

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 764 501**

51 Int. Cl.:

**C12Q 1/6809** (2008.01)

**C12Q 1/6879** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.07.2016 PCT/EP2016/066653**

87 Fecha y número de publicación internacional: **26.01.2017 WO17012954**

96 Fecha de presentación y número de la solicitud europea: **13.07.2016 E 16738770 (3)**

97 Fecha y número de publicación de la concesión europea: **18.12.2019 EP 3326094**

54 Título: **Método y sistema para estimar el género del feto de una hembra embarazada**

30 Prioridad:

**17.07.2015 BE 201505464**

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

**03.06.2020**

73 Titular/es:

**MULTIPLICOM NV (100.0%)**

**Galileilaan 18**

**2845 Niel, BE**

72 Inventor/es:

**VAUTERIN, PAUL;**

**VYVERMAN, MICHAËL y**

**DE SCHRIJVER, JOACHIM**

74 Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

ES 2 764 501 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 estimar el género del feto de una hembra embarazada

### 5 Campo de la invención

El campo de la invención se refiere a la estimación del género de un feto de una hembra embarazada. Las realizaciones específicas de las invenciones se refieren a métodos, sistemas, programas informáticos y productos de programas informáticos para estimar el género del feto de una hembra embarazada.

10

### Antecedentes

El documento WO 2013/057568 en nombre del solicitante divulga métodos de detección prenatal que utilizan técnicas no invasivas. En particular, se refiere al diagnóstico prenatal de una aneuploidía cromosómica fetal mediante la detección de ácidos nucleicos fetales y maternos en una muestra biológica materna. Más específicamente, el documento WO 2013/057568 emplea la PCR múltiple para amplificar fracciones seleccionadas de los respectivos cromosomas de los cromosomas maternos y fetales. Las respectivas cantidades de las regiones cromosómicas con sospecha de aneuploidía y los cromosomas de referencia se determinan a partir de un análisis de secuenciación masiva seguido de un análisis estadístico para detectar una aneuploidía en particular. El documento WO2011/091063 se refiere a métodos de detección de la aneuploidía o del sexo en un feto mediante el análisis del ADN extraído de la sangre de una hembra embarazada.

15

20

Aunque existen varios métodos de detección prenatal, se necesita un método preciso para estimar el género del feto de una hembra embarazada.

25

### Sumario

El objetivo de las realizaciones de la invención es proporcionar un método, sistema y programa informático para estimar el género del feto de una hembra embarazada.

30

De acuerdo con un primer aspecto de la invención, se proporciona un método para estimar el género de un feto de una hembra embarazada. El método comprende el paso de medir las presencias alélicas para una primera pluralidad de marcadores genéticos del cromosoma X y las presencias alélicas para una segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia, diferente de los cromosomas X e Y, en una muestra de ADN libre de células de una hembra embarazada. Cada presencia alélica representa la presencia en un marcador genético de al menos uno de: un alelo de referencia de origen materno o fetal y un alelo alternativo de origen materno o fetal. En aras de la exhaustividad, debe tenerse en cuenta que esto no implica que se pueda hacer una distinción entre un alelo de origen materno y un alelo de origen fetal, sino simplemente que puede medirse un alelo de cualquier origen. El método comprende además: en base a las presencias alélicas medidas para dicha primera pluralidad, determinar una primera fracción ( $F_X$ ) de la misma que se asocia a marcadores genéticos puramente homocigotos, en base a dichas presencias alélicas medidas para dicha segunda pluralidad, determinar una segunda fracción ( $F_R$ ) de la misma que se asocia a marcadores genéticos puramente homocigóticos y estimar el género de dicho feto a partir de dicha primera y segunda fracción. El término "marcador genético puramente homocigoto" se refiere a un marcador genético que es a la vez homocigoto en el ADN de origen materno y en el ADN de origen fetal.

35

40

45

Las realizaciones de la invención se basan, entre otras cosas, en el conocimiento inventivo de que la primera fracción esperada en el cromosoma X debe ser mayor en el caso de un feto del género masculino en comparación con un feto del género femenino. Para un feto del género femenino, se espera que esta primera fracción sea idéntica a la segunda fracción obtenida de los cromosomas no X/Y. Para un feto del género masculino, se espera que la primera fracción sea más alta que la segunda fracción obtenida de al menos un cromosoma de referencia diferente de los cromosomas X e Y. El género se puede estimar con precisión utilizando esta primera y segunda fracción.

50

Estimar el género de un feto de acuerdo con las realizaciones de la invención permite verificar el género como se estima en métodos de estimación conocidos. Esta verificación mejorará la precisión de una estimación del género de un feto en comparación con los métodos de estimación conocidos.

55

En el contexto de esta memoria descriptiva, un "marcador genético" es una posición en el genoma que se sabe que adquiere varios estados posibles en los individuos de una población.

60

En una realización preferida, la determinación de la primera fracción comprende: en base a dicha medida de las presencias alélicas para la primera pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha primera pluralidad; y determinar como la primera fracción a la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado. Un experto en la materia entenderá que la tasa de error de un dispositivo de secuenciación es un estándar industrial, y será de sobra conocido para un fabricante o instrumento específico, pero en una realización específica, el margen de error predeterminado es  $<0,02$ . En particular, puede ser  $<0,015$ ,  $<0,01$  o  $<0,005$ . De igual

65

manera la determinación de la segunda fracción puede comprender: en base a dicha medida de la presencia alélica para la segunda pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha segunda pluralidad; y determinar como segunda fracción a la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado. Un experto en la materia entenderá que la tasa de error de un dispositivo de secuenciación es un estándar industrial, y será de sobra conocido para un fabricante o instrumento específico, pero en una realización específica, el margen de error predeterminado es  $<0,02$ . En particular, puede ser  $<0,015$ ,  $<0,01$  o  $<0,005$ . Las presencias alélicas se representan típicamente como recuentos de lecturas en función de la frecuencia alélica, y la primera/segunda fracción corresponde a los recuentos de lecturas de los puntos de datos de las variables concentrados alrededor de la frecuencia alélica 0 y alrededor de la frecuencia alélica 1.

