

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 761 633**

51 Int. Cl.:

C12Q 1/6881 (2008.01)

C12Q 1/6827 (2008.01)

G01N 33/49 (2006.01)

G16B 20/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.07.2016 PCT/EP2016/066654**

87 Fecha y número de publicación internacional: **20.04.2017 WO17063769**

96 Fecha de presentación y número de la solicitud europea: **13.07.2016 E 16738771 (1)**

97 Fecha y número de publicación de la concesión europea: **27.11.2019 EP 3362928**

54 Título: **Método y sistema para la estimación de si una hembra está gestante basándose en una muestra de sangre**

30 Prioridad:

14.10.2015 BE 201505658

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

20.05.2020

73 Titular/es:

**MULTIPLICOM NV (100.0%)
Galileilaan 18
2845 Niel, BE**

72 Inventor/es:

**DEL FAVERO, JURGEN PETER LODE;
VYVERMAN, MICHAËL y
VAUTERIN, PAUL**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 761 633 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

Método y sistema para la estimación de si una hembra está gestante basándose en una muestra de sangre

5 **Campo de la invención**

El campo de la invención se refiere a la estimación de si una hembra está gestante o no basándose en una muestra de sangre. Realizaciones particulares de la invención se refieren a métodos, sistemas, programas de computadora y productos de programación de computadora para la estimación de si una hembra está gestante o no.

10

Antecedentes

El documento WO 2013/057568 en nombre del solicitante desvela métodos de detección prenatal utilizando técnicas no invasivas. En particular, se refiere al diagnóstico prenatal de una aneuploidía cromosómica fetal detectando ácidos nucleicos fetal y maternal en una muestra biológica maternal. Más particularmente, el documento WO 2013/057568 aplica una PCR múltiple para amplificar fracciones seleccionadas de los respectivos cromosomas de maternales y fetales. Las cantidades respectivas de regiones cromosómicas sospechosas de aneuploidía y los cromosomas se determinaron a partir de un análisis de secuenciación masivo seguido por el análisis estadístico para detectar una aneuploidía particular.

15

20

Tynan et al. J Mol Diagn 2011, 13:382-389 describe la detección de alelos paternos en el ADN plasmático maternal.

25

Aunque existen distintos métodos de detección prenatal, existe la necesidad de un método mejorado para la estimación de si una hembra está gestante o no el caso de que solo se proporcione una muestra de sangre de la hembra.

Sumario

30

La invención se define por las reivindicaciones. El objetivo de las realizaciones de la invención es proporcionar un método, sistema y un programa de computadora para la estimación de si una hembra está gestante o no basándose en una muestra de sangre.

35

De acuerdo con un primer aspecto de la invención se proporciona un método para la estimación de si una hembra está gestante. El método comprende la medición de presencias alélicas (D) para una pluralidad de marcadores genéticos de al menos un cromosoma, diferente del cromosoma X e Y, en una muestra de un ADN libre de células de una hembra potencialmente gestante. Cada presencia alélica representa la presencia en un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal. Con el fin de completarlo, se señala que no implica que se pueda hacer una distinción entre un alelo de origen maternal y un alelo de origen fetal, sino simplemente que se puede medir un alelo de cualquier origen. El método comprende adicionalmente, basándose en dichas presencias alélicas medidas, la determinación de una fracción homocigota (F_{ho}) en las mismas que se asocia con marcadores genéticos puramente homocigotos; y la estimación de si la hembra está gestante basándose en dicha fracción donde la estimación comprende la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado y la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado, donde el primer y segundo valores predeterminados pueden ser el mismo o donde el segundo valor predeterminado puede ser mayor que el primer valor predeterminado. En un evento de una muestra de hembra gestante, la expresión "marcador genético puramente homocigoto" se refiere a un marcador genético que es homocigoto en el ADN de origen maternal como en el ADN de origen fetal. En un evento de una muestra de hembra no gestante, la expresión "marcador genético puramente homocigoto" se refiere a un marcador genético que es homocigoto en el ADN de la hembra.

40

45

50

55

Las realizaciones de la invención se basan *inter alia* en la visión inventiva de que se espera que la fracción homocigota sea mayor para las muestras de la mujer no gestante que para las muestras de la mujer gestante. La razón es que la presencia del ADN fetal desplazará una parte del estado puramente homocigoto debido a la presencia de ADN basado en el ADN del padre. Esta fracción se puede determinar sin la detección actual de la presencia de alelos paternos.

60

En el contexto de la presente memoria descriptiva, un 'marcador genético' es una posición del genoma que se sabe que tiene varios estados posibles en los individuos de una población.

65

En una realización ejemplar, la determinación de la fracción homocigota comprende: basándose en dicha presencia alélica medida para la pluralidad de marcadores genéticos, el cálculo de un número correspondiente de frecuencia alélica para dicha pluralidad y la determinación como fracción homocigota la fracción de dicha presencia alélica medida para la que la frecuencia de alelos es 0 o 1 dentro de un margen de error predeterminado. Este margen de error predeterminado puede referirse a la tasa de error intrínseco de un dispositivo de secuenciación y se puede

determinar empíricamente. Un experto en la técnica entenderá que la tasa de error de un dispositivo de secuenciación es una convención industrial y estará bien reconocido para un fabricante o instrumento específico, pero en una realización específica el margen de error predeterminado es $< 0,02$. Particularmente, puede ser $< 0,015$, $< 0,01$ o $< 0,005$. Normalmente, la presencia del alelo se grafica como recuentos de lectura en función de la frecuencia del alelo, y la fracción homocigota se corresponde con los recuentos de lectura de los puntos de datos variantes concentrados alrededor de la frecuencia de alelo 0 y alrededor de la frecuencia de alelo 1.

La estimación comprende: la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado; y la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado. El primer y segundo valores predeterminados pueden ser el mismo. De manera alternativa, el segundo valor predeterminado puede ser mayor que el primer valor predeterminado. En una realización posible el primer umbral y el segundo umbral tienen un valor constante, opcionalmente, el mismo valor constante, que se puede haber determinado empíricamente. Utilizando los métodos que se describen en el presente documento, los valores predeterminados para la determinación de si un individuo se puede clasificar como gestante están entre 0,50-0,75, particularmente entre 0,55-0,70, más particularmente entre 0,60-0,68. Sin embargo, se entiende que otros métodos y otras realizaciones pueden necesitar diferentes valores del umbral. Un experto en la técnica será capaz de determinar valores predeterminados adecuados para la fracción homocigota para la población específica que se va a ensayar, por ejemplo, a partir del ensayo de un pequeño conjunto de individuos cuyo estado de gestación se conoce (por ejemplo, utilizando un conjunto de datos de 5-15 hembras gestantes y 5-15 hembras no gestantes, utilizando específicamente 10 hembras gestantes y 10 no gestantes).

En una realización ejemplar el método comprende adicionalmente, basándose en dicha presencia de alelo medida, la determinación de una fracción heterocigota (F_{he}) de la misma que se asocia con marcadores genéticos que son heterocigotos en el ADN de la hembra. La estimación puede comprender entonces si la hembra está gestante basándose en la fracción homocigota y dicha fracción heterocigota. En una realización desarrollada adicionalmente del mismo, el método comprende el cálculo de un valor de umbral en función de la fracción heterocigota. La estimación comprende: la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un valor de umbral calculado; y la estimación de que la hembra no está gestante si la fracción homocigota está por encima del valor de umbral calculado. De esta manera, la precisión del método se puede mejorar adicionalmente debido a que esto puede compensar diferentes niveles de homocigosis total en el ADN de la hembra.

