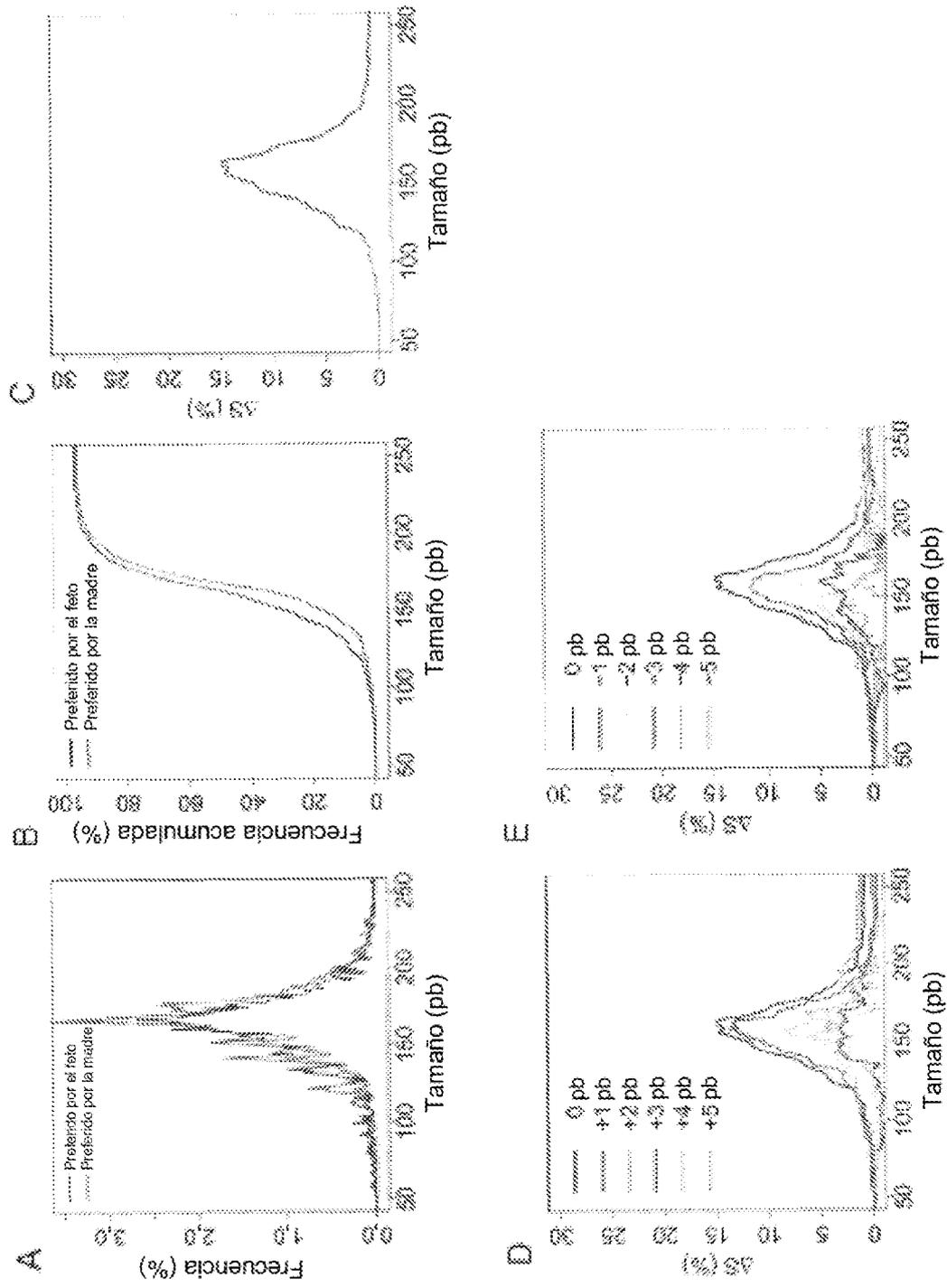


FIG. 24

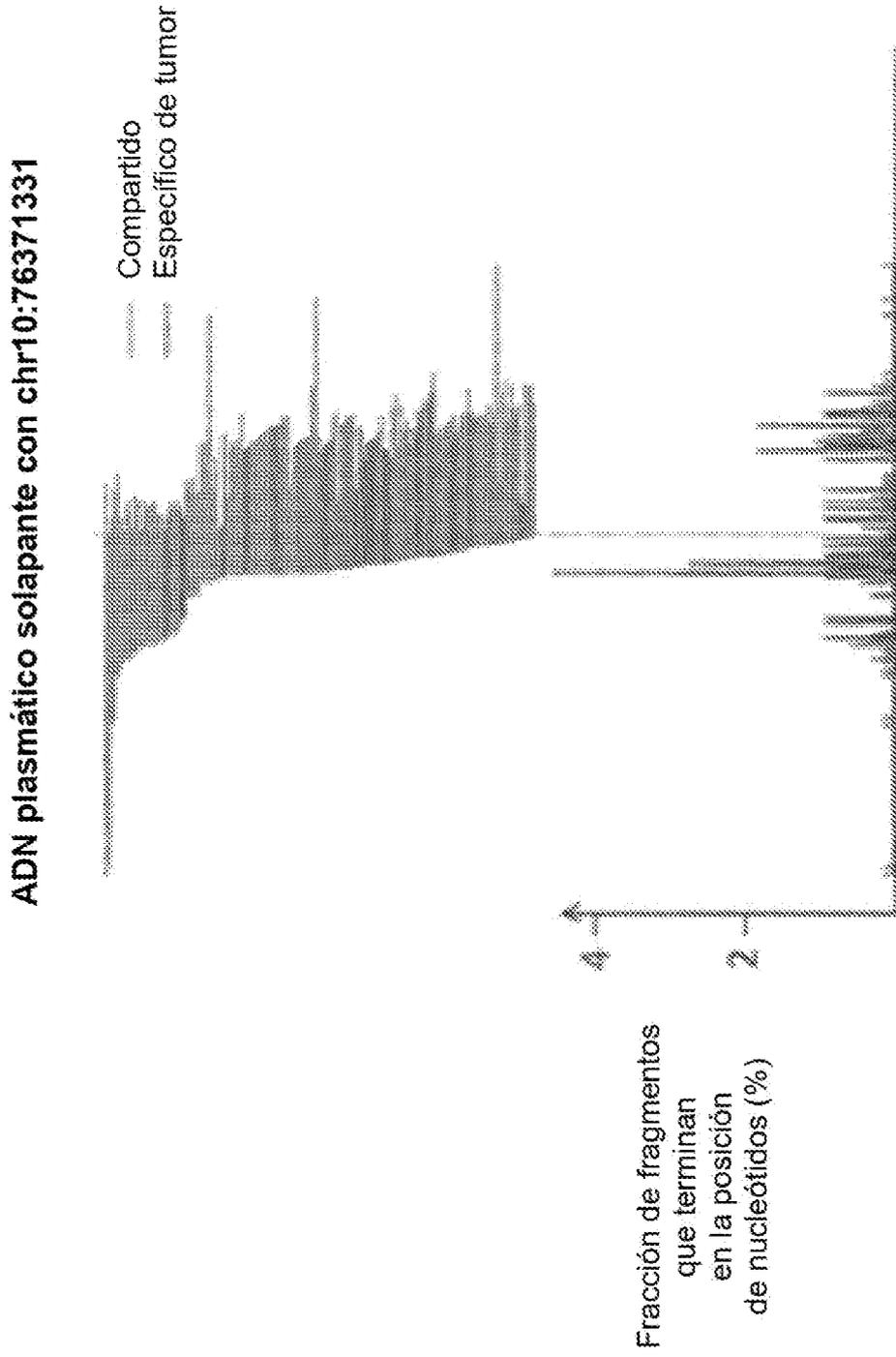
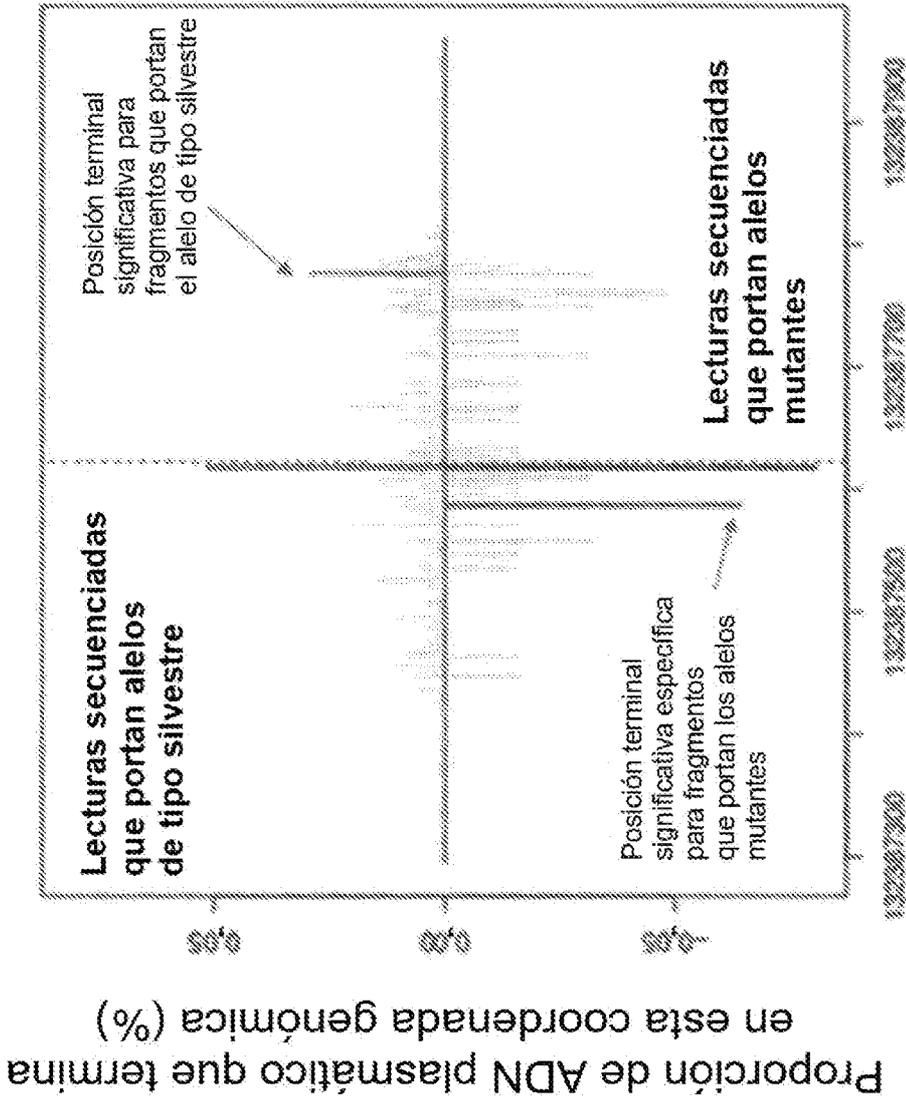