En una realización preferida, la estimación comprende estimar un primer estimador de género utilizando una proporción entre la primera fracción ( $F_X$ ) y la segunda fracción ( $F_R$ ). El valor umbral para determinar los niveles umbral apropiados para la proporción para estimar un feto de género masculino o femenino debe calcularse empíricamente para cada realización. Un experto en la materia será capaz de calcular dichos umbrales utilizando los métodos descritos en el presente documento teniendo en cuenta los niveles de ruido y la heterocigosidad de las muestras como se describe en el presente documento. En realizaciones específicas, cuando la proporción entre  $F_X$  y  $F_R$  es inferior a 1,1, se estima que el feto es de género femenino. En particular, la proporción es inferior a 1,05, inferior a 1,03, inferior a 1,01 o inferior a 1,00. En una realización particular donde la proporción es aproximadamente 1, se estima que el feto es de género femenino. En realizaciones específicas, cuando la proporción entre  $F_X$  y  $F_R$  es superior a 1,1, se estima que el feto es de género masculino, en particular, la proporción es superior a 1,15, superior a 1,17, superior a 1,18, superior a 1,19 o superior a 1,20. En una realización particular, cuando la proporción es significativamente superior a 1, se estima que el feto es de género masculino.

En una realización preferida los pasos de medición y determinación se realizan para un lote que comprende una pluralidad de muestras; en donde para cada muestra del lote, se calculan la primera y la segunda fracción; y en donde el paso de estimación del género se basa además en dicha primera y segunda fracción de cada muestra. De esta manera, la precisión del resultado de la estimación puede mejorarse aún más.

En otra realización desarrollada, el paso de estimación de un género comprende: determinar un primer estimador de género ( $E_{H1}$ ) para cada muestra en base a dicha primera y segunda fracción para dicha muestra; para cada muestra para la cual el primer estimador de género indica que el género del feto es femenino, determinar una proporción ( $N_{Xi}$ ) entre las presencias alélicas medidas para la primera pluralidad y las presencias alélicas medidas para la segunda pluralidad de marcadores genéticos, y determinar los parámetros de distribución estadística de dicha pluralidad de proporciones para dichas muestras asociadas a un primer estimador de género que indica que el feto es de género femenino; y utilizar dicha distribución estadística para determinar un segundo estimador de género. De esta manera, la información estadística de un lote de muestras se utiliza para mejorar la estimación de género para una única muestra del lote. En el caso de que el número de muestras del lote de análisis para el cual el primer estimador de género indica que el género del feto es femenino sea demasiado pequeño para obtener una estimación fiable de la distribución, este conjunto puede aumentarse utilizando muestras para las cuales el primer estimador de género indica que el género del feto es masculino, utilizando un factor de corrección.

El paso de estimación de un género puede entonces comprender de manera adicional: determinar un estimador de género combinado utilizando dicho primer y segundo estimador de género.

De acuerdo con otro aspecto de la invención, se proporciona un sistema para estimar el género del feto de una hembra embarazada. El sistema comprende un dispositivo de medición, un módulo de determinación y un módulo de estimación. El dispositivo de medición está configurado para medir las presencias alélicas ( $D_X$ ) para una primera pluralidad de marcadores genéticos del cromosoma X y las presencias alélicas ( $D_R$ ) para una segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia, diferente de los cromosomas X e Y, en una muestra de ADN libre de células de una hembra embarazada. Puede tratarse de cualquier dispositivo de medición disponible en el mercado que sea adecuado para realizar tales mediciones. El módulo de determinación está configurado para determinar una primera fracción ( $F_X$ ) de dichas presencias alélicas medidas para dicha primera pluralidad que se asocia a marcadores genéticos puramente homocigotos, y una segunda fracción ( $F_R$ ) de dichas presencias alélicas medidas para dicha segunda pluralidad que se asocia a marcadores genéticos puramente homocigotos. El módulo de estimación está configurado para estimar el género de dicho feto a partir de dicha primera y segunda fracción. En una realización característica, el módulo de determinación y el módulo de estimación se implementan como software.

En una realización preferida el módulo de determinación se configura para, en base a dicha medida de las presencias alélicas para la primera pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha primera pluralidad; y para determinar como la primera fracción la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado. De manera similar, el módulo de determinación puede configurarse para, en base a dicha medida de la presencia alélica para la segunda pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha segunda pluralidad; y determinar como segunda fracción a la fracción de dichas

presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado. Un experto en la materia entenderá que la tasa de error de un dispositivo de secuenciación es un estándar industrial, y será de sobra conocido para un fabricante o instrumento específico, pero en una realización específica, el margen de error predeterminado es  $<0,02$ . En particular, puede ser  $<0,015$ ,  $<0,01$  o  $<0,005$ .

5 En una realización preferida, el módulo de estimación se configura para estimar un primer estimador de género utilizando una proporción entre la primera fracción ( $F_X$ ) y la segunda fracción ( $F_R$ ). El valor umbral para determinar los niveles umbral apropiados para la proporción para estimar un feto de género masculino o femenino debe calcularse empíricamente para cada realización. Un experto en la materia será capaz de calcular dichos umbrales  
10 utilizando los métodos descritos en el presente documento teniendo en cuenta los niveles de ruido y la heterocigosidad de las muestras como se describe en el presente documento. En realizaciones específicas, cuando la proporción entre  $F_X$  y  $F_R$  es inferior a 1,1, se estima que el feto es de género femenino, En particular, la proporción es inferior a 1,05, inferior a 1,03, inferior a 1,01, inferior a 1,00. En una realización particular donde la proporción es aproximadamente 1, se estima que el feto es de género femenino. En realizaciones específicas, cuando la  
15 proporción entre  $F_X$  y  $F_R$  es superior a 1,1, se estima que el feto es de género masculino, en particular, la proporción es superior a 1,15, superior a 1,17, superior a 1,18, superior a 1,19 o superior a 1,20. En una realización particular, cuando la proporción es significativamente superior a 1, se estima que el feto es de género masculino.