En una realización ejemplar la determinación de una fracción homocigota (F_{ho}) que se asocia con marcadores genéticos puramente homocigotos comprende, basándose en dicha presencia de alelo medida una pluralidad de marcadores genéticos, el cálculo del número correspondiente de frecuencia de alelos para dicha pluralidad de marcadores genéticos; y la determinación de recuentos de lectura en función de la frecuencia de alelos para dicha pluralidad de marcadores genéticos; la selección de una parte de los recuentos de lectura, preferentemente una parte en la que se hayan retirado los valores de recuento de lecturas más altos y/o los más bajos; y la determinación de la fracción homocigota (F_{ho}) de dicha parte seleccionada, que se asocia con marcadores genéticos puramente homocigotos. Seleccionando una parte de los recuentos de lectura en vez de utilizar todos los recuentos de lectura, el método se puede mejorar adicionalmente. Más en particular, retirando los recuentos de lecturas con los valores más bajos, la precisión se mejora ya que los recuentos de lectura más altos tienen una frecuencia de alelo más precisa, debido a que hay más datos de medición; y retirando los recuentos de lectura con los valores más altos, se evita que estos recuentos de lectura muy altos tengan una influencia desproporcionada sobre la determinación de la fracción. En una realización particular, un experto en la técnica identificará antes de ejecutar el experimento los umbrales para retirar dichos valores "atípicos" teniendo en cuenta la media y la desviación típica de los valores de recuentos de lectura. En una realización alternativa, el umbral más bajo para desechar los valores de recuento de lectura más bajos se puede seleccionar basándose en la tasa de error específico del método o dispositivo de secuenciación utilizado. En una realización particular, se retira el 10 % más bajo de los valores de recuento de lectura, más particularmente se retira el 5 % más bajo, el 3 % más bajo, el 1 % más bajo o el 0,5 % más bajo de los valores de recuento de lectura. En una realización particular, se retira el 10 % más alto de los valores de recuento de lectura, más particularmente se retira el 5 % más alto, el 3 % más alto, el 1 % más alto o el 0,5 % más alto de los valores de recuento de lectura.

En una realización ejemplar, las etapas de medición y determinación se llevan a cabo para un lote que comprende una pluralidad de muestras de la hembra; en el que se calcula la fracción homocigota cada muestra del lote, y la estimación se basa adicionalmente en dicha fracción homocigota de cada muestra.

De acuerdo con otro aspecto de la invención, se proporciona un sistema para la estimación de si una hembra está gestante. El sistema comprende un dispositivo de medición, un módulo de determinación, y un módulo de estimación. El dispositivo de medición se configura para la medición de presencias alélicas para una pluralidad de marcadores genéticos de al menos un cromosoma, diferente del cromosoma X e Y, en una muestra de ADN libre de células de una hembra potencialmente gestante; representando cada presencia alélica la presencia de un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal. Este puede ser cualquier dispositivo de medición disponible en el mercado que sea adecuado para llevar a cabo dichas mediciones. El módulo de determinación se configura para la determinación, basándose en

- dichas presencias alélicas medidas, de una fracción homocigota de las mismas que se asocian con marcadores genéticos puramente homocigotos. El módulo de estimación se configura para la estimación de si la hembra está gestante basándose en dicha fracción, donde el módulo de estimación se configura para la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado y la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado, donde el primer y segundo valores predeterminados pueden ser el mismo o donde el segundo valor predeterminado puede ser mayor que el primer valor predeterminado. En una realización típica, el módulo de determinación y el módulo de estimación se implementan como un software.
- 10 En una realización ejemplar, el módulo de determinación se configura para calcular un número correspondiente de frecuencias alélicas para dicha pluralidad de marcadores genéticos basándose en dichas presencias alélicas medidas; y para la determinación como fracción homocigota la fracción de dicha presencia alélica medida para la que la frecuencia de alelos es 0 o 1 dentro de un margen de error predeterminado. Un experto en la técnica entenderá que la tasa de error de un dispositivo de secuenciación es una convención industrial y estará bien reconocido para un fabricante o instrumento específico, pero en una realización específica el margen de error predeterminado es $< 0,02$. Particularmente, puede ser $< 0,015$, $< 0,01$ o $< 0,005$.
- 20 El módulo de estimación se configura para la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado y la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado, donde el primer y segundo valores predeterminados pueden ser el mismo o donde el segundo valor predeterminado puede ser mayor que el primer valor predeterminado. Utilizando los métodos que se describen en el presente documento, los valores predeterminados para la determinación de si un individuo se puede clasificar como gestante están entre 0,50-0,75, particularmente entre 0,55-0,70, más particularmente entre 0,60-0,68. Sin embargo, se entiende que otros métodos y otras realizaciones pueden necesitar diferentes valores del umbral. Un experto en la técnica será capaz de determinar valores predeterminados adecuados para la fracción homocigota para la población específica que se va a ensayar, por ejemplo, a partir del ensayo de un pequeño conjunto de individuos cuyo estado de gestación se conoce (por ejemplo, utilizando un conjunto de datos de 5-15 hembras gestantes y 5-15 hembras no gestantes, utilizando específicamente 10 hembras gestantes y 10 no gestantes).
- 30 En una realización ejemplar el módulo de determinación se configura adicionalmente para, basándose en dichas presencias alélicas medidas, la determinación de una fracción heterocigota (F_{he}) de las mismas que se asocia con marcadores genéticos heterocigotos. El módulo de estimación puede configurarse adicionalmente entonces para la estimación de si la hembra está gestante basándose en la fracción homocigota y dicha fracción heterocigota.
- 35 En una realización ejemplar, el módulo de determinación se configura adicionalmente para el cálculo de un valor de umbral en función de la fracción heterocigota. El módulo de estimación se puede configurar adicionalmente entonces para la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un valor de umbral calculado; y la estimación de que la hembra no está gestante si la fracción homocigota está por encima del valor de umbral calculado. La fracción heterocigota se determina como una función lineal con una pendiente negativa y una intersección positiva. Los métodos ejemplares para la determinación de la fracción heterocigota se describen adicionalmente en los ejemplos, y en la Figura 4. Un experto en la técnica es capaz de utilizar las enseñanzas del presente documento para determinar el umbral en función de la fracción heterocigota, utilizando los métodos descritos o alternativas que estén dentro de su conocimiento general común.
- 45 En una realización ejemplar, el dispositivo de medición y el módulo de determinación se configuran para llevar a cabo la medición y determinación de un lote que comprende una pluralidad de muestra, donde para cada muestra del lote, se calcula la fracción homocigota. El módulo de estimación puede configurarse entonces para la estimación de si la hembra está gestante basándose en dicha fracción homocigota de cada muestra.
- 50 En una realización preferida, el módulo de estimación se configura para la estimación del estado de gestación de la hembra de una manera como se ha descrito en cualquiera de las realizaciones ejemplares del método.
- 55 En una realización preferida, el dispositivo de medición se configura para medir las presencias alélicas utilizando al menos una de las siguientes: reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa, amplificación basada en la secuencia de ácidos nucleicos (NASBA), y métodos de ADN ramificado; y preferentemente PCR.
- 60 En realizaciones ejemplares de la invención, la medición de presencias alélicas puede comprender la medición de presencias alélicas de SNP y/o la medición de presencias alélicas de inserciones y/o eliminaciones cortas.
- Se desvelan realizaciones preferidas del método y sistema de la invención en las reivindicaciones dependientes adjuntas.
- 65 De acuerdo con un aspecto adicional de la invención, se proporciona un programa de computadora que comprenden instrucciones ejecutables por computadora para llevar a cabo, cuando se ejecuta el programa en una computadora,

de al menos la etapa de estimación de las realizaciones del método desvelado anteriormente. De acuerdo con otro aspecto se proporciona un dispositivo de almacenamiento de datos que codifica un programa en una forma leíble por la máquina y ejecutable por la máquina para llevar a cabo al menos la etapa de estimación de una cualquiera de las realizaciones del método desvelado anteriormente.