FIG. 25

ADN plasmático solapante con chr10:132387623



Coordenadas genómicas

FIG. 26

**Posiciones terminales de fragmentos de ADN
entre las mutaciones presentes en el plasma**

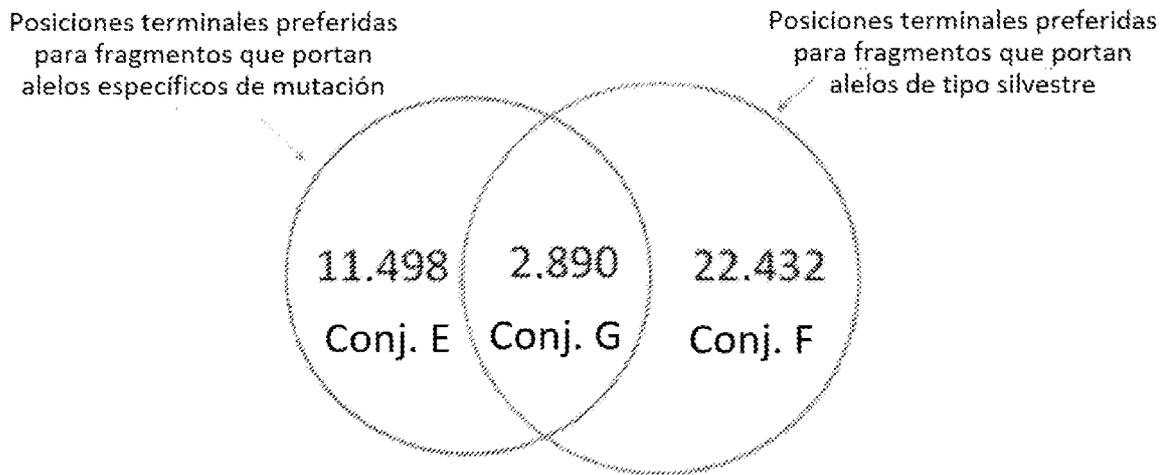


FIG. 27A

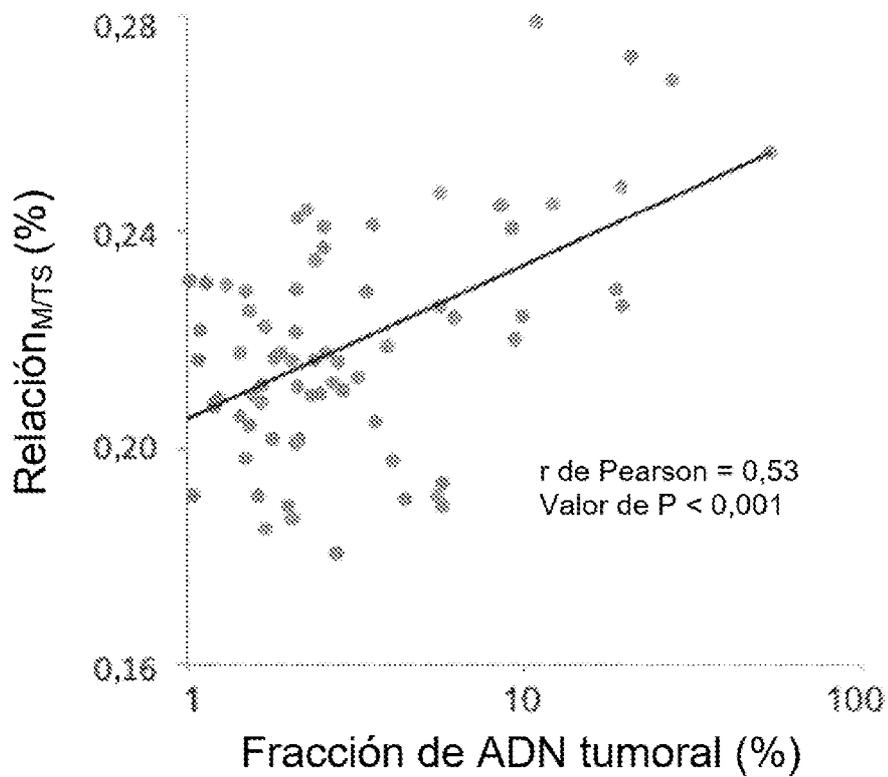


FIG. 27B

Sitios terminales preferidos de fragmentos de ADN plasmático

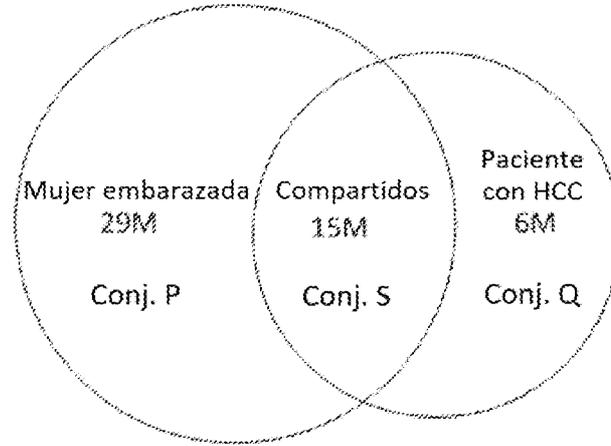


FIG. 28A

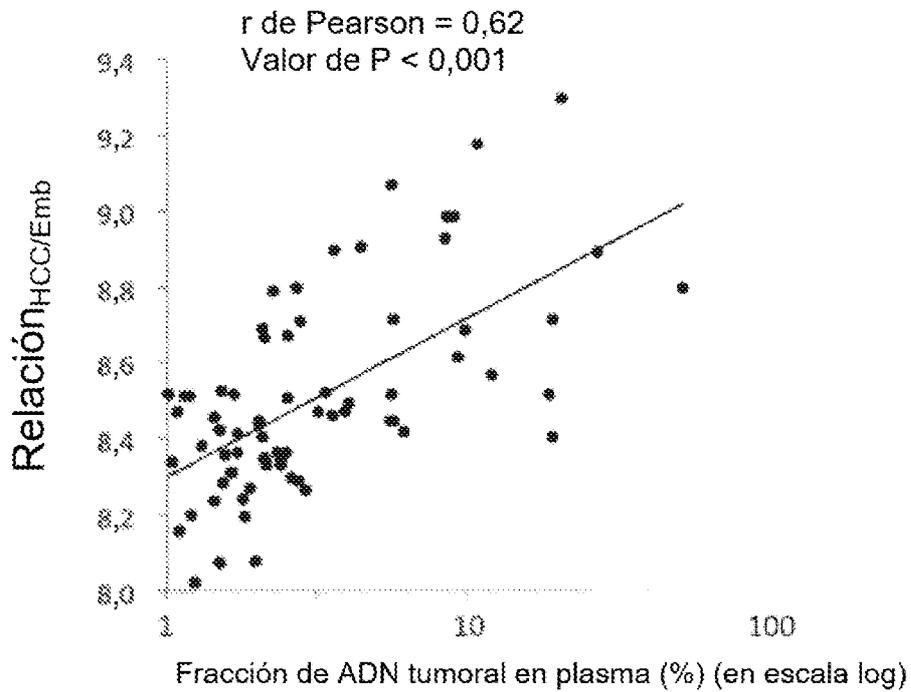


FIG. 28B

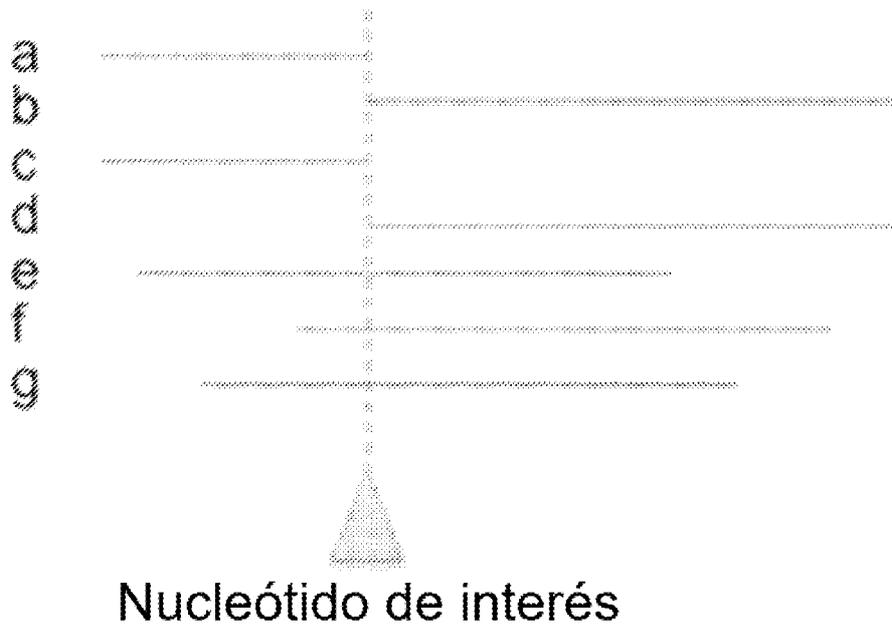


FIG. 29A

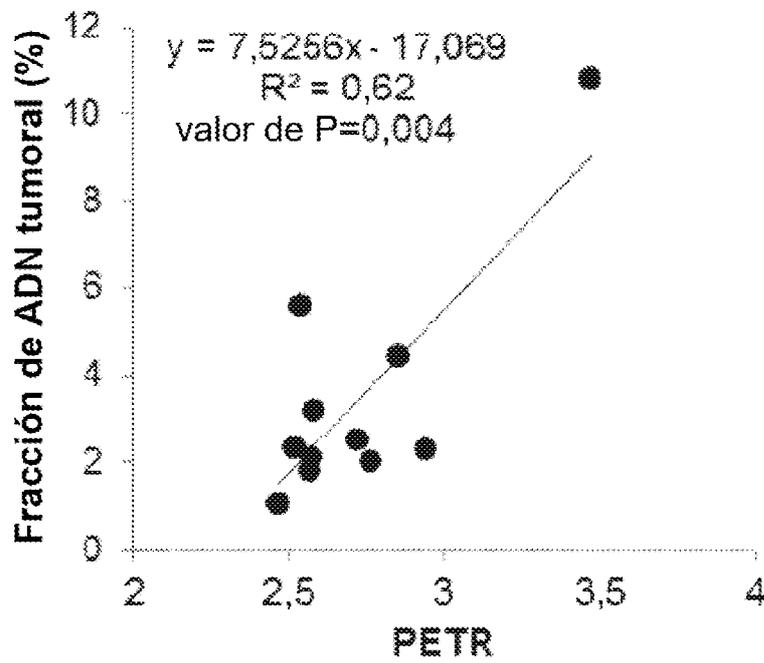


FIG. 29B

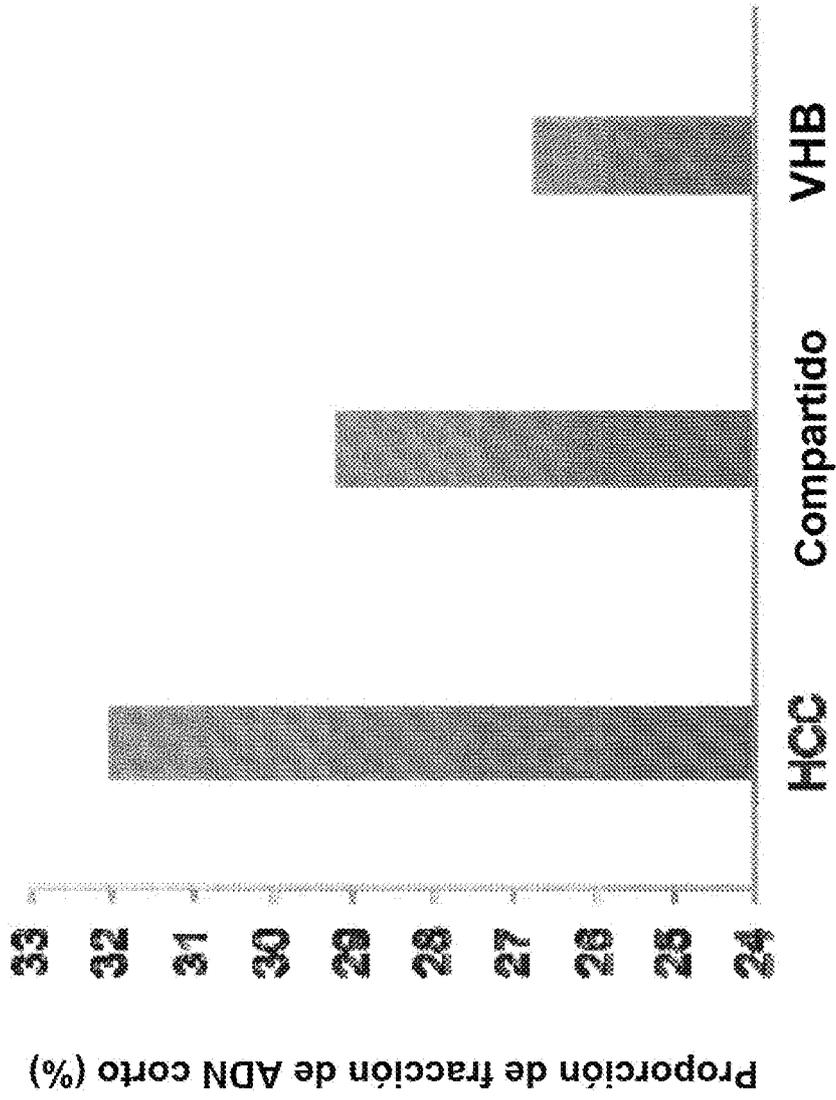


FIG. 30

Posición terminal preferida

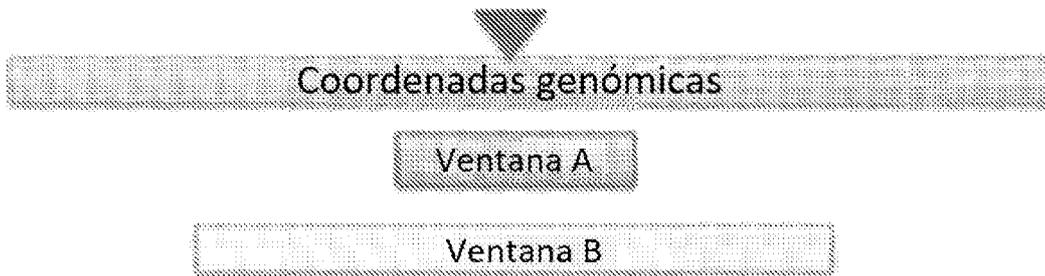


FIG. 31A

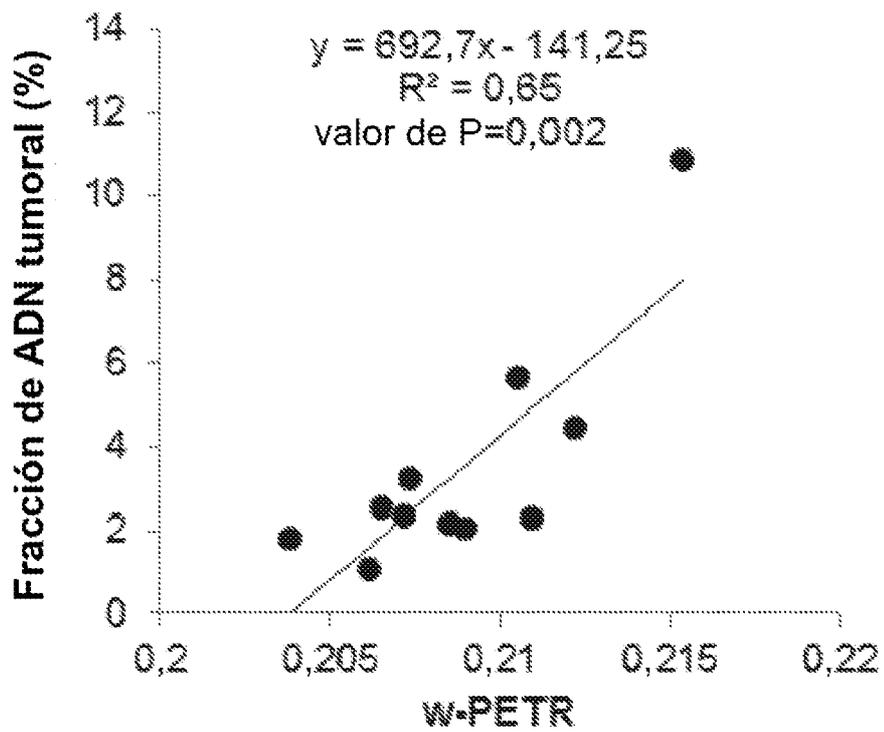


FIG. 31B

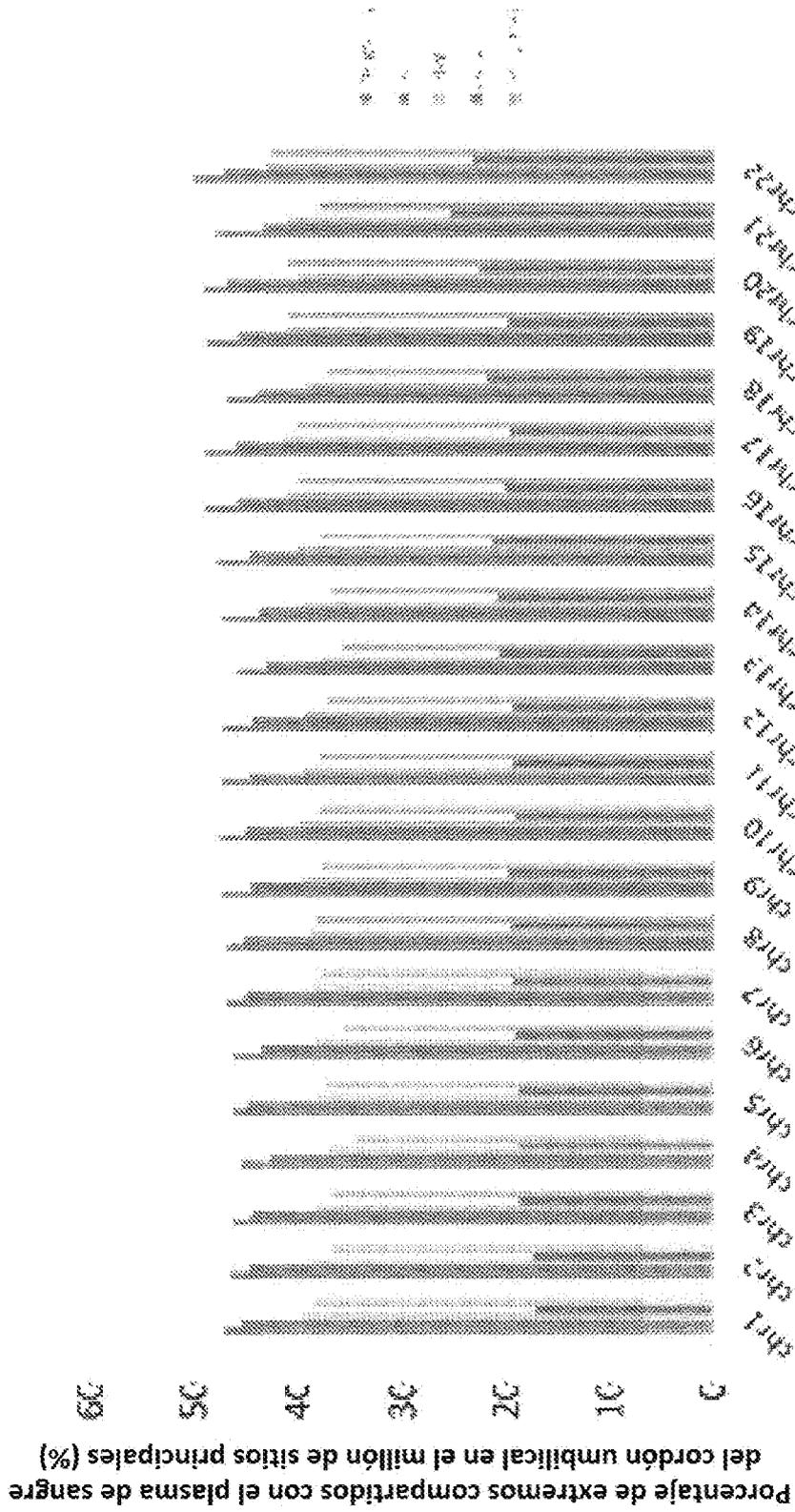


FIG. 32

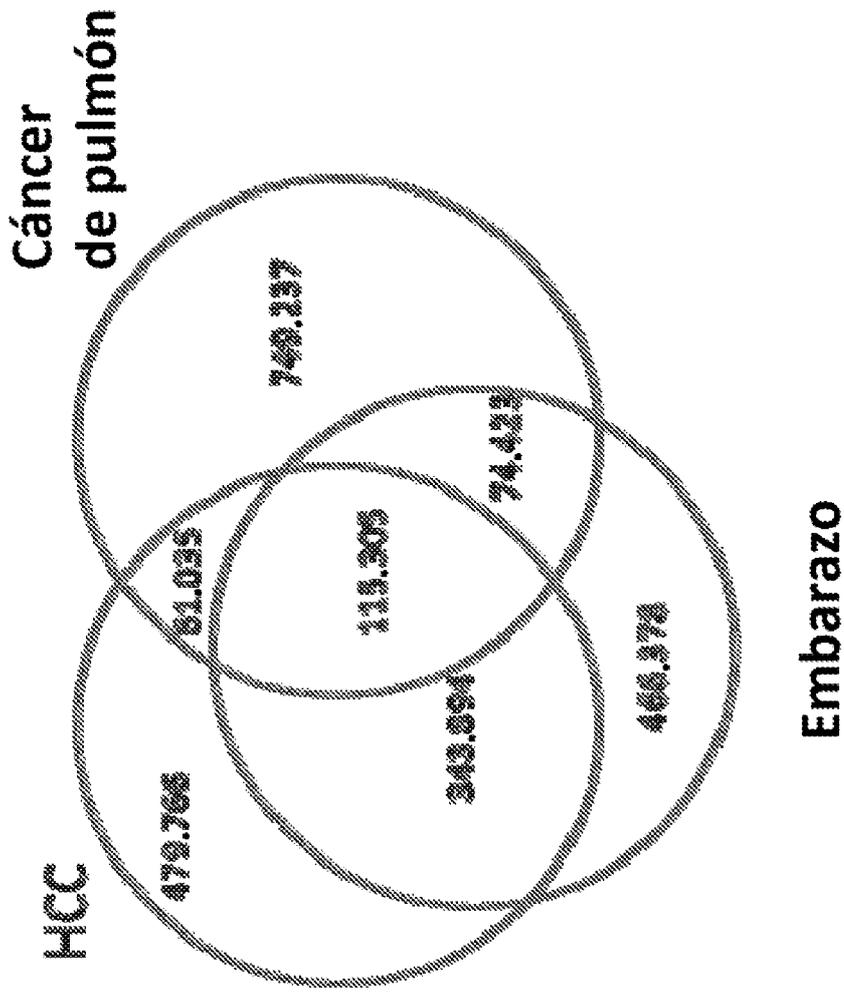


FIG. 33

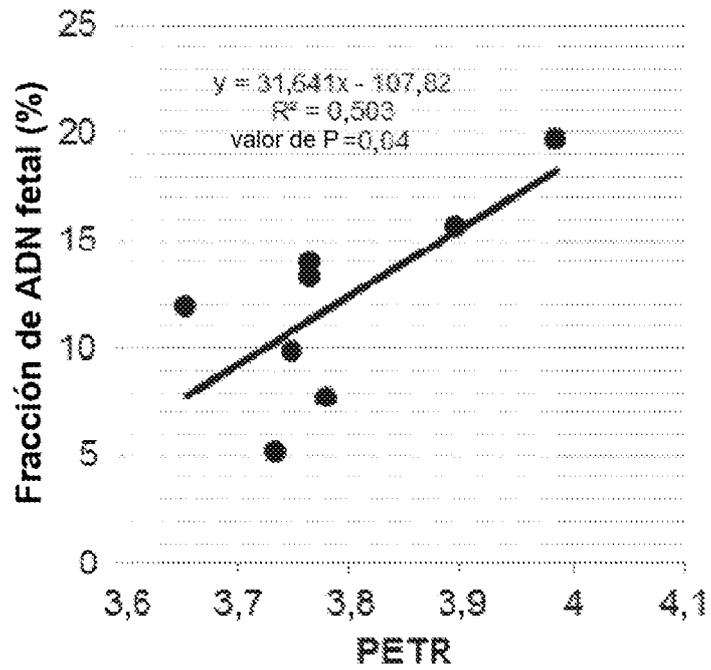


FIG. 34A

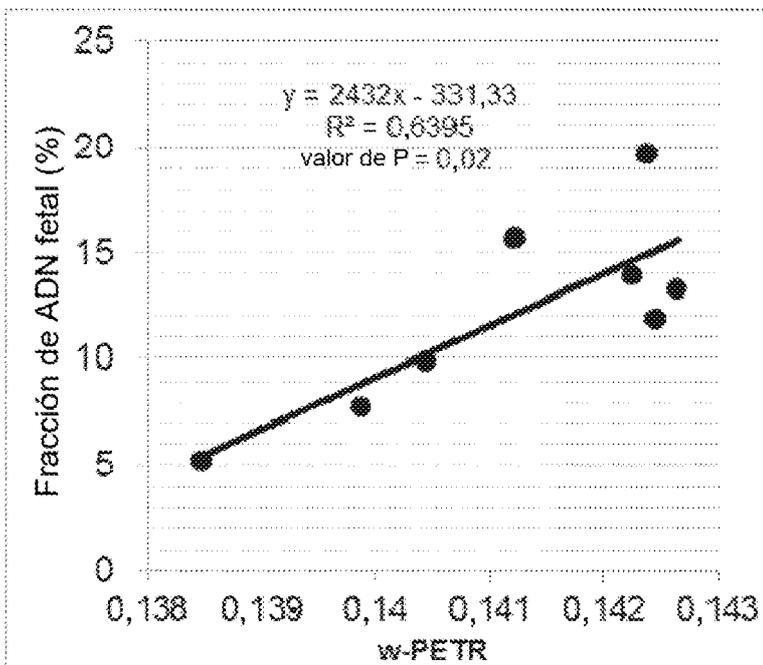


FIG. 34B

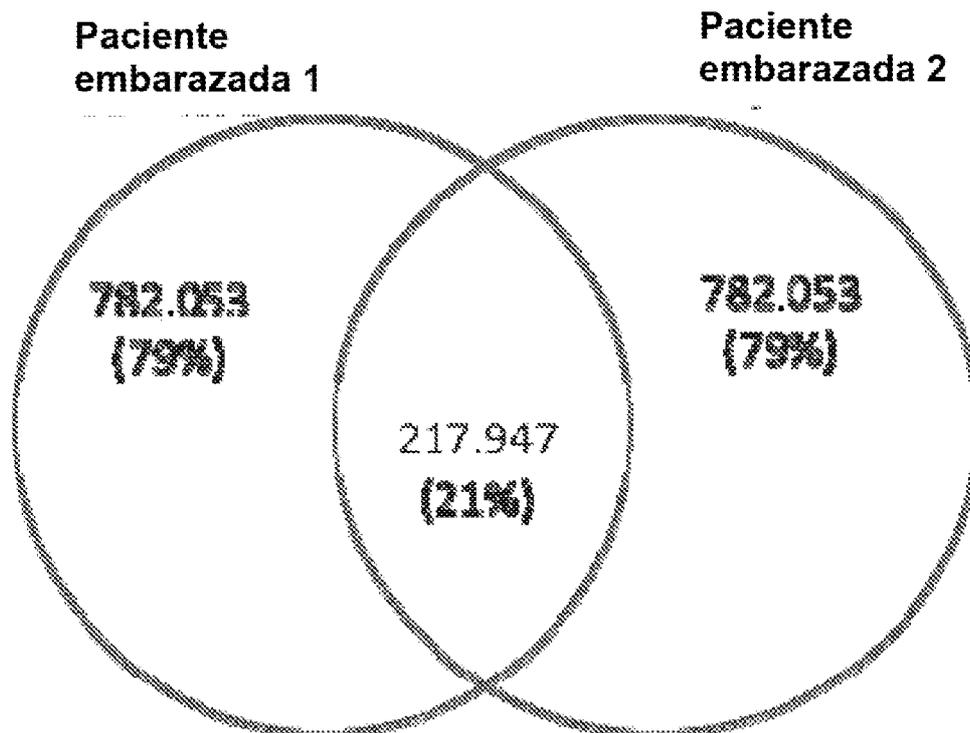


FIG. 35A

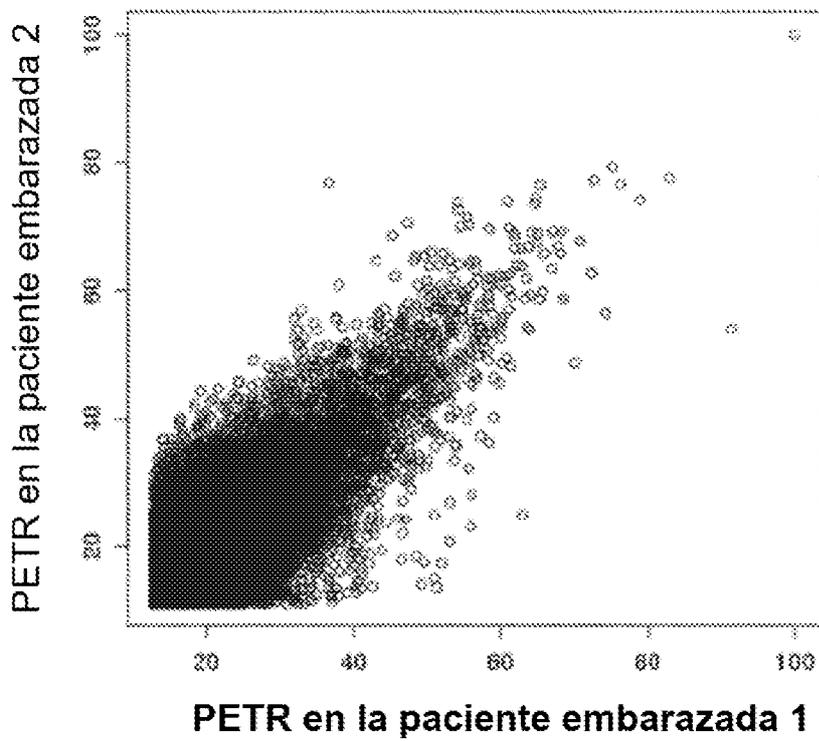


FIG. 35B

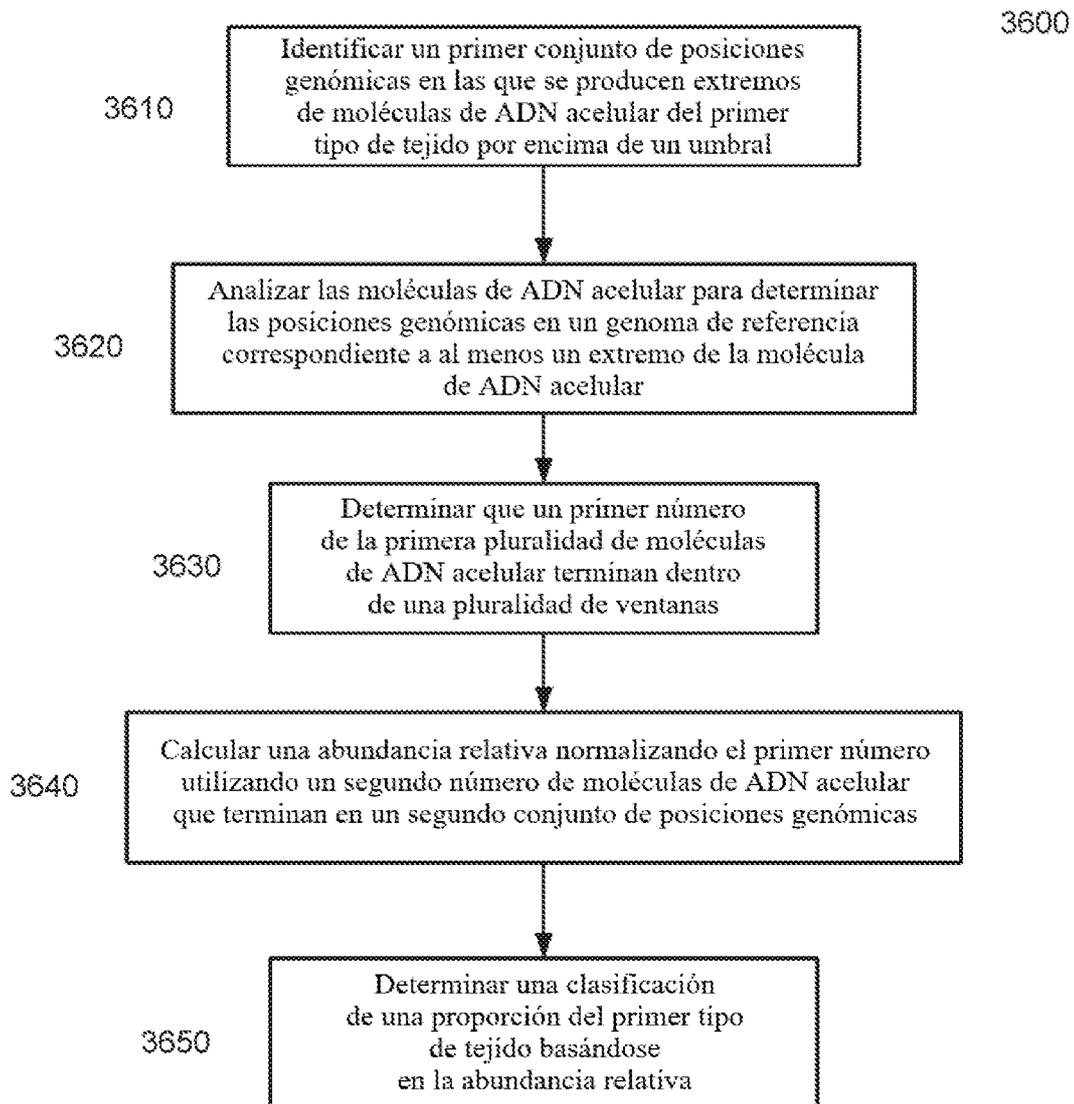


FIG. 36

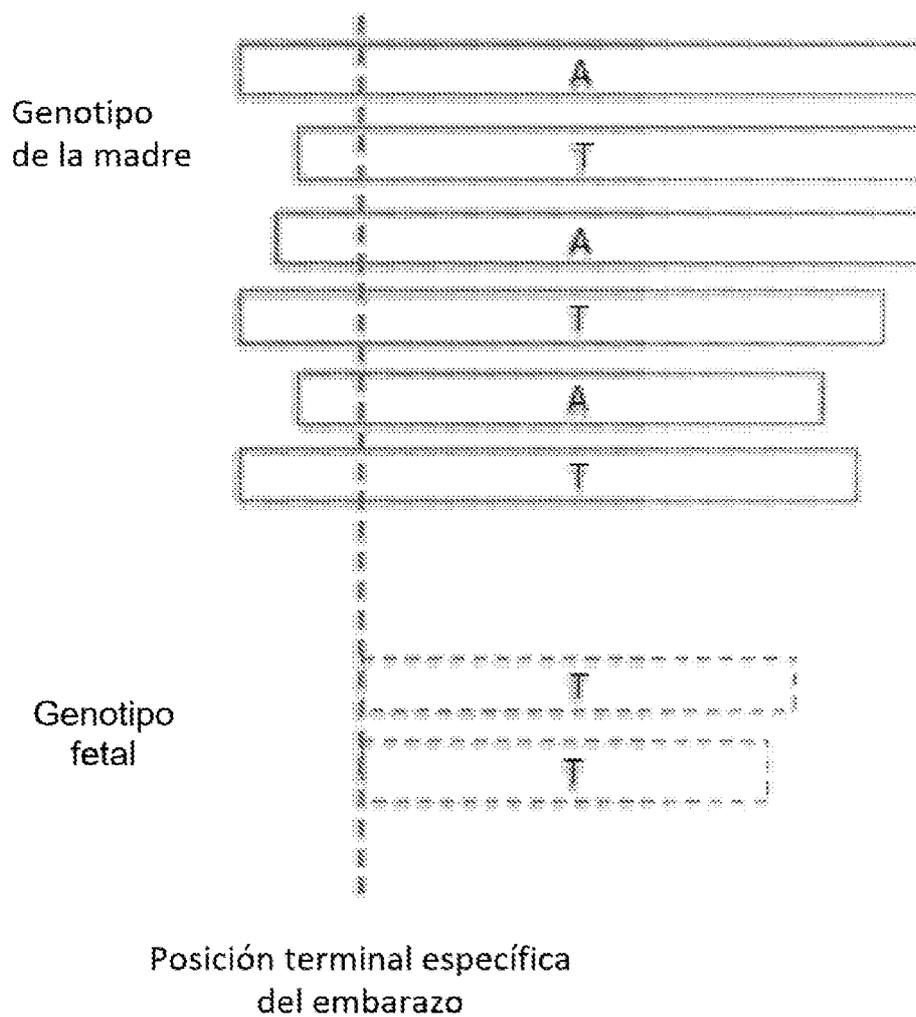


FIG. 37

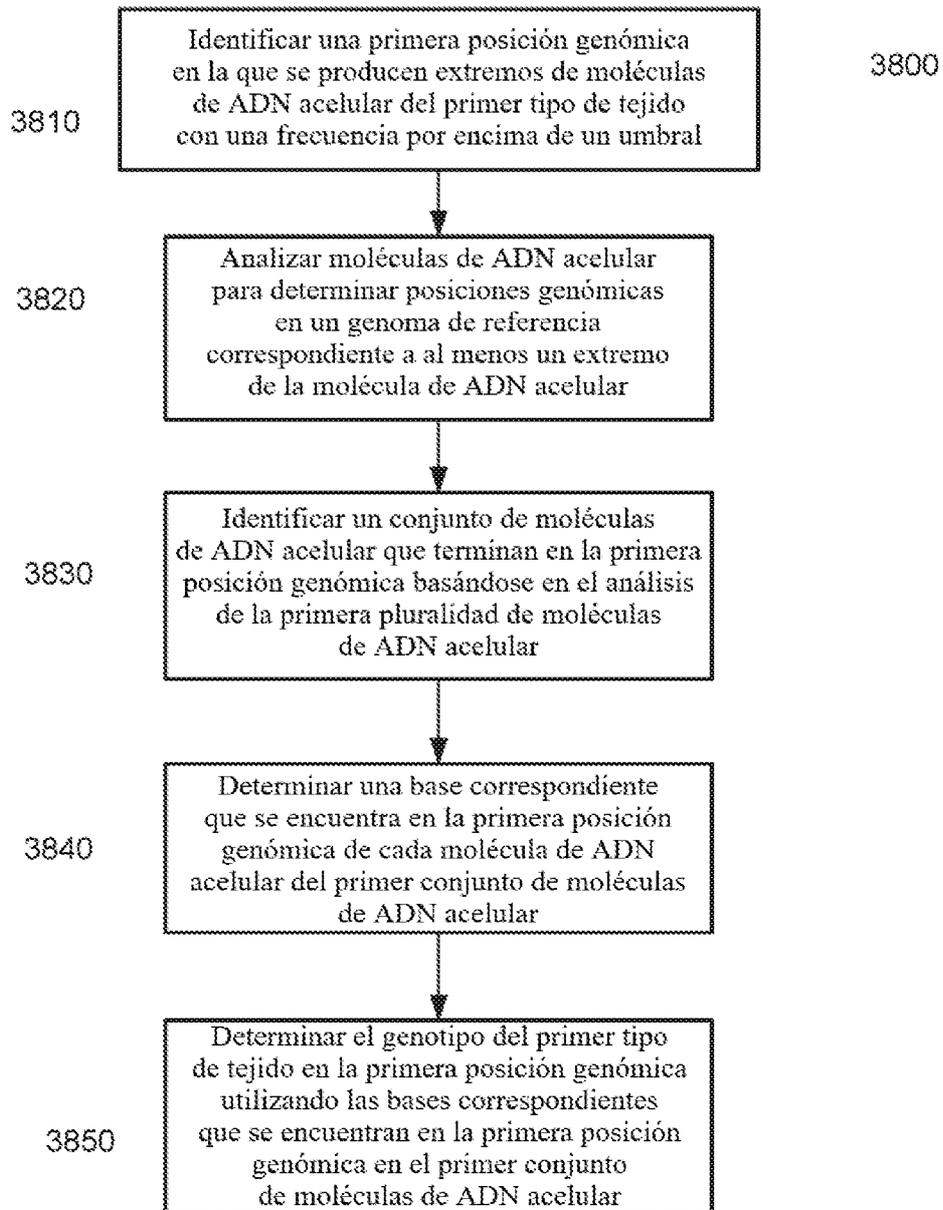


FIG. 38

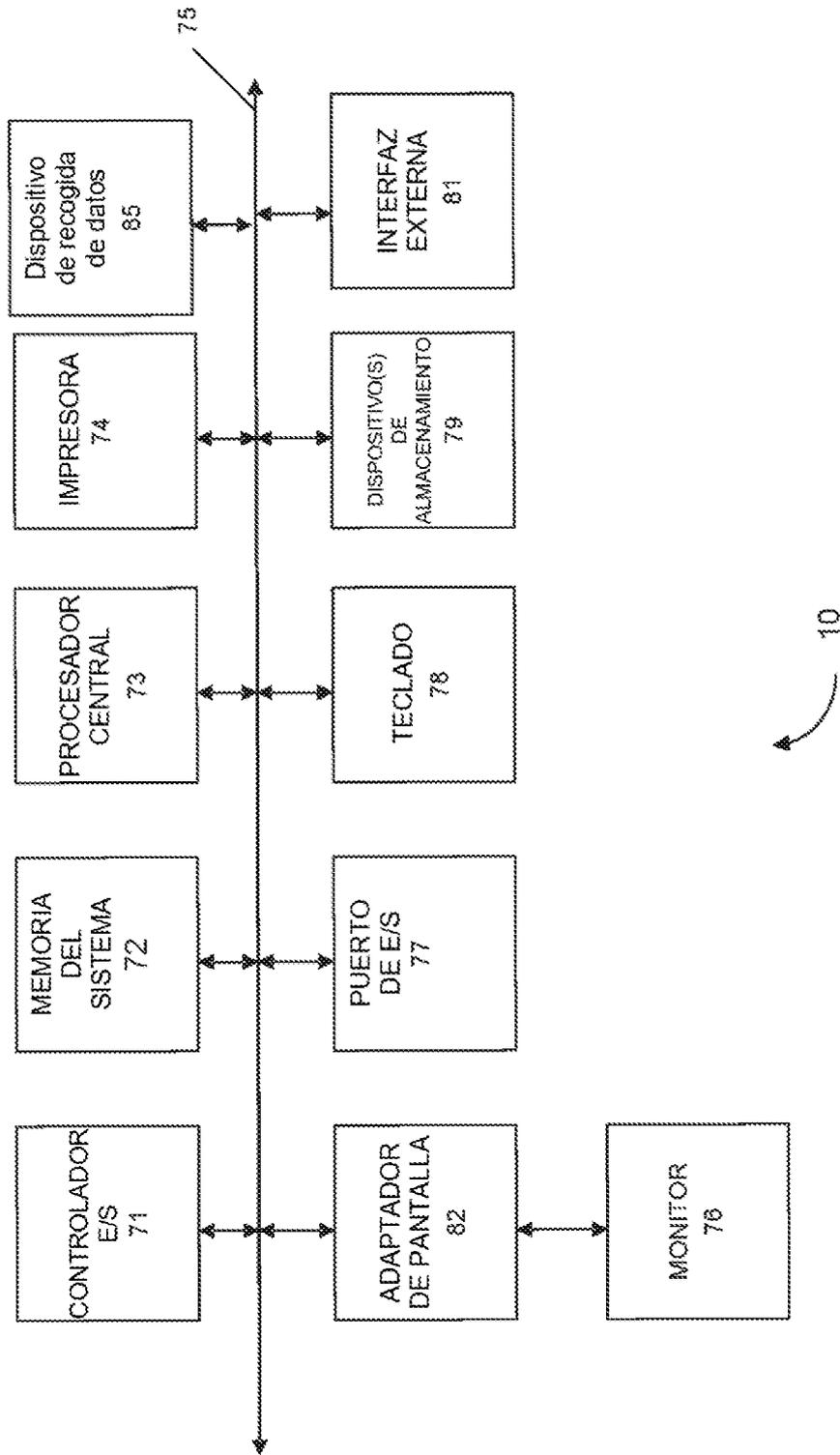


FIG. 39

Table with multiple columns containing alphanumeric strings, likely representing data points or identifiers. The table is extremely dense and contains a large volume of data.