20 En una realización preferida, el dispositivo de medición y el módulo de determinación se configuran para realizar la medición y determinación para un lote que comprende una pluralidad de muestras, en donde para cada muestra del lote, se calculan la primera y la segunda fracción; y en donde el módulo de estimación se configura para estimar el género en base a dicha primera y segunda fracción de cada muestra.

25 En una realización preferida, el módulo de estimación se configura para estimar el género de una manera tal y como se describe en cualquiera de las realizaciones ejemplares del método.

30 En una realización preferida, el módulo de medición se configura para medir las presencias alélicas utilizando al menos uno de los 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.

En realizaciones ejemplares de la invención, la medición de las presencias alélicas puede comprender la medición de las presencias alélicas de SNP y/o la medición de las presencias alélicas para inserciones y/o deleciones cortas.

35 Las realizaciones preferidas del método y el sistema de la invención se divulgan en las reivindicaciones dependientes adjuntas.

40 De acuerdo con un aspecto adicional de la invención, se proporciona un programa de ordenador que comprende instrucciones ejecutables por ordenador para su realización, cuando el programa se ejecuta en un ordenador, uno o más pasos y, en particular, el paso de estimación de las realizaciones del método descrito anteriormente. De acuerdo con un aspecto adicional de la invención, se proporciona un dispositivo informatizado u otro dispositivo de hardware programado para realizar uno o más pasos y, en particular, el paso de estimación de una cualquiera de las realizaciones del método divulgado anteriormente. De acuerdo con otro aspecto, se proporciona un dispositivo de almacenamiento de datos que codifica un programa legible por una máquina y ejecutable por una máquina para  
45 realizar uno o más pasos de una cualquiera de las realizaciones del método divulgado anteriormente. La referencia a las instrucciones/formas ejecutables por ordenador tiene que interpretarse de forma que comprenda tanto el código ejecutable por una máquina, el código que debe ser compilado para ser ejecutado, como el código que es interpretado en lugar de ejecutado *per se*.

## 50 Breve descripción de las figuras

Los dibujos adjuntos se utilizan para ilustrar las realizaciones ejemplares no limitativas preferidas actualmente de un método y de un sistema de la presente invención. Las ventajas anteriores y otras ventajas de las características y los objetivos de la invención se harán más evidentes y la invención se entenderá mejor a partir de la siguiente  
55 descripción detallada cuando se lea junto con los dibujos adjuntos, en los cuales:

la figura 1 es un gráfico que muestra la cobertura total de SNP frente a la frecuencia alélica en el cromosoma X para la muestra M;

60 la figura 2 es un gráfico que muestra la cobertura total de SNP frente a la frecuencia alélica en el cromosoma X para la muestra F;

la figura 3 es un gráfico que muestra, para un conjunto de muestras, el estimador de género  $E_{H1}$  en el eje X frente a la fracción fetal en el eje Y;

la figura 4 es un gráfico que representa, para el conjunto de muestras de la figura 3, la proporción  $N_{Xi}$  en el eje X frente a la fracción fetal en el eje Y;

65 la figura 5 es un gráfico que representa, para el conjunto de muestras de la figura 3, la proporción  $N_{Xi}$  para las muestras clasificadas como femeninas (de acuerdo con  $E_{H1}$ ), y  $N'_{Xi}$  para las muestras clasificadas como

masculinas (de acuerdo con  $E_{H1}$ ) en el eje X frente a la fracción fetal en el eje Y;

la figura 6 es un gráfico que representa, para el conjunto de muestras de la figura 3, el valor-z  $Z_{oi}$  en el eje X frente a la fracción fetal en el eje Y;

5 la figura 7 es un gráfico que representa, para el conjunto de muestras de la figura 3, el segundo estimador de género  $E_{H2}$  en el eje X frente a la fracción fetal en el eje Y;

la figura 8 es un gráfico que representa el primer estimador de género  $E_{H1}$  en el eje X frente al segundo estimador de género  $E_{H2}$  en el eje Y para el conjunto de muestras de la figura 3;

la figura 9 es un gráfico que representa, para el conjunto de muestras de la figura 3, el estimador de género combinado  $E_H$  en el eje X frente a la fracción fetal en el eje Y; y

10 la figura 10 es un dibujo esquemático de la realización de un sistema.

### Descripción detallada de realizaciones ejemplares

15 En una prueba prenatal no invasiva (NIPT), conocida en la técnica anterior, el ADN libre de células (ADNcf) en una muestra de suero o plasma materno de una hembra embarazada se secuencian para detectar la presencia de aneuploidías cromosómicas en el feto, tales como la trisomía del cromosoma 21. De acuerdo con las realizaciones ejemplares de la invención, se proporciona un método para estimar el género del feto.

20 En una realización característica, se obtiene una muestra de suero o plasma 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, puede preferirse utilizar volúmenes más grandes. La preparación del suero o del plasma a partir de la muestra de sangre materna puede llevarse a cabo utilizando técnicas estándar. Entre las técnicas adecuadas se encuentran la centrifugación y/o las técnicas basadas en matriz. En posibles realizaciones, se puede utilizar un método de enriquecimiento basado en secuencias en el suero o plasma maternos para enriquecer específicamente las  
25 muestras en secuencias de ácido nucleico fetal.

Las realizaciones del método de la invención pueden realizarse para una muestra que contiene ADN fetal en una concentración de fracción fetal de la cantidad total 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  
30 utilizar cualquier método de amplificación conocido por el experto, tal como un método de PCR.

En una realización preferida se utilizan datos de la prueba de Clarigo del solicitante que no implica la detección de alelos de SNP (o polimorfismo de un solo nucleótido), 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 embarazada. La prueba de Clarigo consiste en la secuenciación dirigida de una serie de regiones del genoma humano (en otras palabras, se dirige a marcadores genéticos específicos), utilizando los SNP conocidos (polimorfismos de un solo nucleótido) con una alta prevalencia poblacional (por ejemplo, mayor del 1%, preferentemente mayor del 10%) y dos posibles alelos (sc. un alelo de referencia, también conocido como REF; y un alelo alternativo, también conocido como ALT). Se pueden encontrar más detalles sobre la prueba de Clarigo en Internet en <http://www.multiplicom.com/product/clarigo>,  
35 y en el documento WO 2013/057568 en nombre del solicitante.