5 La referencia a instrucciones/forma ejecutable por computadora se tiene que considerar como que comprende tanto un código ejecutable por la máquina directamente, código que se debe compilar para ejecutarse, como el código que se interpreta en vez de ejecutarse *per se*.

10 Breve descripción de las figuras

Los dibujos adjuntos se utilizan para ilustrar las realizaciones ejemplares no limitantes actuales de un método y sistema de la presente invención. Las ventajas anteriores y otras de las características y objetivos de la invención se volverán más evidentes y la invención se comprenderá mejor a partir de la siguiente descripción detallada cuando se lea en conjunción con los dibujos adjuntos, en los que:

- La Figura 1 es un gráfico de dispersión que presenta la cobertura total de SNP (eje y) frente a la frecuencia alélica (eje x) para una muestra no gestante NP;
- La Figura 2 es un gráfico de dispersión que presenta la cobertura total de SNP (eje y) frente a la frecuencia alélica (eje x) para una muestra gestante P;
- La Figura 3 es un histograma de las fracciones homocigotas de un conjunto de muestras de hembras gestantes y no gestantes;
- La Figura 4 es un gráfico de dispersión que muestra la F_{ho} en el eje y, y F_{he} en el eje x para un conjunto de muestras de hembras gestantes y no gestantes, así como una línea horizontal que indica el valor del umbral F_T calculado empíricamente;
- La Figura 5 es un gráfico de dispersión que muestra la F_{ho} en el eje y, y F_{he} en el eje x para un conjunto de muestras de hembras gestantes y no gestantes, así como una línea que indica el valor del umbral calculado $F_T (F_{Rhe})$;
- La Figura 6 es un dibujo esquemático de una realización de un sistema.

30 Descripción detallada de las realizaciones ejemplares

En un Ensayo Prenatal No Invasivo (NIPT), conocido en la técnica anterior, se secuencian el ADN libre de células (cfADN) de una muestra de suero o plasma materno de una hembra gestante con el fin de explorar la presencia de aneuploidías cromosómicas en el feto, tal como la trisomía del cromosoma 21. De acuerdo con realizaciones ejemplares de la invención, se proporciona un método para la estimación de si una hembra está gestante.

En una realización típica, la muestra de suero o plasma materno se deriva de la sangre materna. Esta puede ser una pequeña cantidad de suero o plasma, por ejemplo, de 1 a 20 ml. Dependiendo de la precisión deseada se puede preferir la utilización de volúmenes mayores. La preparación del suero o plasma a partir de la muestra de sangre materna se puede llevar a cabo utilizando técnicas convencionales. Las técnicas adecuadas incluyen la centrifugación y las técnicas basadas en matriz. En realizaciones posibles, se puede utilizar un método de enriquecimiento basado en la secuencia sobre el suero o plasma materno para enriquecer específicamente las secuencias de ácido nucleico fetal.

Las realizaciones del método de la invención se pueden llevar a cabo para una muestra que contenga ADN fetal con una concentración de la fracción fetal de la cantidad de ADN por encima de un umbral predeterminado. En realizaciones preferidas, se lleva a cabo una amplificación de las secuencias de ADN fetal en la muestra. Se puede utilizar cualquier método de amplificación conocido por el experto, tal como un método de PCR.

En una realización preferida solamente se utilizan los datos del ensayo Clarigo del solicitante que no implican la detección de alelos con SNP (o polimorfismos de nucleótido único, es decir, un marcador genético que comprende un único nucleótido variable) en el ADN fetal que no están presentes en el ADN de la hembra gestante. El ensayo Clarigo consiste en una secuenciación direccionada de varias regiones del genoma humano (en otras palabras, direccionando marcadores genéticos específicos), utilizando SNP (polimorfismos de nucleótidos únicos) conocidos con una alta prevalencia (por ejemplo, mayor del 1 %, preferentemente mayor del 10 %) y dos posibles alelos (un alelo de referencia conocido como REF; y un alelo alternativo conocido como ALT). Se pueden encontrar más detalles sobre el ensayo Clarigo en internet en <http://www.multiplicom.com/product/clarigo> y en el documento WO 2013/057568 a nombre del solicitante.

Ahora se expondrá con detalle una realización ejemplar de un método para la estimación de si una hembra está gestante. En una primera etapa de medición, se miden las presencias alélicas de una pluralidad (D_R) de marcadores genéticos de al menos un cromosoma, diferentes del cromosoma X e Y, en una muestra de ADN libre de células de una hembra. Cada presencia alélica representa la presencia en un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal. En una segunda etapa de

cálculo, basándose en dichas presencias alélicas medidas para dicha pluralidad, se determina una fracción homocigota (F_{ho}) de las mismas, que se asocian con marcadores genéticos puramente homocigotos. En una tercera etapa se estima basándose en dicha fracción homocigota si la hembra está gestante.

5 Medición y determinación de la fracción

Una manera ventajosa para representar los resultados de la medición de presencias alélicas para un marcador genético es asociar la siguiente información con un punto de datos variante para ese marcador genético. Un punto de datos variante (que es un punto de datos asociado con un número de variantes, tales como alelos) se utiliza en esta memoria descriptiva como una representación conveniente para un marcador genético, y por lo tanto representa los resultados de la medición de la presencia alélica en un número de amplicones para los marcadores genéticos. Un amplicón es un trozo de ADN o ARN que es (la fuente y/o) el producto de eventos de amplificación o replicación. En otras palabras, un amplicón es un trozo biofísico de material de replicación, diseñado para que contenga una posición de SNP conocida con alta prevalencia en la población. Cada punto de datos variantes se asocia de esta manera con un SNP conocido con alta prevalencia en la población y con dos alelos posibles (un alelo de referencia conocido como REF; y un alelo alternativo conocido como ALT). Para cada punto de datos variante A_i , se pueden determinar los siguientes números utilizando, por ejemplo, una tubería bioinformática convencional aplicada sobre los datos de secuenciación:

- El número de lecturas que contienen el alelo REF sobre la posición conocida de SNP, C_{Ri} .
- El número de lecturas que contienen el alelo ALT sobre la posición conocida de SNP, C_{Ai} .
- La cobertura total $C_{Ti} = C_{Ri} + C_{Ai}$.
- La frecuencia alélica, o la fracción de lecturas del alelo ALT sobre la cobertura total

$$F_i = C_{Ai} / (C_{Ri} + C_{Ai}).$$

Por lo tanto, para un marcador genético i determinado, se pueden medir las presencias alélicas del alelo REF, el alelo ALT, y ambos alelos, midiendo el número de lecturas que contienen el alelo REF, el alelo ALT y ambos alelos REF y ALT, respectivamente. Basándose en las presencias alélicas medidas, se calcula un número de frecuencias alélicas correspondiente para el número predeterminado de marcadores genéticos.

Para cada posición del genoma (es decir, para cada locus genético), excluyendo los cromosomas X e Y, y asumiendo que no hay trastornos cromosómicos relevantes, en el caso de una muestra de hembra gestante, hay cuatro copias presentes en la muestra (asumiendo que la posición no es parte de una región de aneuploidía), que determina el número total de lecturas: dos copias del ADN materno y dos copias del ADN fetal.