Main table containing multiple columns of alphanumeric data, likely representing a large dataset or ledger. The data is organized into rows and columns, with some entries appearing to be numerical values or alphanumeric codes.

151

Table with multiple columns of numbers, likely a data list or index. The numbers are arranged in rows and columns, with some numbers appearing to be grouped or separated by specific markers.

2273052, 3538349, 3624848, 3724222, 3828359, 3934422, 4042474, 4152478, 4263478, 4375478, 4488478, 4602478, 4716478, 4831478, 4947478, 5063478, 5180478, 5298478, 5416478, 5535478, 5655478, 5775478, 5896478, 6017478, 6138478, 6260478, 6383478, 6506478, 6630478, 6754478, 6879478, 7004478, 7130478, 7256478, 7383478, 7511478, 7640478, 7770478, 7901478, 8033478, 8165478, 8298478, 8432478, 8566478, 8701478, 8836478, 8972478, 9108478, 9245478, 9383478, 9521478, 9660478, 9799478, 9939478, 10079478, 10220478, 10361478, 10503478, 10645478, 10788478, 10931478, 11075478, 11218478, 11362478, 11506478, 11651478, 11796478, 11942478, 12088478, 12234478, 12380478, 12526478, 12673478, 12820478, 12968478, 13116478, 13264478, 13413478, 13562478, 13711478, 13861478, 14011478, 14161478, 14312478, 14463478, 14614478, 14766478, 14918478, 15071478, 15224478, 15377478, 15531478, 15685478, 15840478, 15995478, 16150478, 16306478, 16462478, 16618478, 16775478, 16932478, 17090478, 17248478, 17407478, 17566478, 17726478, 17886478, 18047478, 18208478, 18369478, 18531478, 18693478, 18856478, 19019478, 19182478, 19346478, 19510478, 