Ahora se discutirá en detalle una realización ejemplar de un método para estimar el género del feto de una hembra embarazada. En un primer paso de medición, se miden las presencias alélicas para una primera pluralidad ( $D_X$ ) de marcadores genéticos del cromosoma X y para una segunda pluralidad ( $D_R$ ) de marcadores genéticos de al menos un cromosoma de referencia, diferente del cromosoma X e Y, en una muestra de ADN libre de células de una hembra embarazada. Cada presencia alélica representa la presencia en un marcador genético de al menos uno de los siguientes: un alelo de referencia de origen materno o fetal y un alelo alternativo de origen materno o fetal. En un segundo paso de cálculo, en base a las presencias alélicas medidas para dicha primera pluralidad, se determina una primera fracción ( $F_X$ ) de la misma que se asocia a marcadores genéticos puramente homocigotos, y se basa en las  
45 presencias alélicas medidas para dicha segunda pluralidad, y se determina una segunda fracción ( $F_R$ ) de la misma que se asocia a marcadores genéticos puramente homocigotos. En un tercer paso se estima el género del feto en base a dicha primera y segunda fracción. En la realización ejemplar expuesta a continuación, el paso de estimación comprende el cálculo de un primer estimador de género para una muestra específica, el cálculo de un segundo estimador de género utilizando también datos de otras muestras y el cálculo de un estimador de género combinado  
50 para la muestra específica en base al primer y segundo estimador de género.

### Medición y determinación de fracciones

Una manera ventajosa de representar los resultados de la medición de presencias alélicas para un marcador genético, es asociar la siguiente información a un punto de datos de las variables para ese marcador genético. Un punto de datos de las variables (que es un punto de datos asociado con una serie de variables, tales como los alelos) se utiliza en esta memoria descriptiva como una representación conveniente para un marcador genético y, por lo tanto, representa el resultado de la medición de presencias alélicas en varios amplicones para marcadores genéticos. Un amplicón es un fragmento de ADN o ARN que es el producto (y/o fuente) de acontecimientos de amplificación o replicación. En otras palabras, un amplicón es un fragmento biofísico de material de replicación,  
65 diseñado para contener una posición conocida de SNP con alta prevalencia poblacional. Cada punto de datos de las

variables se asocia, por lo tanto, con un SNP conocido con alta prevalencia poblacional y con dos alelos posibles (sc. un alelo de referencia, también conocido como REF; y un alelo alternativo, también conocido como ALT). Para cada punto de datos de las variables  $A_i$ , se pueden determinar los siguientes números utilizando, por ejemplo, un flujo de trabajo (pipeline) de bioinformática estándar aplicado a los datos de secuenciación:

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

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

Por lo tanto, para un determinado marcador genético  $i$ , pueden medirse las presencias alélicas para el alelo REF, para el alelo ALT y para ambos alelos, midiendo los números de lecturas que contienen al alelo REF, al alelo ALT y tanto el alelo REF como el alelo ALT, respectivamente. En base a las presencias alélicas medidas, se calcula un número correspondiente de frecuencias alélicas para el número predeterminado de marcadores genéticos.

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

Para un punto de datos de una variable individual (es decir, para un marcador genético individual), imaginemos que 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 estados correspondientes para el ADN fetal. Esto significa que los puntos de datos de las variables pueden estar en los posibles estados enumerados en la Tabla 1:

Tabla 1

estado del punto de datos de las variables	Lecturas de ALT de la fracción esperadas ( $F_i$ )
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

A modo de una ilustración de las presencias alélicas medidas, los diagramas de dispersión ilustrados en las figuras 1 y 2 muestran los datos de una muestra M (feto del género masculino) y una muestra F (feto del género femenino) para las que se han medido las presencias alélicas para los marcadores genéticos del cromosoma X, en donde:

- Cada punto es un punto de datos de las variables (que representa el resultado de las mediciones realizadas en los amplicones) para los marcadores genéticos del cromosoma X.
- El eje horizontal muestra la fracción de las lecturas con el alelo ALT SNP (es decir, la frecuencia alélica  $F_i$ ).
- El eje vertical muestra la cobertura total de lectura  $C_{Ti}$ .

En la figura 1 se puede ver que los puntos de datos de las variables específicas están asociados con una frecuencia alélica específica. Por ejemplo, el punto de datos de las variables 101 se muestra a la izquierda y representa un marcador genético para el que se realizaron aproximadamente 3300 lecturas. Todas o casi todas las lecturas de este marcador genético han medido la presencia alélica para indicar que están presentes alelos REF (es decir, A), pero no que están presentes alelos ALT (es decir, B). Por lo tanto, el punto de datos de las variables 101 se ha representado a la izquierda en la figura 1, donde la frecuencia de los alelos es aproximadamente 0 y probablemente representa un marcador genético homocigoto AAaa. El punto de datos de las variables 102 se muestra a la derecha y representa un marcador genético para el cual se realizaron aproximadamente 2500 lecturas. Todas o casi todas las lecturas de este marcador genético han medido la presencia de alelos para indicar que no están presentes alelos REF (es decir, A), pero que están presentes alelos ALT (es decir, B). Por lo tanto, el punto de datos de las variables 102 se ha representado a la derecha en la figura 1, donde la frecuencia de los alelos es aproximadamente 1 y probablemente representa un marcador genético homocigoto BBbb.