Para los puntos de datos variantes individuales (es decir, para un marcador genético individual), A y B denotan el alelo REF y ALT para el SNP conocido en el ADN materno para ese marcador genético, y a y b los correspondientes estados para el ADN fetal de una hembra gestante. Para una muestra no gestante, el punto de datos variantes puede estar en tres estados posibles AA, AB, BB con la fracción esperada respectiva de lecturas del alelo ALT para cada uno de estos estados (F_i) 0, 0.5 and 1. Para una muestra gestante, el punto de datos variante puede estar en los posibles estados que se enumeran en la Tabla 1:

Tabla 1

Estado del punto de datos variantes	Fracción esperada de lecturas ALT (F_i)
AAaa	0
AAab	$FF/2$
ABaa	$0,5 - FF/2 \leftarrow (1 - FF) / 2$
ABab	0,5
ABbb	$0,5 + FF/2 \leftarrow (1 - FF) / 2 + FF$
BBab	$1 - FF/2$
BBbb	1

Como ilustración de las presencias alélicas medidas, se midieron los gráficos de dispersión ilustrados en las Figuras 1 y 2 muestran datos para una muestra NP (hembra no gestante) y una muestra P (hembra gestante) para cuyas presencias alélicas para marcadores genéticos para al menos un cromosoma, que sea diferente del cromosoma X y el cromosoma Y, donde:

- Cada punto es un punto de datos variante (que representa el resultado de las mediciones hechas sobre

amplicones) para los marcadores genéticos de al menos un cromosoma, que sea diferente del cromosoma X y el cromosoma Y.

- El eje horizontal muestra la fracción de lecturas con el SNP del alelo ALT (es decir, la frecuencia alélica F_i).
- El eje vertical muestra la cobertura de lectura total C_{π} .

5 Se puede ver en la Figura 1 que los puntos de datos variantes se asocian con una frecuencia alélica específica. Por ejemplo, el punto de datos variante 101 se muestra a la izquierda y representa un marcador genético para el cual se llevaron a cabo aproximadamente 2300 lecturas. Todas o casi todas las lecturas para este marcador han medido la presencia alélica para indicar que los alelos REF (es decir, A) están presentes, pero no están presentes los alelos ALT (es decir, B). Por lo tanto, el punto de datos variantes 101 se ha representado a la izquierda en la Figura 1, donde la frecuencia alélica es aproximadamente 0, y probablemente representa un marcador homocigoto AA. El punto de datos variante 102 se muestra a la derecha y representa un marcador genético para el cual se llevaron a cabo aproximadamente 2300 lecturas. Todas o casi todas las lecturas para este marcador han medido la presencia alélica para indicar que los alelos REF (es decir, A) no están presentes pero que los alelos ALT (es decir, B) están presentes. Por lo tanto, el punto de datos variantes 102 se ha representado a la derecha en la Figura 1, donde la frecuencia alélica es aproximadamente 1, y probablemente representa un marcador homocigoto BB.

En la Figura 2 el punto de datos 103 representa un marcador genético para el que se llevaron a cabo aproximadamente 2300 lecturas. Las presencias alélicas medidas permiten calcular la frecuencia alélica, que es relativamente baja, pero no cero. Por lo tanto, el punto de datos variantes 103 probablemente representa un marcador genético que es maternamente homocigoto para el par de alelos de referencia, pero tiene un par de alelos heterocigotos de origen fetal (por tanto, AAab), mientras que la fracción de ADN materno presente en la muestra para el marcador genético es (más) grande que la fracción del ADN fetal presente en la muestra para el marcador genético. El punto de datos variante 104 se muestra a la derecha y representa un marcador genético para el cual se llevaron a cabo aproximadamente 2100 lecturas. Con el mismo razonamiento, el punto de datos variantes 104 representa probablemente un marcador que es maternamente homocigoto para el par de alelos alternativos, pero tiene un par de alelos heterocigotos de origen fetal (por tanto, BBab). Nótese que al remarcar que 103 y 104 representan probablemente un marcador que es homocigoto en el ADN materno y heterocigoto en el ADN fetal solo se hace para ayudar al entendimiento de las figuras. Las realizaciones ejemplares de la invención no se basan en dicha evaluación, ya que los marcadores genéticos representados por 103 y 104 sol ose marcan como no homocigotos.

En la Figura 1 el punto de datos 105 representa un marcador genético para el que se llevaron a cabo aproximadamente 2300 lecturas. Las presencias alélicas medidas permiten calcular la frecuencia alélica, que se ha descubierto que es aproximadamente de 0,5. Por lo tanto, el punto de datos variante 105 probablemente representa un marcador genético que es heterocigoto (por lo tanto, AB).

Por lo tanto, en la Figura 2, se pueden distinguir tres grupos de puntos de datos variantes (11A y 11B, 12A y 12B, y 13):

- puntos de datos variantes 11A y 11B que son homocigotos en el ADN materno y fetal (AAaaa, BBbbb);
- puntos de datos variantes 12A y 12B que son homocigotos en el ADN materno y heterocigotos en el ADN fetal (AAab, BBab). Nótese que en estos casos el ADN fetal contiene un alelo que se heredó del padre y que no está presente en el ADN materno. En otras palabras, para un feto macho este grupo de puntos de datos variantes no estará presente para el cromosoma X, véase la Figura 1;
- puntos de datos variantes 13 que son heterocigotos en el ADN materno (ABaa, ABab, ABbb).

De nuevo, nótese que esta clasificación se hace solo para ayudar al entendimiento de una realización ejemplar del método de la invención, y que la realización ejemplar no se basa exactamente en esta distinción.

Se señala que los múltiples puntos de datos variantes pueden tener la misma (o casi la misma) frecuencia alélica, especialmente cuando son parte del mismo grupo. Esto significa que se han medido (casi) el mismo número de presencias de alelos para ellos, con respecto al número total de lecturas.

También se señala que, en las Figuras 1 y 2 los puntos de datos variantes con una cobertura de lectura total más alta (cerca del máximo del gráfico) generalmente tienen una frecuencia de alelos más precisa, simplemente porque hay más datos de mediciones. Esta propiedad se puede tener en cuenta cuando se determina una fiabilidad estadística para un punto de datos variantes determinado. Más en particular, en una realización ejemplar, la precisión del método se puede mejorar restringiendo el conjunto de amplicones basándose en la cobertura total C_{π} . Esto puede incluir la necesidad de una cobertura total mínima y/o una cobertura total máxima, retirando una fracción del número total de amplicones con la mayor y/o menor cobertura total. La razón es que los amplicones con mayor cobertura total tienen una frecuencia alélica más precisa, debido a que hay más datos de mediciones. Sin embargo, los amplicones con una cobertura total que sea mucho mayores que la media del conjunto tiene una influencia desproporcionada sobre la determinación de la fracción homocigota.

Se calcularon los siguientes parámetros a partir de los resultados de la medición obtenidos:

- el número total D de lecturas sobre todos los puntos de datos variantes en el al menos un cromosoma (esto puede ser en todos los cromosomas excluyendo el cromosoma X y el cromosoma Y); es decir, el número total de presencias alélicas medidas para la pluralidad de marcadores genéticos de el al menos un cromosoma;
- 5 - el número total D_{ho} de lecturas en el al menos un cromosoma, correspondiente a los puntos de datos variantes que son puramente homocigotos, es decir, una frecuencia alélica que es de 0 o 1, dentro de un margen de error predeterminado.

10 Para la muestra NP, se obtuvo el siguiente conjunto de valores por el método (véase la Figura 1 para los puntos de datos variantes):

$$D = 2000000 \text{ (dividido entre 4000 amplicones),}$$

$$D_{ho} = 1400000 \text{ (dividido entre un subconjunto de 2900 amplicones homocigotos).}$$

15 Para la muestra P, se identificó el siguiente conjunto de valores por el método (véase la Figura 2 para los puntos de datos variantes):

$$D = 2000000 \text{ (dividido entre 4000 amplicones),}$$

$$D_{ho} = 1100000 \text{ (dividido entre un subconjunto de 2200 amplicones puramente homocigotos).}$$

20 En la realización ejemplar, para la identificación del conjunto de amplicones puramente homocigotos para ambas muestras P y NP, se utilizaron los límites de frecuencia alélica de 0,01 y 0,99.