19674478, 19839478, 20003478, 20168478, 20333478, 20498478, 20664478, 20830478, 20996478, 21162478, 21329478, 21496478, 21664478, 21832478, 22001478, 22170478, 22340478, 22510478, 22680478, 22851478, 23022478, 23193478, 23365478, 23537478, 23710478, 23883478, 24056478, 24230478, 24404478, 24579478, 24754478, 24930478, 25106478, 25282478, 25459478, 25636478, 25814478, 25992478, 26170478, 26348478, 26527478, 26706478, 26885478, 27064478, 27244478, 27424478, 27604478, 27784478, 27964478, 28144478, 28324478, 28504478, 28684478, 28864478, 29044478, 29224478, 29404478, 29584478, 29764478, 29944478, 30124478, 30304478, 30484478, 30664478, 30844478, 31024478, 31204478, 31384478, 31564478, 31744478, 31924478, 32104478, 32284478, 32464478, 32644478, 32824478, 33004478, 33184478, 33364478, 33544478, 33724478, 33904478, 34084478, 34264478, 34444478, 34624478, 34804478, 34984478, 35164478, 35344478, 35524478, 35704478, 35884478, 36064478, 36244478, 36424478, 36604478, 36784478, 36964478, 37144478, 37324478, 37504478, 37684478, 37864478, 38044478, 38224478, 38404478, 38584478, 38764478, 38944478, 39124478, 39304478, 39484478, 39664478, 39844478, 40024478, 40204478, 40384478, 40564478, 40744478, 40924478, 41104478, 41284478, 41464478, 41644478, 41824478, 42004478, 42184478, 42364478, 42544478, 42724478, 42904478, 43084478, 43264478, 43444478, 43624478, 43804478, 43984478, 44164478, 44344478, 44524478, 44704478, 44884478, 45064478, 45244478, 45424478, 45604478, 45784478, 45964478, 46144478, 