En la figura 2, los puntos de datos de las variables 103 representan un marcador genético para el cual se realizaron aproximadamente 1800 lecturas. Las presencias alélicas medidas permiten calcular la frecuencia alélica, que es

- relativamente baja, pero no cero. Por lo tanto, el punto de datos de las variables 103 representa probablemente un marcador genético que es homocigoto por parte materna para el par de alelos de referencia pero tiene un par de alelos heterocigotos de origen fetal (por lo tanto es AAab), dado que la fracción de ADN materno presente en la muestra para el marcador genético es (mucho) mayor que la fracción de ADN fetal presente en la muestra para el
- 5 marcador genético. El punto de datos de las variables 104 se muestra a la derecha y representa un marcador genético para el cual se realizaron aproximadamente 2250 lecturas. Por el mismo razonamiento, el punto de datos de las variables 104 representa probablemente un marcador genético que es homocigoto por parte materna para el par de alelos alternativos pero tiene un par de alelos heterocigotos de origen fetal (por lo tanto es BBab).
- 10 En la figura 1, los puntos de datos de las variables 105 representan un marcador genético para el cual se realizaron aproximadamente 2750 lecturas. Las presencias alélicas medidas permiten calcular la frecuencia alélica, que se ha observado que es de aproximadamente 0,48. Por lo tanto, el punto de datos de las variables 105 representa probablemente un marcador genético que es heterocigoto por parte materna (por lo tanto AB).
- 15 Por lo tanto, pueden distinguirse tres grupos de puntos de datos de las variables (11A y 11B, 12A y 12B, y 13):
- los puntos de datos de las variables 11A y 11B que son homocigotos en el ADN materno y fetal (AAaa, BBbb);
  - los puntos de datos de las variables 12A y 12B que son homocigotos en el ADN materno y heterocigotos en el ADN fetal (AAab, BBab). Debe tenerse en cuenta que, en estos casos, el ADN fetal contiene un alelo que fue heredado del padre y que no está presente en el ADN materno. En otras palabras, para un feto del género masculino, este grupo de puntos de datos de las variables no estará presente para el cromosoma X, véase la figura 1;
  - los puntos de datos de las variables 13 que son heterocigotos en el ADN materno (ABaa, ABab, ABbb).
- 20
- 25 Se observa que múltiples puntos de datos de las variables pueden tener la misma (o casi la misma) frecuencia alélica, especialmente cuando son parte del mismo grupo. Esto significa que se ha medido (de manera muy próxima) el mismo número de presencias de alelos para ellos, en relación al número total de lecturas.
- También se observa que, en las figuras 1 y 2, Los puntos de datos de las variables con una mayor cobertura total de lectura (más cerca de la parte superior del gráfico) tienen una frecuencia alélica más precisa, simplemente porque hay más datos de medición. Esta propiedad puede tenerse en cuenta al determinar la fiabilidad estadística para un punto de datos de las variables determinado.
- 30

Estimador de género de la fracción homocigota (primer estimador de género)

- 35 Un primer estimador de género se calcula investigando la fracción de puntos de datos de las variables que son puramente homocigotos en la muestra para el cromosoma X y para al menos un cromosoma de referencia que no incluye el cromosoma X/Y. En una realización preferida se pueden utilizar todos los cromosomas de referencia.
- 40 A partir de los resultados obtenidos de la medición se calculan los siguientes parámetros:
- el número total de lecturas  $D_X$  en todos los puntos de datos de las variables del cromosoma X, es decir, el número total de presencias alélicas medidas para la primera pluralidad de marcadores genéticos del cromosoma X;
  - 45 - el número total de lecturas  $D_{Xh}$  en el cromosoma X, correspondiente a puntos de datos de las variables que son puramente homocigotos, es decir, la frecuencia alélica es 0 o 1, dentro de un margen de error predeterminado, es decir, los grupos 11A y 11B de las figuras 1 y 2;
  - el número total de lecturas  $D_R$  sobre todos los puntos de datos de las variables en al menos un cromosoma de referencia (estos pueden ser todos los cromosomas excepto el cromosoma X y el cromosoma Y); es decir, el
  - 50 número total de presencias alélicas medidas para la segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia;
  - el número total de lecturas  $D_{Rh}$  en al menos un cromosoma de referencia, correspondiente a puntos de datos de las variables que son puramente homocigotos, es decir, una frecuencia alélica que es 0 o 1, dentro de un margen de error predeterminado;
- 55 Para la muestra M, el siguiente conjunto de valores se obtuvo por el método (véanse en la figura 1 los puntos de datos de las variables en el cromosoma X):
- 60  $D_X = 227542$  (dividido entre 390 amplicones),  
 $D_{Xh} = 162151$  (dividido entre un subconjunto de 285 amplicones),  
 $D_R = 883695$  (dividido entre 1681 amplicones),  
 $D_{Rh} = 491752$  (dividido entre un subconjunto de 917 amplicones).
- 65 Para la muestra F, se identifica el siguiente conjunto de valores por el método (véase la figura 2 para los puntos de datos de las variables en el cromosoma X):

$D_R = 942355$  (dividido entre 1680 amplicones),  
 $D_{Rh} = 541801$  (dividido entre un subconjunto de 953 amplicones),  
 $D_X = 255394$  (dividido entre 391 amplicones),  
 $D_{Xh} = 143376$  (dividido entre un subconjunto de 228 amplicones).

5 A continuación, se calculan las siguientes fracciones:

- una primera fracción  $F_X$  de lecturas para la primera pluralidad de marcadores genéticos, que se asocia con marcadores genéticos puramente homocigotos:

10 
$$F_X = D_{Xh} / D_X,$$

- una segunda fracción  $F_R$  de lecturas para la segunda pluralidad de marcadores genéticos, que se asocia con marcadores genéticos puramente homocigotos:

15 
$$F_R = D_{Rh} / D_R.$$

Un primer estimador preliminar de género puede calcularse como:

20 
$$E_{H0} = F_X / F_R.$$

En el caso de un feto del género femenino, hay dos copias presentes del cromosoma X y de al menos un cromosoma de referencia. Por lo tanto, se espera que  $E_{H0}$  sea 1. En caso de un feto del género masculino, solo hay una copia del cromosoma X, pero dos copias de todos los cromosomas de referencia. Por lo tanto, se espera que  $E_{H0}$  sea mayor que 1. En otras palabras, el primer estimador preliminar de género permite estimar el género.