25 A continuación, se calcula una fracción homocigota F_{ho} de lecturas para la pluralidad de marcadores genéticos, que están asociados con los marcadores genéticos puramente homocigotos:

$$F_{ho} = D_{ho} / D.$$

30 Para la muestra NP la fracción homocigota $F_{ho} = 0,70$ y para la muestra P $F_{ho} = 0,55$. Está claro que la fracción homocigota F_{ho} es mayor para la muestra NP que para la muestra P.

35 Se espera que el valor de F_{ho} es mayor para las muestras de mujeres no gestantes que para las muestras de mujeres gestantes. La razón es que la presencia del ADN fetal desplazará una fracción de amplicones del estado puramente homocigoto debido a la presencia de un SNP introducido por el ADN del padre. Esta desviación se mide sin la detección actual de los SNP paternos.

Estimación

40 Para automatizar la estimación de si una hembra está gestante o no, se puede establecer un valor de umbral empírico F_T , que distingue entre una condición no gestante y gestante, optimizando la discriminación utilizando datos de referencia con una condición de gestación conocida.

45 En una realización ejemplar, los cálculos anteriores se repiten para las muestras con un estado de gestación determinado que es conocido para establecer un valor de umbral F_T . En la figura 3 se da un histograma de la fracción homocigota F_{ho} para un conjunto de 175 muestras (117 muestras gestantes y 58 muestras no gestantes). Basándose en el ejemplo de la Figura 3, un valor de $F_T = 0,635$ es un buen valor de umbral empírico.

La estimación se puede llevar a cabo entonces de la siguiente manera:

50 si $F_{ho} < F_T,$

se determina que la muestra es gestante;

55 si $F_{ho} \geq F_T,$

se determina que la muestra es no gestante.

60 Basándose en este valor de umbral, la muestra NP se clasifica como no gestante y la muestra P se clasifica como gestante.

65 El valor de F_{ho} para una muestra no solo está influenciado por el estado de gestación sino también por el nivel total de heterocigosis del ADN de la hembra, que a su vez depende de la etnia del individuo. Esto se puede corregir utilizando la siguiente estrategia:

- determinar el número de lecturas total D_{he} sobre los cromosomas no sexuales que cubren amplicones con una

frecuencia alélica heterocigota. Los amplicones heterocigotos se pueden determinar fijando los umbrales de F_i , por ejemplo, entre 0,3 y 0,7;

- calcular un estimador para la heterocigosis, por ejemplo, una fracción heterocigota $F_{he} = D_{he}/D$.

5 En una realización ejemplar, la precisión se puede mejorar haciendo la F_T dependiente del nivel de heterocigosis total del ADN de la hembra, estimado por F_{he} . Esta relación $F_T (F_{he})$ puede optimizarse utilizando los datos de referencia para las muestras gestantes y no gestantes.

La estimación de gestación se puede llevar a cabo entonces de la siguiente manera:

10

si

$$F_{ho} < F_T (F_{he}),$$

15

se determina que la muestra es gestante;

si

$$F_{ho} \geq F_T (F_{he}),$$

20

se determina que la muestra es no gestante.

Los inventores han observado que el nivel total de heterocigosis en el ADN materno también tiene influencia sobre el valor de F_{ho} . Esto se puede ver en la Figura 4, que muestra la fracción homocigota F_{ho} en el eje y, y la fracción heterocigota F_{he} en el eje x para las 175 muestras. Además, en la Figura 4 se puede ver que para las 117 muestras gestantes (véase la referencia numérica 42) la fracción homocigota F_{ho} es menor que la $F_{Te} = 0,635$ (véase la línea horizontal 40), y que para las 58 muestras no gestantes (véase la referencia numérica 41) la fracción homocigota F_{ho} es mayor que la $F_{Te} = 0,635$.

25

Los valores de la fracción heterocigota F_{he} se pueden utilizar para estimar un valor de umbral $F_T (F_{he})$, véase la línea 50 en la Figura 5. Las muestras con un nivel total de heterocigosis más bajo tenderán a tener un valor de $F_T (F_{he})$ mayor y viceversa. Aunque un umbral fijo ($F_{Te} = 0,635$) separa perfectamente los grupos de muestras gestantes y no gestantes, un umbral $F_T (F_{he})$ puede mejorar la fiabilidad del método. Basándose en el conjunto de datos, se puede establecer un umbral empírico $F_T (F_{he}) = 0,86 - 0,75 * F_{he}$. Este umbral $F_T (F_{he})$ se muestra en la Figura 5, véase la línea 50.

30

35

La Figura 6 ilustra una realización de un sistema de la invención para la estimación de si una hembra está gestante o no. El sistema comprende un dispositivo de medición 1001, un módulo de determinación 1002, y un módulo de estimación 1003. El dispositivo de medición 1001 se configura para la medición de presencia alélica (D) para una pluralidad de marcadores genéticos de al menos un cromosoma, diferentes del cromosoma X e Y, en una muestra de ADN libre de células de una hembra potencialmente gestante; representando la presencia de cada alelo la presencia de un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal. El módulo de determinación 1002 se configura para la determinación, basándose en dicha presencia alélica medida, de una fracción homocigota (F_{ho}) de la misma que se asocia con marcadores genéticos puramente homocigotos. El módulo de estimación 103 se configura para la estimación de si la hembra está gestante, basándose en la determinación de dicha fracción. La estimación se puede hacer de acuerdo con una cualquiera de las realizaciones ejemplares descritas anteriormente.

40

45

Un experto en la técnica reconocerá fácilmente que las etapas de los distintos métodos descritos anteriormente se pueden llevar a cabo mediante computadoras programadas. En el presente documento, algunas realizaciones también tienen la intención de cubrir los dispositivos de almacenamiento de programas, por ejemplo, medios de almacenamiento de datos digitales, que pueden ser leídos por una máquina o computadora y que codifican programas de instrucciones ejecutables por una máquina o ejecutables por una computadora, donde dichas instrucciones llevan a cabo algunas o todas las etapas de dichos métodos descritos anteriormente. Los dispositivos de almacenamiento de programas pueden ser, por ejemplo, memorias digitales, medios de almacenamiento magnéticos tales como discos magnéticos y cintas magnéticas, discos duros, o medios de almacenamiento de datos digitales que se pueden leer ópticamente. También se describe en el presente documento computadoras programadas para llevar a cabo dichas etapas de los métodos descritos anteriormente.

50

55

Las funciones de los distintos elementos que se muestran en las figuras, incluyendo cualquiera de los bloques funcionales marcados como "módulos", puede proporcionarse mediante el uso de hardware dedicado, así como hardware capaz de ejecutar en softwares en asociación con el software adecuado. Cuando se proporcionan mediante un procesador, las funciones se pueden proporcionar mediante un único procesador dedicado, mediante un único procesador compartido, o mediante una pluralidad de procesadores individuales, algunos de los cuales se pueden compartir. Además, el uso explícito del término "módulo" no se debería considerar que se refiere exclusivamente al hardware capaz de ejecutar el software, y puede incluir implícitamente, sin limitación, hardware de procesador de señal digital (DSP), de procesador en red, un circuito integrado específico de la aplicación (ASIC),

60

65

ES 2 761 633 T3

matriz de regulación de campo programable (FPGA), una memoria solo de lectura (ROM) para el almacenamiento del software, una memoria de acceso aleatorio (RAM), y un almacenamiento no volátil. También se puede incluir otro hardware, convencional y/o a medida.