46324478, 46504478, 46684478, 46864478, 47044478, 47224478, 47404478, 47584478, 47764478, 47944478, 48124478, 48304478, 48484478, 48664478, 48844478, 49024478, 49204478, 49384478, 49564478, 49744478, 49924478, 50104478, 50284478, 50464478, 50644478, 50824478, 51004478, 51184478, 51364478, 51544478, 51724478, 51904478, 52084478, 52264478, 52444478, 52624478, 52804478, 52984478, 53164478, 53344478, 53524478, 53704478, 53884478, 54064478, 54244478, 54424478, 54604478, 54784478, 54964478, 55144478, 55324478, 55504478, 55684478, 55864478, 56044478, 56224478, 56404478, 56584478, 56764478, 56944478, 57124478, 57304478, 57484478, 57664478, 57844478, 58024478, 58204478, 58384478, 58564478, 58744478, 58924478, 59104478, 59284478, 59464478, 59644478, 59824478, 60004478, 60184478, 60364478, 60544478, 60724478, 60904478, 61084478, 61264478, 61444478, 61624478, 61804478, 62004478, 62184478, 62364478, 62544478, 62724478, 62904478, 63084478, 63264478, 63444478, 63624478, 63804478, 63984478, 64164478, 64344478, 64524478, 64704478, 64884478, 65064478, 65244478, 65424478, 65604478, 65784478, 65964478, 66144478, 66324478, 66504478, 66684478, 66864478, 67044478, 67224478, 67404478, 67584478, 67764478, 67944478, 68124478, 68304478, 68484478, 68664478, 68844478, 69024478, 69204478, 69384478, 69564478, 69744478, 69924478, 70104478, 70284478, 70464478, 70644478, 70824478, 71004478, 71184478, 71364478, 71544478, 71724478, 71904478, 72084478, 72264478, 72444478, 72624478, 72804478, 72984478, 73164478, 