Para la realización ejemplar de la figura 1 (muestra M), las fracciones que son puramente homocigotas y el estimador preliminar del género  $E_{H0}$  pueden calcularse como:

30 
$$F_X = D_{Xh}/D_X \square 0,71,$$

$$F_R = D_{Rh}/D_R \square 0,55,$$

35 
$$E_{H0} = F_X/F_R \square 1,28.$$

Como  $E_{H0}$  está muy por encima de 1, se puede estimar que la muestra M es masculina.

Para la realización ejemplar de la figura 2 (muestra F), las fracciones que son puramente homocigotas y el estimador preliminar del género  $E_{H0}$  pueden calcularse como:

40 
$$F_X = D_{Xh}/D_X \square 0,56,$$

$$F_R = D_{Rh}/D_R \square 0,57,$$

45 
$$E_{H0} = F_X/F_R \square 0,97.$$

Como  $E_{H0}$  es aproximadamente 1, se puede estimar que la muestra F es femenina.

50 Esto puede entenderse más como se indica a continuación. Dadas dos copias del cromosoma X, alelo REF A y alelo ALT B, hay cuatro combinaciones posibles diferentes de esos alelos para la madre que se espera que aparezcan: AA, AB, BA, BB. En base al género del feto, las combinaciones anteriores pueden dividirse aún más debido a la presencia del ADN fetal:

55 1. Feto del género masculino (figura 1, muestra M): el contenido de ADN del cromosoma X consiste en dos copias de la madre y una sola copia del feto heredada de la madre, en una cantidad igual a la mitad de la fracción fetal. Se pueden distinguir las siguientes combinaciones alélicas:

- a. AAa, BBb: puramente homocigoto.
- b. ABa, ABb, BAa, BAb: heterocigoto.

60 2. Feto del género femenino (figura 2, muestra F): el contenido de ADN del cromosoma X consiste en dos copias de la madre y dos copias del feto, una de las cuales se hereda de la madre y otra se hereda del padre, con una cantidad igual a la fracción fetal. Se pueden distinguir los siguientes grupos:

- 65 a. AAaa, BBbb: puramente homocigoto.
- b. AAab, BBab: pares de alelos homocigotos de origen materno y pares de alelos heterocigotos de origen



fetal.

c. ABaa, ABab, ABba, ABbb, BAaa, BAab, BAbA, BAbb: pares de alelos heterocigotos de origen materno.

5 A partir de lo anterior, se deduce que la primera fracción esperada de los SNP puramente homocigotos en los cromosomas X debería ser mayor en el caso de un feto del género masculino que en el de un feto del género femenino. Para un feto del género femenino, se espera que esta primera fracción sea idéntica a la segunda fracción obtenida de los cromosomas no sexuales. Para un feto del género masculino, se espera que la primera fracción sea más alta que la segunda.

10 El primer estimador preliminar de género  $E_{H0}$  puede estandarizarse a un primer estimador de género  $E_{H1}$  que se espera que sea 0 para un feto del género femenino y 1 para un feto del género masculino:

$$E_{H1} = (E_{H0} - 1) / (E_{He} - 1),$$

15 con un valor predictivo  $E_{He}$  para  $E_{H0}$  en caso de un feto del género masculino.

Este valor predictivo  $E_{He}$  se puede obtener estimando el nivel de heterocigosidad de los SNP de la población utilizando la fracción de SNP que se observa como heterocigota en el ADN de la madre en la muestra. Esta estimación se puede hacer de la siguiente manera. Se introducen las siguientes definiciones:

20 - el número total  $D_{Rhet}$  de lecturas en al menos un cromosoma de referencia, correspondiente a los puntos de datos de las variables que son heterocigotos en el ADN materno. Para los puntos de datos de las variables heterocigotos, hay 8 posibilidades: ABaa, ABbb, ABab, ABba, BAaa, BAbb, BAab y BAbA.

25 La fracción de puntos de datos de las variables heterocigotos para al menos un cromosoma de referencia puede calcularse como:

$$F_{Rhet} = D_{Rhet} / D_R.$$

30 Esto puede entenderse más como se indica a continuación. Supongamos que en promedio, el alelo REF A puede encontrarse en una fracción  $F_A$  de todas las incidencias de un SNP y el alelo ALT en una fracción  $F_B = 1 - F_A$ . Si se hace una generalización que establece que todos los SNP de los puntos de datos de las variables comparten la misma fracción  $F_A$  del alelo REF, se pueden estimar los niveles esperados de homocigosidad y heterocigosidad. El nivel de homocigosidad, o de manera equivalente, el nivel de incidencia de las combinaciones AA, y BB se estima por:

$$F_A^2 + F_B^2 = F_A^2 + (1 - F_A)^2.$$

40 El nivel de heterocigosidad, o de manera equivalente, el nivel de incidencia de las combinaciones AB, BA es

$$F_{Rhet} = 2F_A F_B - 2F_A(1 - F_A).$$

45 Las combinaciones AAa, BBb tienen el mismo nivel de incidencia que las combinaciones AA, BB. La aparición de las combinaciones AAaa, y BBbb es  $F_A^3 + F_B^3 = F_A^3 + (1 - F_A)^3$ . La incidencia de las otras combinaciones puede estimarse de manera similar.

A partir de lo anterior, se deduce que un valor predictivo para  $E_{H0}$  en caso de un feto del género masculino puede estimarse como:

$$50 \quad E_{He} = F_A^2 + (1 - F_A)^2 / F_A^3 + (1 - F_A)^3.$$

El método estima el valor de  $F_A$  a partir del nivel de heterocigosidad observado en al menos un cromosoma de referencia  $F_{Rhet}$ , resolviendo la siguiente ecuación de segundo grado y tomando la raíz más grande:

$$2F_A(1 - F_A) - F_{Rhet} = 0.$$

55 Si no existe una raíz real,  $F_A$  se ajusta a 0,5.

Se puede ejecutar una primera clasificación de género basada en  $E_{H1}$ , utilizando un umbral fijo  $T_0$  ( $\leq 0,5$ ):

60  $E_{H1} < T_0$  : género femenino,

$E_{H1} \geq T_0$  and  $E_{H1} \leq 1 - T_0$  : desconocido,

$E_{H1} > 1 - T_0$  : género masculino.