REIVINDICACIONES

1. Un método para la estimación de si una hembra está gestante, comprendiendo dicho método:

- 5 - la medición de presencias alélicas (D) para una pluralidad de marcadores genéticos de al menos un cromosoma, diferente del cromosoma X e Y, en una muestra de ADN libre de células de una hembra potencialmente gestante; representando cada presencia alélica la presencia en un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal;
- 10 - basándose en dichas presencias alélicas medidas, la determinación de una fracción homocigota (F_{ho}) de las mismas que se asocian con marcadores genéticos puramente homocigotos; y
- la estimación de si la hembra está gestante basándose en dicha fracción, donde la estimación comprende:

- la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado; y
- 15 la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado, donde el primer y segundo valores predeterminados pueden ser el mismo o donde el segundo valor predeterminado puede ser mayor que el primer valor predeterminado.

2. El método de la reivindicación 1, donde la determinación de la fracción homocigota comprende:

- 20 - basándose en dichas presencias alélicas medidas para la pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha pluralidad;
- la determinación como fracción homocigota la fracción de dichas presencias alélicas medidas para la cual la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado.

3. El método de la reivindicación 1 o 2 que comprende adicionalmente:

- 25 - basándose en dichas presencias alélicas medidas, la determinación de una fracción heterocigota (F_{he}) de las mismas que se asocia con marcadores genéticos heterocigotos; y

en el que la estimación comprende:

- la estimación de si la hembra está gestante basándose en la fracción homocigota y dicha fracción heterocigota.

4. El método de la reivindicación 3 que comprende adicionalmente:

- el cálculo de un valor de umbral en función de la fracción heterocigota;
- donde la estimación comprende:
- 40 la estimación de que la hembra está gestante si la fracción homocigota está por debajo del valor de umbral calculado; y
- la estimación de que la hembra no está gestante si la fracción homocigota está por encima del valor de umbral calculado.

5. El método de la reivindicación 1, donde el primer y segundo valores predeterminados son un valor constante que se ha determinado empíricamente.

6. El método de una cualquiera de las reivindicaciones previas, donde la determinación de una fracción homocigota (F_{ho}) que está asociada con marcadores genéticos puramente homocigotos comprende:

- 50 basándose en dichas presencias alélicas medidas para la pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha pluralidad de marcadores genéticos; y determinar los recuentos de lectura en función de la frecuencia alélica para dicha pluralidad de marcadores genéticos;
- seleccionar una parte de los recuentos de lectura, preferentemente una parte en la que se han eliminado los recuentos de lectura con los valores más altos y/o más bajos; y
- 55 determinar una fracción homocigota (F_{ho}) de dicha parte seleccionada, que se asocia con marcadores genéticos puramente homocigotos.

7. El método de una cualquiera de las reivindicaciones previas, donde:

- 60 (i) las etapas de medición y determinación se llevan a cabo para un lote que comprende una pluralidad de muestras de la hembra; donde para cada muestra del lote, se calcula la fracción homocigota; y donde la estimación se basa adicionalmente en dicha fracción homocigota de cada muestra;
- (ii) la muestra es sangre, plasma, orina, líquido cefalorraquídeo, suero, saliva o es un fluido del lavado transcervical materno; y/o
- 65 (iii) dicha etapa de medición comprende al menos una de las siguientes: reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa, amplificación basada en la secuencia de ácidos nucleicos (NASBA), y

métodos de ADN ramificado; y preferentemente PCR.

8. El método de una cualquiera de las reivindicaciones previas, que comprenden adicionalmente la provisión de la muestra *in vitro*.

5 9. Un sistema para la estimación de si una hembra está gestante, comprendiendo dicho sistema:

10 - un dispositivo de medición configurado para la medición de presencias alélicas (D) para una pluralidad de marcadores genéticos de al menos un cromosoma, diferente del cromosoma X e Y, en una muestra de ADN libre de células de una hembra potencialmente gestante; representando la presencia de cada alelo la presencia de un marcador genético de al menos uno de: un alelo de referencia de origen maternal o fetal, y un alelo alternativo de origen maternal o fetal;

15 - un módulo de determinación configurado para la determinación, basándose en dichas presencias alélicas medidas, de una fracción homocigota (F_{ho}) de las mismas que se asocia con marcadores genéticos puramente homocigotos;

- un módulo de estimación configurado para la estimación de si la hembra está gestante basándose en dicha fracción, donde el módulo de estimación se configura para:

20 la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un primer valor predeterminado; y
la estimación de que la hembra no está gestante si la fracción homocigota está por encima de un segundo valor predeterminado, donde el primer y segundo valores predeterminados pueden ser el mismo o donde el segundo valor predeterminado puede ser mayor que el primer valor predeterminado.

25 10. El Sistema de la reivindicación 9, donde el módulo de determinación se configura para:

- basándose en dichas presencias alélicas medidas, calcular un número correspondiente de frecuencias alélicas para dicha pluralidad de marcadores genéticos;

30 - la determinación como fracción homocigota la fracción de dichas presencias alélicas medidas para la cual la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado.

11. El Sistema de la reivindicación 9 o 10, donde el módulo de determinación se configura adicionalmente para:

35 - basándose en dichas presencias alélicas medidas, la determinación de una fracción heterocigota (F_{he}) de las mismas que se asocia con marcadores genéticos heterocigotos; y

donde el módulo de estimación se configura adicionalmente para:

40 - la estimación de si la hembra está gestante basándose en la fracción homocigota y dicha fracción heterocigota.

12. El sistema de la reivindicación 11, donde el módulo de determinación se configura adicionalmente para: el cálculo de un valor de umbral en función de la fracción heterocigota;

donde el módulo de estimación se configura adicionalmente para:

45 la estimación de que la hembra está gestante si la fracción homocigota está por debajo de un valor de umbral calculado; y
la estimación de que la hembra no está gestante si la fracción homocigota está por encima del valor de umbral calculado.

50 13. El Sistema de una cualquiera de las reivindicaciones 9-12, donde:

(i) el dispositivo de medición y el módulo de determinación se configuran para llevar a cabo la medición y determinación de un lote que comprende una pluralidad de muestras, donde para cada muestra del lote, se calcula la fracción homocigota, y donde el módulo de estimación se configura para la estimación de si la hembra está gestante basándose en dicha fracción homocigota para cada muestra; y/o

55 (ii) el dispositivo de medición se configura para medir las presencias alélicas utilizando al menos uno de las siguientes: reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa, amplificación basada en la secuencia de ácidos nucleicos (NASBA), y métodos de ADN ramificado; y preferentemente PCR.

60 14. Un programa de computadora que comprende instrucciones ejecutables por computadora para llevar a cabo, cuando se ejecuta el programa en una computadora, al menos la etapa de estimación del método de una cualquiera de las reivindicaciones 1-8.

65 15. Un medio de almacenamiento de datos digital que codifica un programa ejecutable por computadora de instrucciones para llevar a cabo al menos la etapa de estimación del método de una cualquiera de las reivindicaciones 1-8.