73344478, 73524478, 73704478, 73884478, 74064478, 74244478, 74424478, 74604478, 74784478, 74964478, 75144478, 75324478, 75504478, 75684478, 75864478, 76044478, 76224478, 76404478, 76584478, 76764478, 76944478, 77124478, 77304478, 77484478, 77664478, 77844478, 78024478, 78204478, 78384478, 78564478, 78744478, 78924478, 79104478, 79284478, 79464478, 79644478, 79824478, 80004478, 80184478, 80364478, 80544478, 80724478, 80904478, 81084478, 81264478, 81444478, 81624478, 81804478, 81984478, 82164478, 82344478, 82524478, 82704478, 82884478, 83064478, 83244478, 83424478, 83604478, 83784478, 83964478, 84144478, 84324478, 84504478, 84684478, 84864478, 85044478, 85224478, 85404478, 85584478, 85764478, 85944478, 86124478, 86304478, 86484478, 86664478, 86844478, 87024478, 87204478, 87384478, 87564478, 87744478, 87924478, 88104478, 88284478, 88464478, 88644478, 88824478, 89004478, 89184478, 89364478, 89544478, 89724478, 89904478, 90084478, 90264478, 90444478, 90624478, 90804478, 90984478, 91164478, 91344478, 91524478, 91704478, 91884478, 92064478, 92244478, 92424478, 92604478, 92784478, 92964478, 93144478, 93324478, 93504478, 93684478, 93864478, 94044478, 94224478, 94404478, 94584478, 94764478, 94944478, 95124478, 95304478, 95484478, 95664478, 95844478, 96024478, 96204478, 96384478, 96564478, 96744478, 96924478, 97104478, 97284478, 97464478, 97644478, 97824478, 98004478, 98184478, 98364478, 98544478, 98724478, 98904478, 99084478, 99264478, 99444478, 99624478, 99804478, 1000000000