## ES 2 764 501 T3

Para la muestra M, el primer estimador de género se puede calcular de la siguiente manera. Se determina el primer  $D_{Rhet}$ :  $D_{Rhet} = 259134$  (dividido entre 498 amplicones), y a continuación se calcula la fracción:

5 
$$F_{Rhet} = D_{Rhet}/D_R \square 0,29.$$

A partir de esta fracción se puede calcular el valor predictivo  $E_{He}$ :

10 
$$F_A \square 0,82,$$
  
$$E_{He} \square 1,25.$$

El primer estimador de género se convierte entonces en:

15 
$$E_{H1} = (E_{H0} - 1) / (E_{He} - 1) \square 1,07.$$

En base a este valor, una primera clasificación de la muestra M es masculina.

20 Para la muestra F, el primer estimador de género se puede calcular de la siguiente manera. Se determina el primer  $D_{Rhet}$ :  $D_{Rhet} = 250259$  (dividido entre 436 amplicones), y a continuación se calcula la fracción:

$$F_{Rhet} = D_{Rhet}/D_R \square 0,26.$$

25 A partir de esta fracción se puede calcular el valor predictivo  $E_{He}$ :

$$F_A \square 0,84,$$
  
$$E_{He} \square 1,22.$$

30 El primer estimador de género se convierte entonces en:

$$E_{H1} = (E_{H0} - 1) / (E_{He} - 1) \square -0,10.$$

35 En base a este valor, una primera clasificación de la muestra F es femenina.

La figura 3 ilustra un primer estimador de género  $E_{H1}$  (eje X) para un lote de muestras con diferentes fracciones fetales (eje Y). Las muestras (cada punto representa una muestra) a la izquierda de la línea vertical 301 se corresponden con un primer estimador que indica un feto del género femenino, y las muestras a la derecha de la línea vertical 301 se corresponden con un primer estimador que indica un feto del género masculino.

40 Estimador de género de cobertura (segundo estimador de género)

45 Un segundo estimador de género se obtiene comparando la cobertura del cromosoma X con la cobertura de un conjunto de al menos un cromosoma de referencia diferente de los cromosomas X e Y. Este conjunto puede ser el mismo conjunto que el utilizado para determinar el primer estimador de género, o un conjunto diferente.

Los siguientes parámetros se calculan a partir de los puntos de datos de las variables de una muestra  $i$  de un lote:

- 50 - el número total de lecturas  $D_{Xi}$  sobre todos los puntos de datos de las variables del cromosoma X, es decir, el número total de presencias alélicas medidas para la primera pluralidad de marcadores genéticos del cromosoma X; y
- 55 - el número total de lecturas  $D_{Ri}$  en todos los puntos de datos de las variables en al menos un cromosoma de referencia, es decir, el número total de presencias alélicas medidas para la segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia.

A continuación, se calcula una proporción  $N_{Xi}$  de las lecturas  $D_{Xi}$  y  $D_{Ri}$  de la siguiente manera para la muestra  $i$ :

$$N_{Xi} = D_{Xi} / D_{Ri}.$$

60 La proporción  $N_{Xi}$  puede calcularse para todas las muestras de un lote de análisis, por ejemplo, todas las muestras secuenciadas durante un solo ensayo. A continuación, se pueden seleccionar los valores  $N_{Xi}$  para las muestras que se determinaron como "femeninas" utilizando la primera clasificación de género, y se pueden calcular parámetros de distribución normal para esos valores, por ejemplo, un valor medio  $\mu$  y una desviación estándar  $\sigma$ .

65 En caso de que el número de muestras del lote de análisis que se clasificaron como "femeninas" sea demasiado pequeño para obtener una estimación fiable de la distribución, este conjunto puede aumentarse usando muestras

clasificadas como "masculinas", usando un valor corregido:

$$N'_{xi} = N_{xi} \times 1 / (1 - FF/2).$$

5 con FF la fracción fetal estimada de la muestra. La fracción fetal puede estimarse usando una realización del método divulgado en la solicitud de patente BE 2015/5460 en nombre del solicitante que se incluye en el presente documento por referencia. También se pueden utilizar otros métodos existentes para estimar la fracción fetal.

Utilizando esta distribución, para todas las muestras i del lote de análisis se puede calcular un valor-z  $Z_{Oi}$ :

10

$$Z_{Oi} = \frac{N_{xi} - \mu}{\sigma}.$$

Para cada muestra, determinar  $Z_{Mi}$ , el valor esperado del valor-z bajo el supuesto de que la muestra es "masculina". Este valor se basa en la fracción fetal estimada de la muestra:

$$Z_{Mi} = -\frac{\mu}{\sigma} \times \frac{FF}{2}.$$

15

Un segundo estimador de género se calcula como:

$$E_{H2} = Z_{Oi} / Z_{Mi}.$$

20 Se espera que este valor sea 0 para un feto del género femenino y 1 para un feto del género masculino.

Para las muestras M y F, la proporción  $N_{xi}$  puede calcularse usando las fórmulas anteriores:

$$\begin{aligned} N_{XM} &= 0,1225806583663057, \\ N_{XF} &= 0,12775398163696797. \end{aligned}$$

25

A continuación, se calculan todos los valores  $N_{xi}$  para todas las muestras del lote de análisis. En la figura 4 se muestran los valores  $N_{xi}$  de todas las muestras en función de su fracción fetal FF. Las muestras clasificadas como femeninas por el primer estimador de género  $E_{H1}$  se muestran en un contorno 401 y las muestras clasificadas preliminarmente como masculinas se muestran en un contorno 402.

30

Las muestras clasificadas como femeninas (es decir, los puntos en el contorno 401) se utilizan para determinar una distribución normal. Estas son las muestras para las cuales  $E_{H1} < T_0$ . Estas muestras pueden encontrarse a la izquierda de la línea de separación vertical 301 en la figura 3.