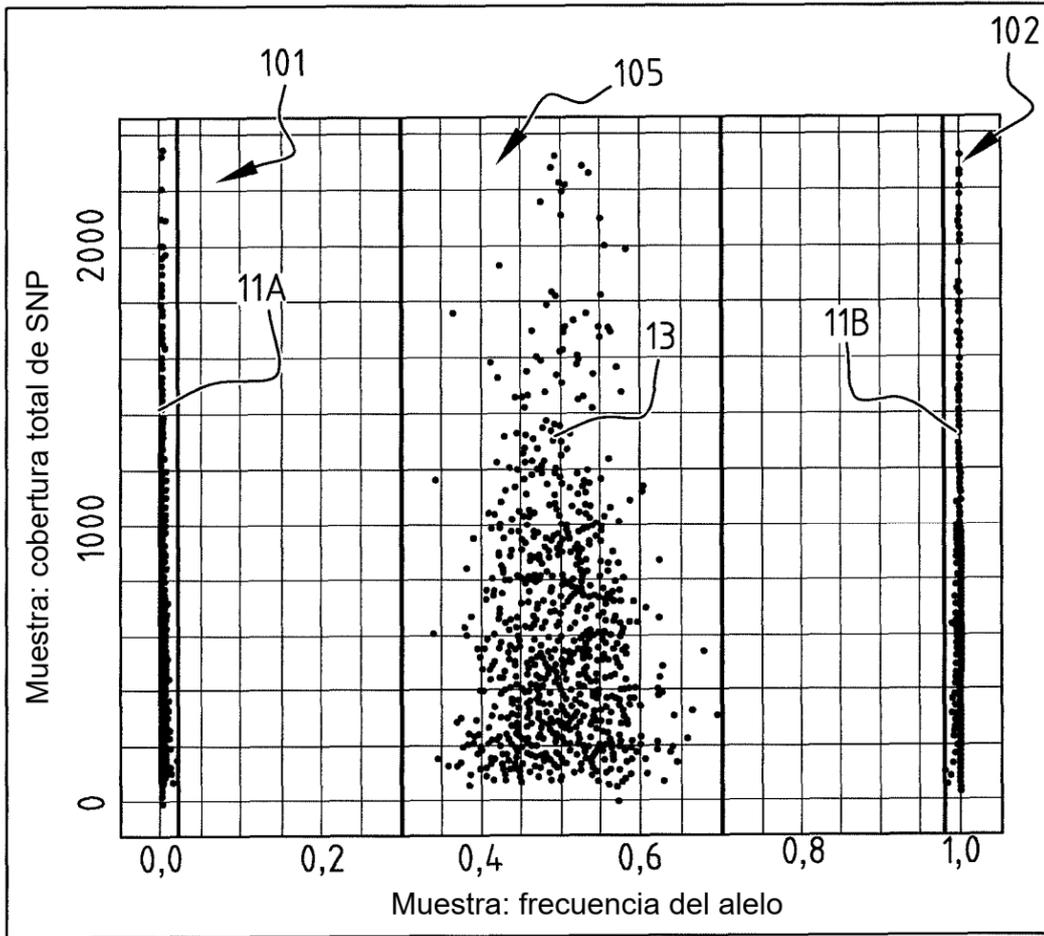


FIG. 1

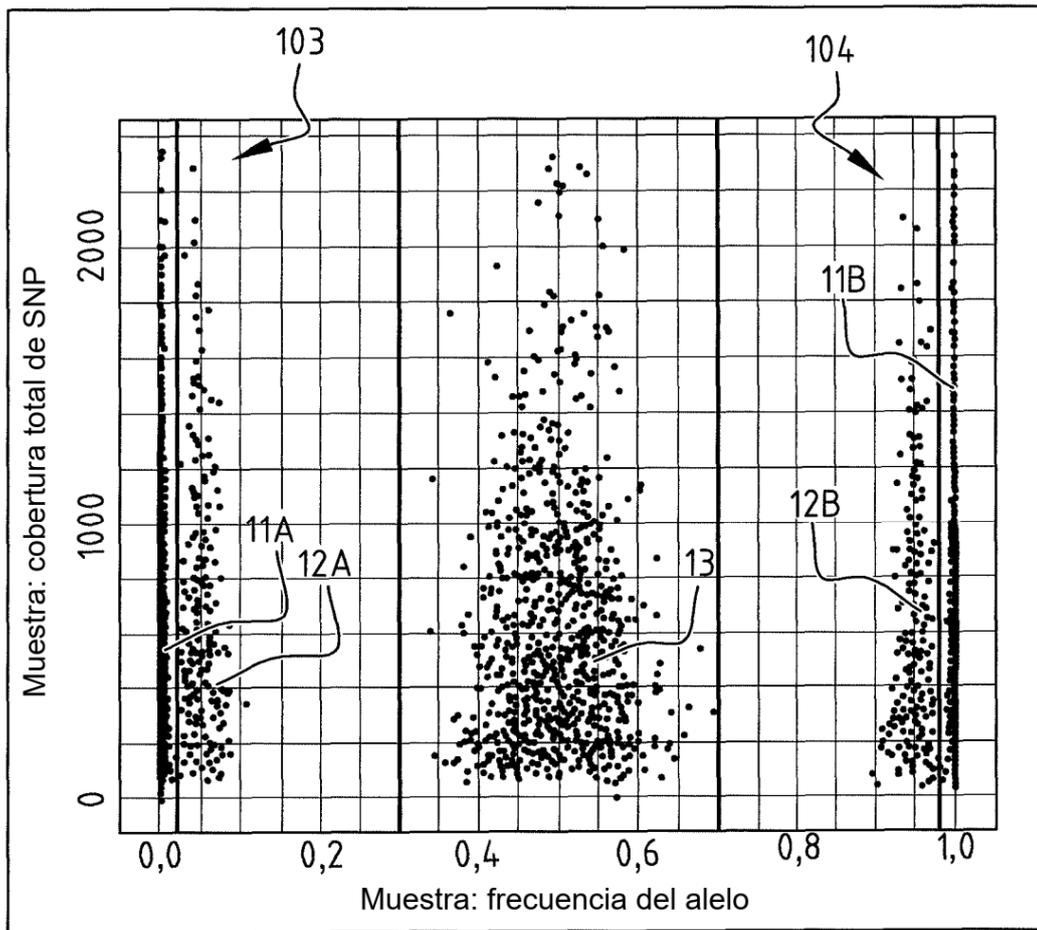


FIG. 2

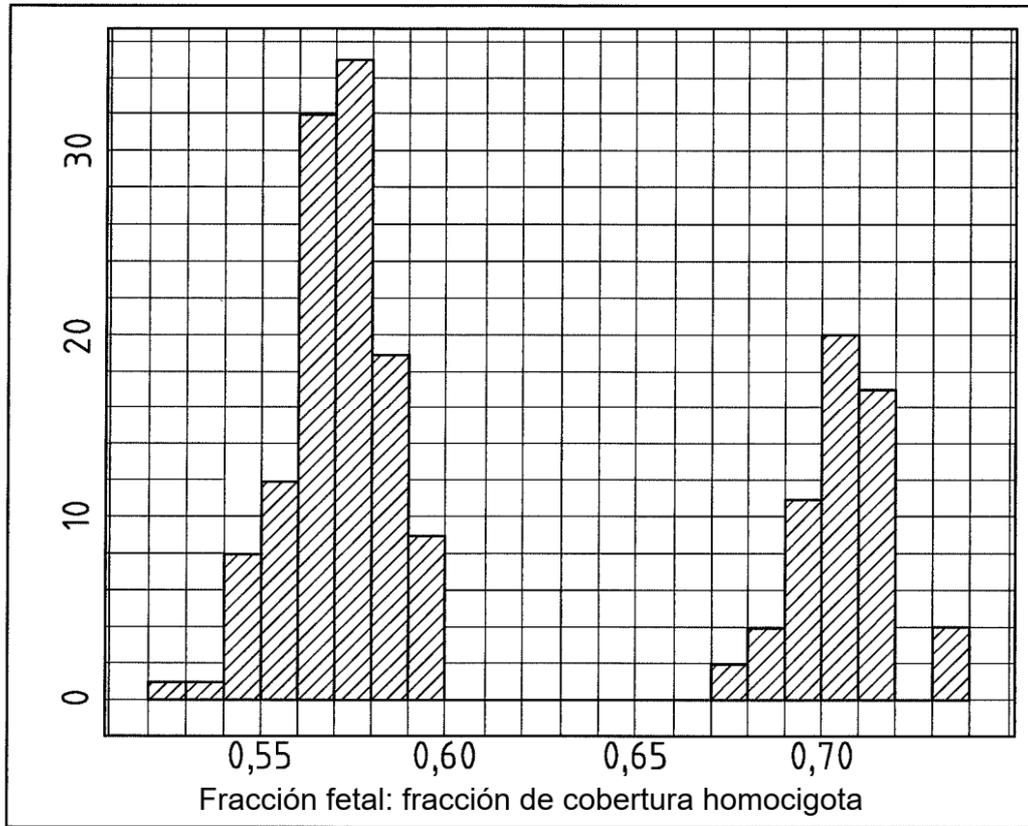


FIG. 3