3978555	3978556	3978557	3978558	3978559	3978560	3978561	3978562	3978563	3978564	3978565	3978566	3978567	3978568	3978569	3978570	3978571	3978572	3978573	3978574	3978575	3978576	3978577	3978578	3978579	3978580	3978581	3978582	3978583	3978584	3978585	3978586	3978587	3978588	3978589	3978590	3978591	3978592	3978593	3978594	3978595	3978596	3978597	3978598	3978599	3978600	3978601	3978602	3978603	3978604	3978605	3978606	3978607	3978608	3978609	3978610	3978611	3978612	3978613	3978614	3978615	3978616	3978617	3978618	3978619	3978620	3978621	3978622	3978623	3978624	3978625	3978626	3978627	3978628	3978629	3978630	3978631	3978632	3978633	3978634	3978635	3978636	3978637	3978638	3978639	3978640	3978641	3978642	3978643	3978644	3978645	3978646	3978647	3978648	3978649	3978650	3978651	3978652	3978653	3978654	3978655	3978656	3978657	3978658	3978659	3978660	3978661	3978662	3978663	3978664	3978665	3978666	3978667	3978668	3978669	3978670	3978671	3978672	3978673	3978674	3978675	3978676	3978677	3978678	3978679	3978680	3978681	3978682	3978683	3978684	3978685	3978686	3978687	3978688	3978689	3978690	3978691	3978692	3978693	3978694	3978695	3978696	3978697	3978698	3978699	3978700	3978701	3978702	3978703	3978704	3978705	3978706	3978707	3978708	3978709	3978710	3978711	3978712	3978713	3978714	3978715	3978716	3978717	3978718	3978719	3978720	3978721	3978722	3978723	3978724	3978725	3978726	3978727	3978728	3978729	3978730	3978731	3978732	3978733	3978734	3978735	3978736	3978737	3978738	3978739	3978740	3978741	3978742	3978743	3978744	3978745	3978746	3978747	3978748	3978749	3978750	3978751	3978752	3978753	3978754	3978755	3978756	3978757	3978758	3978759	3978760	3978761	3978762	3978763	3978764	3978765	3978766	3978767	3978768	3978769	3978770	3978771	3978772	3978773	3978774	3978775	3978776	3978777	3978778	3978779	3978780	3978781	3978782	3978783	3978784	3978785	3978786	3978787	3978788	3978789	3978790	3978791	3978792	3978793	3978794	3978795	3978796	3978797	3978798	3978799	3978800
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