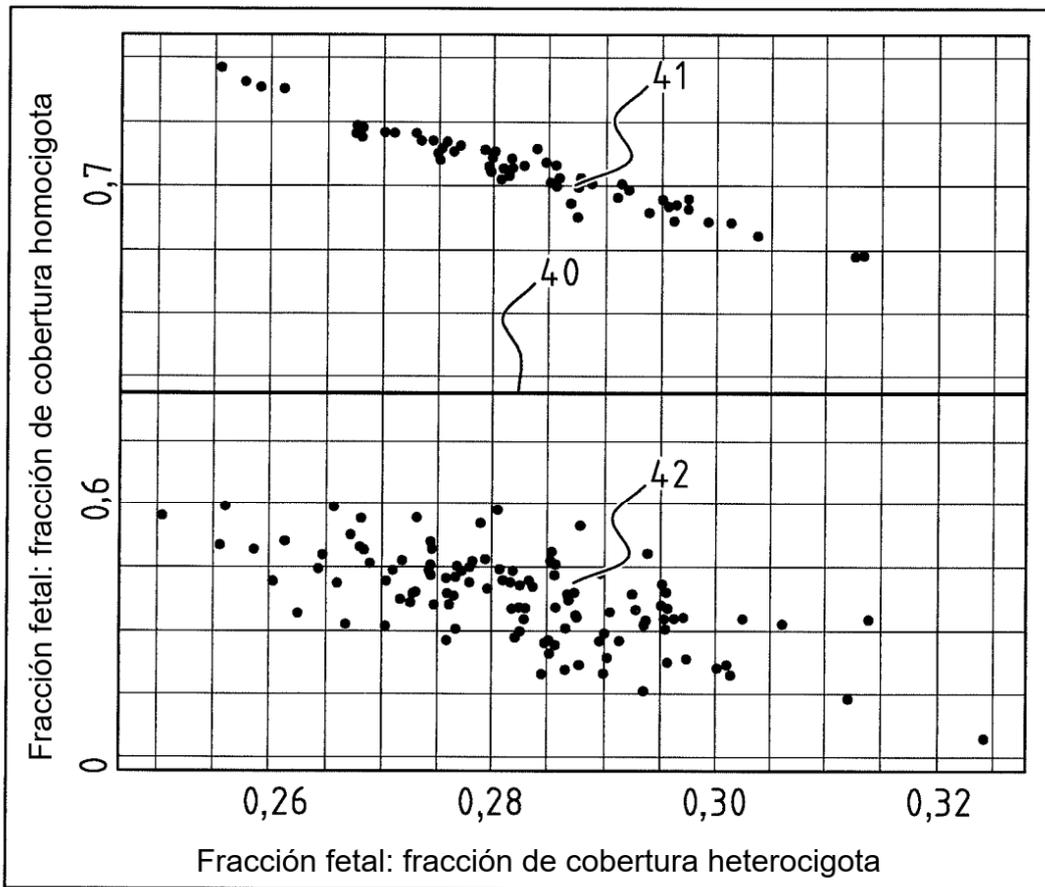


FIG. 4

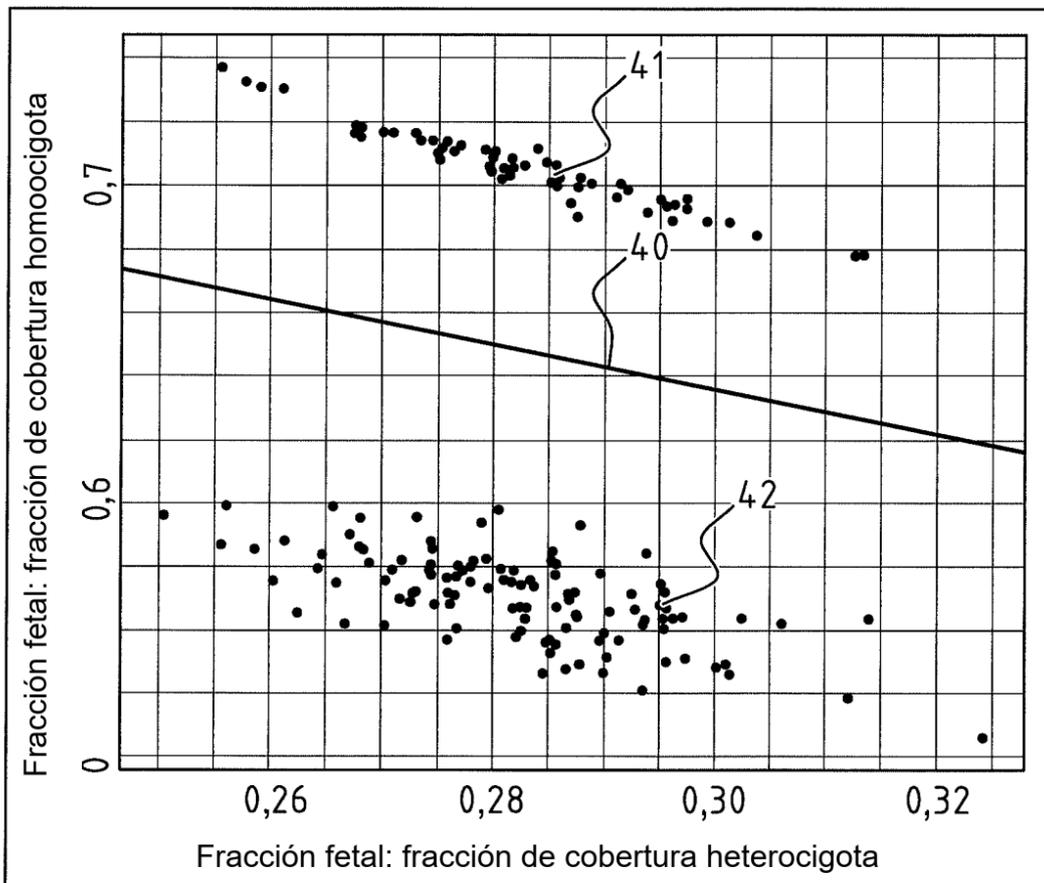


FIG. 5

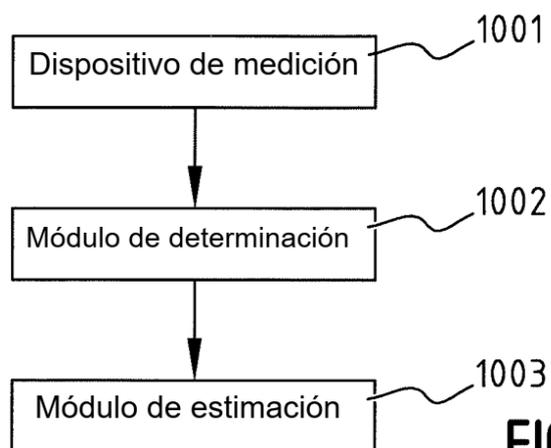


FIG. 6