13915595, 13915604, 13915613, 13915622, 13915631, 13915640, 13915649, 13915658, 13915667, 13915676, 13915685, 13915694, 13915703, 13915712, 13915721, 13915730, 13915739, 13915748, 13915757, 13915766, 13915775, 13915784, 13915793, 13915802, 13915811, 13915820, 13915829, 13915838, 13915847, 13915856, 13915865, 13915874, 13915883, 13915892, 13915901, 13915910, 13915919, 13915928, 13915937, 13915946, 13915955, 13915964, 13915973, 13915982, 13915991, 13916000, 13916009, 13916018, 13916027, 13916036, 13916045, 13916054, 13916063, 13916072, 13916081, 13916090, 13916099, 13916108, 13916117, 13916126, 13916135, 13916144, 13916153, 13916162, 13916171, 13916180, 13916189, 13916198, 13916207, 13916216, 13916225, 13916234, 13916243, 13916252, 13916261, 13916270, 13916279, 13916288, 13916297, 13916306, 13916315, 13916324, 13916333, 13916342, 13916351, 13916360, 13916369, 13916378, 13916387, 13916396, 13916405, 13916414, 13916423, 13916432, 13916441, 13916450, 13916459, 13916468, 13916477, 13916486, 13916495, 13916504, 13916513, 13916522, 13916531, 13916540, 13916549, 13916558, 13916567, 13916576, 13916585, 13916594, 13916603, 13916612, 13916621, 13916630, 13916639, 13916648, 13916657, 13916666, 13916675, 13916684, 13916693, 13916702, 13916711, 13916720, 13916729, 13916738, 13916747, 13916756, 13916765, 13916774, 13916783, 13916792, 13916801, 13916810, 13916819, 13916828, 13916837, 13916846, 13916855, 13916864, 13916873, 13916882, 13916891, 13916900, 13916909, 13916918, 13916927, 13916936, 13916945, 13916954, 13916963, 13916972, 13916981, 13916990, 13916999, 13917008, 13917017, 13917026, 13917035, 13917044, 13917053, 13917062, 13917071, 13917080, 13917089, 13917098, 13917107, 13917116, 13917125, 13917134, 13917143, 13917152, 13917161, 13917170, 13917179, 13917188, 13917197, 13917206, 13917215, 13917224, 13917233, 13917242, 13917251, 13917260, 13917269, 13917278, 13917287, 13917296, 13917305, 13917314, 13917323, 13917332, 13917341, 13917350, 13917359, 13917368, 13917377, 13917386, 13917395, 13917404, 13917413, 13917422, 13917431, 13917440, 13917449, 13917458, 13917467, 13917476, 13917485, 13917494, 13917503, 13917512, 13917521, 13917530, 13917539, 13917548, 13917557, 13917566, 13917575, 13917584, 13917593, 13917602, 13917611, 13917620, 13917629, 13917638, 13917647, 13917656, 13917665, 13917674, 13917683, 13917692, 13917701, 13917710, 13917719, 13917728, 13917737, 13917746, 13917755, 13917764, 13917773, 13917782, 13917791, 13917800, 13917809, 13917818, 13917827, 13917836, 13917845, 13917854, 13917863, 13917872, 13917881, 13917890, 13917899, 13917908, 13917917, 13917926, 13917935, 13917944, 13917953, 13917962, 13917971, 13917980, 13917989, 13917998, 13918007, 13918016, 13918025, 13918034, 13918043, 13918052, 13918061, 13918070, 13918079, 13918088, 13918097, 13918106, 13918115, 13918124, 13918133, 13918142, 13918151, 13918160, 13918169, 13918178, 13918187, 13918196, 13918205, 13918214, 13918223, 13918232, 13918241, 13918250, 13918259, 13918268, 13918277, 13918286, 13918295, 13918304, 13918313, 13918322, 13918331, 13918340, 13918349, 13918358, 13918367, 13918376, 13918385, 13918394, 13918403, 13918412, 13918421, 13918430, 13918439, 13918448, 13918457, 13918466, 13918475, 13918484, 13918493, 13918502, 13918511, 13918520, 13918529, 13918538, 13918547, 13918556, 13918565, 13918574, 13918583, 13918592, 13918601, 13918610, 13918619, 13918628, 13918637, 13918646, 13918655, 13918664, 13918673, 13918682, 13918691, 13918700, 13918709, 13918718, 13918727, 13918736, 13918745, 13918754, 13918763, 13918772, 13918781, 13918790, 13918799, 13918808, 13918817, 13918826, 13918835, 13918844, 13918853, 13918862, 13918871, 13918880, 13918889, 13918898, 13918907, 13918916, 13918925, 13918934, 13918943, 13918952, 13918961, 13918970, 13918979, 13918988, 13918997, 13919006, 13919015, 13919024, 13919033, 13919042, 13919051, 13919060, 13919069, 13919078, 13919087, 13919096, 13919105, 13919114, 13919123, 13919132, 13919141, 13919150, 13919159, 13919168, 13919177, 13919186, 13919195, 13919204, 13919213, 13919222, 13919231, 13919240, 13919249, 13919258, 13919267, 13919276, 13919285, 13919294, 13919303, 13919312, 13919321, 13919330, 13919339, 13919348, 13919357, 13919366, 13919375, 13919384, 13919393, 13919402, 13919411, 13919420, 13919429, 13919438, 13919447, 13919456, 13919465, 13919474, 13919483, 13919492, 13919501, 13919510, 13919519, 13919528, 13919537, 13919546, 13919555, 13919564, 13919573, 13919582, 13919591, 13919600, 13919609, 13919618, 13919627, 13919636, 13919645, 13919654, 13919663, 13919672, 13919681, 13919690, 13919699, 13919708, 13919717, 13919726, 13919735, 13919744, 13919753, 13919762, 13919771, 13919780, 13919789, 13919798, 13919807, 13919816, 13919825, 13919834, 13919843, 13919852, 13919861, 13919870, 13919879, 13919888, 13919897, 13919906, 13919915, 13919924, 13919933, 13919942, 13919951, 13919960, 13919969, 13919978, 13919987, 13919996, 14000005, 14000014, 14000023, 14000032, 14000041, 14000050, 14000059, 14000068, 14000077, 14000086, 14000095, 14000104, 14000113, 14000122, 14000131, 14000140, 14000149, 14000158, 14000167, 14000176, 14000185, 14000194, 14000203, 14000212, 14000221, 14000230, 14000239, 14000248, 14000257, 14000266, 14000275, 14000284, 14000293, 14000302, 14000311, 14000320, 14000329, 14000338, 14000347, 14000356, 14000365, 14000374, 14000383, 14000392, 14000401, 14000410, 14000419, 14000428, 14000437, 14000446, 14000455, 14000464, 14000473, 14000482, 14000491, 14000500, 14000509, 14000518, 14000527, 14000536, 14000545, 14000554, 14000563, 14000572, 14000581, 14000590, 14000599, 14000608, 14000617, 14000626, 14000635, 14000644, 14000653, 14000662, 14000671, 14000680, 14000689, 14000698, 14000707, 14000716, 14000725, 14000734, 14000743, 14000752, 14000761, 14000770, 14000779, 14000788, 14000797, 14000806, 14000815, 14000824, 14000833, 14000842, 14000851, 14000860, 14000869, 14000878, 14000887, 14000896, 14000905, 14000914, 14000923, 14000932, 14000941, 14000950, 14000959, 14000968, 14000977, 14000986, 14000995, 14001004, 14001013, 14001022, 14001031, 14001040, 14001049, 14001058, 14001067, 14001076, 14001085, 14001094, 14001103, 14001112, 14001121, 14001130, 14001139, 14001148, 14001157, 14001166, 14001175, 14001184, 14001193, 14001202, 14001211, 14001220, 14001229, 14001238, 14001247, 14001256, 14001265, 14001274, 14001283, 14001292, 14001301, 14001310, 14001319, 14001328, 14001337, 14001346, 14001355, 14001364, 14001373, 14001382, 14001391, 14001400, 14001409, 14001418, 14001427, 14001436, 14001445, 14001454, 14001463, 14001472, 14001481, 14001490, 14001499, 14001508, 14001517, 14001526, 14001535, 14001544, 14001553, 14001562, 14001571, 14001580, 14001589, 14001598, 14001607, 14001616, 14001625, 14001634, 14001643, 14001652, 14001661, 14001670, 14001679, 14001688, 14001697, 14001706, 14001715, 14001724, 14001733, 14001742, 14001751, 14001760, 14001769, 14001778, 14001787, 14001796, 14001805, 14001814, 14001823, 14001832, 14001841, 14001850, 14001859, 14001868, 14001877, 14001886, 14001895, 14001904, 14001913, 14001922, 14001931, 14001940, 14001949, 14001958, 14001967, 14001976, 14001985, 14001994, 14002003, 14002012, 14002021, 14002030, 14002039, 14002048, 14002057, 14002066, 14002075, 14002084, 14002093, 14002102, 14002111, 14002120, 14002129, 14002138, 14002147, 14002156, 14002165, 14002174, 14002183, 14002192, 14002201, 14002210, 14002219, 14002228, 14002237, 14002246, 14002255, 14002264, 14002273, 14002282, 14002291, 14002300, 14002309, 14002318, 14002327, 14002336, 14002345, 14002354, 14002363, 14002372, 14002381, 14002390, 14002399, 14002408, 14002417, 14002426, 14002435, 14002444, 14002453, 14002462, 14002471, 14002480, 14002489, 14002498, 14002507, 14002516, 14002525, 14002534, 14002543, 14002552, 14002561, 14002570, 14002579, 14002588, 14002597, 14002606, 14002615, 14002624, 14002633, 14002642, 14002651, 14002660, 14002669, 14002678, 14002687, 14002696, 14002705, 14002714, 14002723, 14002732, 14002741, 14002750, 14002759, 14002768, 14002777, 14002786, 14002795, 14002804, 14002813, 14002822, 14002831, 14002840, 14002849, 14002858, 14002867, 14002876, 14002885, 14002894, 14002903, 14002912, 14002921, 14002930, 14002939, 14002948, 14002957, 14002966, 14002975, 14002984, 14002993, 14003002, 14003011, 14003020, 14003029, 14003038, 14003047, 14003056, 14003065, 14003074, 14003083, 14003092, 14003101, 14003110, 14003119, 14003128, 14003137, 14003146, 14003155, 14003164, 14003173, 14003182, 14003191, 14003200, 14003209, 14003218, 14003227, 14003236, 14003245, 14003254, 14003263, 14003272, 14003281, 14003290, 14003299, 14003308, 14003317, 14003326, 14003335, 14003344, 14003353, 14003362, 14003371, 14003380, 14003389, 14003398, 14003407, 14003416, 14003425, 14003434, 14003443, 14003452, 14003461, 14003470, 14003479, 14003488, 14003497, 14003506, 14003515, 14003524, 14003533, 14003542, 14003551, 14003560, 14003569, 14003578, 14003587, 14003596, 14003605, 14003614, 14003623, 14003632, 14003641, 14003650, 14003659, 14003668, 14003677, 14003686, 14003695, 14003704, 14003713, 14003722, 14003731, 14003740, 14003749, 14003758, 14003767, 14003776, 14003785, 14003794, 14003803, 14003812, 14003821, 14003830, 14003839, 14003848, 14003857, 14003866, 14003875, 14003884, 14003893, 14003902, 14003911, 14003920, 14003929, 14003938, 14003947, 14003956, 14003965, 14003974, 14003983, 14003992, 14004001, 14004010, 14004019, 14004028, 14004037, 14004046, 14004055, 14004064, 14004073, 14004082, 14004091, 14004100, 14004109, 14004118, 14004127, 14004136, 14004145, 14004154, 14004163, 14004172, 14004181, 14004190, 14004199, 14004208, 14004217, 14004226, 14004235, 14004244, 14004253, 14004262, 14004271, 14004280, 14004289, 14004298, 14004307, 14004316, 14004325, 14004334, 14004343, 14004352, 14004361, 14004370, 14004379, 14004388, 14004397, 14004406, 14004415, 14004424, 14004433, 14004442, 14004451, 14004460, 14004469, 14004478, 14004487, 14004496, 14004505, 14004514, 14004523, 14004532, 14004541, 14004550, 14004559, 14004568, 14004577, 14004586, 14004595, 14004604, 14004613, 14004622, 14004631, 14004640, 14004649, 14004658, 14004667, 14004676, 14004685, 14004694, 14004703, 14004712, 14004721, 14004730, 14004739, 14004748, 14004757, 14004766, 14004775, 14004784, 14004793, 14004802, 14004811, 14004820, 14004829, 14004838, 14004847, 14004856, 14004865, 14004874, 14004883, 14004892, 14004901, 14004910, 14004919, 14004928, 14004937, 14004946, 14004955, 14004964, 14004973, 14004982, 14004991, 14005000, 14005009, 14005018, 14005027, 14005036, 14005045, 14005054, 14005063, 14005072, 14005081, 14005090, 14005099, 14005108, 14005117, 14005126, 14005135, 14005144, 14005153, 14005162, 14005171, 14005180, 14005189, 14005198, 14005207, 14005216, 14005225, 14005234, 14005243, 14005252, 14005261, 14005270, 14005279, 14005288, 14005297, 14005306, 14005315, 14005324, 14005333, 14005342, 14005351, 14005360, 14005369, 14005378, 14005387, 14005396, 14005405, 14005414, 14005423, 14005432, 14005441, 14005450, 14005459, 14005468, 14005477, 14005486, 14005495, 14005504, 14005513, 14005522, 14005531, 14005540, 14005549, 14005558, 14005567, 14005576, 14005585, 14005594, 14005603, 14005612, 14005621, 14005630, 14005639, 14005648, 14005657, 14005666, 14005675, 14005684, 14005693, 14005702, 14005711, 14005720, 14005729, 14005738, 14005747, 14005756, 14005765, 14005774, 14005783, 14005792, 14005801, 14005810, 14005819, 14005828, 14005837, 14005846, 14005855, 14005864, 14005873, 14005882, 14005891, 14005900, 14005909, 14005918, 14005927, 14005936, 14005945, 14005954, 14005963, 14005972, 14005981, 14005990, 14006000, 14006009, 14006018, 14006027, 14006036, 14006045, 14006054, 14006063, 14006072, 14006081, 14006090, 14006099, 14006108, 14006117, 14006126, 14006135, 14006144, 14006153, 14006162, 14006171, 14006180, 14006189, 14006198, 14006207, 14006216, 14006225, 14006234, 14006243, 14006252, 14006261, 14006270, 14006279, 14006288, 14006297, 14006306, 14006315, 14006324, 14006333, 14006342, 14006351, 14006360, 14006369, 14006378, 14006387, 14006396, 14006405, 14006414, 14006423, 14006432, 14006441, 14006450, 14006459, 14006468, 14006477, 14006486, 14006495, 14006504, 14006513, 14006522, 14006531, 14006540, 14006549, 14006558, 14006567, 14006576, 14006585, 14006594, 14006603, 14006612, 14006621, 14006630, 14006639, 14006648, 14006657, 140066

3888488 4076061 4216142 4261282 4301393 4340504 4379615 4418726 4457837 4496948 4536059 4575170 4614281 4653392 4692503 4731614 4770725 4809836 4848947 4888058 4927169 4966280 5005391 5044502 5083613 5122724 5161835 5200946 5240057 5279168 5318279 5357390 5396501 5435612 5474723 5513834 5552945 5592056 5631167 5670278 5709389 5748500 5787611 5826722 5865833 5904944 5944055 5983166 6022277 6061388 6100499 6139610 6178721 6217832 6256943 6296054 6335165 6374276 6413387 6452498 6491609 6530720 6569831 6608942 6648053 6687164 6726275 6765386 6804497 6843608 6882719 6921830 6960941 7000052 7039163 7078274 7117385 7156496 7195607 7234718 7273829 7312940 7352051 7391162 7430273 7469384 7508495 7547606 7586717 7625828 7664939 7704050 7743161 7782272 7821383 7860494 7899605 7938716 7977827 8016938 8056049 8095160 8134271 8173382 8212493 8251604 8290715 8329826 8368937 8408048 8447159 8486270 8525381 8564492 8603603 8642714 8681825 8720936 8760047 8799158 8838269 8877380 8916491 8955602 8994713 9033824 9072935 9112046 9151157 9190268 9229379 9268490 9307601 9346712 9385823 9424934 9464045 9503156 9542267 9581378 9620489 9659600 9698711 9737822 9776933 9816044 9855155 9894266 9933377 9972488 10000000

Main body of the document containing dense alphanumeric data, likely a list of identifiers or coordinates, organized in vertical columns.