35

En el ejemplo de las figuras 3 y 4, hay suficientes muestras femeninas para estimar con precisión la distribución normal. Si este no fuera el caso, se pueden utilizar los valores  $N'_{xi}$  para las muestras masculinas para estimar la distribución. Para la muestra M (con FF = 5,8%), esto significaría:

$$40 \quad N'_{XM} = N_{XM} \times 1 / (1 - FF/2) = 0,12626348981879823.$$

La figura 5 muestra el efecto de la normalización de los valores  $N_{xi}$  a  $N'_{xi}$  para las muestras clasificadas como masculinas por el primer estimador de género.

45 En base a los valores  $N_{xi}$ , la distribución normal se calcula con:

$$\begin{aligned} \text{valor medio } \mu &= 0,12792346387904036, \text{ y} \\ \text{desviación estándar } \sigma &= 0,0011732807075874412. \end{aligned}$$

50 Utilizando esta distribución, se calcula un valor-z  $Z_{OM} = -4,55$  para la muestra M y  $Z_{OF} = -0,14$  para la muestra F. El valor-z de todas las muestras del lote de análisis se muestra en la figura 6.

El valor esperado del valor-z bajo la suposición de que la muestra es "masculina" para las dos muestras de ejemplo son  $Z_{MM} = -3,18$  y  $Z_{MF} = -2,84$ . Usando ambos valores-z, la segunda estimación de género de las muestras M y F son:

55

Muestra M:

$$E_{H2} = Z_{OM} / Z_{MM} = 1,43,$$

60

Muestra F:

$$E_{H2} = Z_{OF} / Z_{MF} = 0,05.$$

5 Los valores del segundo estimador de género para todas las muestras del lote de análisis se muestran en el eje x de la figura 7. Las muestras clasificadas como femeninas se muestran en un contorno 701 y las muestras clasificadas como masculinas se muestran en un contorno 702.

Estimador de género combinado

10 Tanto el primero como el segundo estimador de género pueden combinarse en un solo estimador conjunto.

En el caso de un feto del género femenino, se espera que el par de estimadores combinados tenga los siguientes valores:

15  $(E_{H1}, E_{H2}) = (0,0).$

En caso de un feto del género masculino, se espera que el par de estimadores combinados tenga los siguientes valores:

20  $(E_{H1}, E_{H2}) = (1,1).$

Para una muestra específica, se pueden calcular las distancias euclidianas entre los pares de valores observados del estimador y ambos pares de valores esperados:

$$D_F = \sqrt{E_{H1}^2 + E_{H2}^2}$$

$$D_M = \sqrt{(E_{H1} - 1)^2 + (E_{H2} - 1)^2}$$

25 A partir de esto, un único estimador combinado se calcula como

$$E_H = (D_M - D_F) / \sqrt{2}.$$

30 Se espera que este valor sea -1 para un feto del género masculino y +1 para un feto del género femenino.

Usando el estimador de género final, una estimación final puede realizarse de la siguiente manera, usando un umbral predeterminado T ( $\geq 0$ ):

35  $E_H < -T$  : género masculino,

$E_H \geq -T$  y  $E_H \leq +T$  : desconocido,

$E_H > +T$  : género femenino.

40 Los estimadores de género dieron como resultado valores (1,07, 1,43) para la muestra M y valores (-0,10, 0,05) para la muestra F. La figura 8 muestra los valores de ambos estimadores para todas las muestras del lote de análisis. Las muestras cuyos valores del estimador están centrados en torno a (0,0) se clasificarán como femeninas y se muestran en un contorno 801. Las muestras cuyos valores del estimador están centrados alrededor de (1,1) se clasificarán como masculinas y se muestran en un contorno 802.

45 Finalmente, se puede calcular un valor del estimador combinado para las muestras M y F:

Muestra M:

50  $E_H = -0,95,$

Muestra F:

55  $E_H = 0,94.$

En base a estos valores y al umbral T=0,3, la muestra M se clasifica como masculina y la muestra F como femenina. El valor de  $E_H$  para todas las muestras del lote de análisis de ejemplo se muestra en la figura 9. Las muestras clasificadas como femeninas se muestran en el lado izquierdo de la línea vertical 901, y las muestras clasificadas como masculinas se muestran en el lado derecho de la línea vertical 902. Cualquier punto de muestra entre las líneas 901 y 902 (ninguna estaba presente en el presente ejemplo) se clasificaría como desconocido. El valor de umbral T se determinó empíricamente utilizando un gran conjunto de datos.

- La figura 10 ilustra una realización de un sistema de la invención para estimar el género de un feto de una hembra embarazada. El sistema comprende un dispositivo de medición 1001, un módulo de determinación 1002 y un módulo de estimación 1003. El dispositivo de medición 1001 está configurado para medir las presencias alélicas ( $D_X$ ) para una primera pluralidad de marcadores genéticos del cromosoma X y las presencias alélicas ( $D_R$ ) para una segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia, diferente de los cromosomas X e Y, en una muestra de ADN libre de células de una hembra embarazada. El módulo de determinación 1002 está configurado para determinar una primera fracción ( $F_X$ ) de dicha presencia alélica medida para dicha primera pluralidad que se asocia a marcadores genéticos puramente homocigotos, y una segunda fracción ( $F_R$ ) de dicha presencia alélica medida para dicha segunda pluralidad que se asocia a marcadores genéticos puramente homocigotos. El módulo de estimación 1003 está configurado para estimar el género de dicho feto a partir de dicha primera y segunda fracción. La estimación puede realizarse de acuerdo con cualquiera de las realizaciones ejemplares descritas anteriormente.
- 15 Un experto en la materia reconocería fácilmente que los pasos de varios métodos descritos anteriormente pueden realizarse por ordenadores programados. En el presente documento, se pretende que algunas realizaciones abarquen dispositivos de almacenamiento de programas, por ejemplo, medios de almacenamiento de datos digitales, que son legibles para una máquina u ordenador y codifican programas de instrucciones ejecutables por una máquina o por un ordenador, en donde dichas instrucciones realizan algunos o todos los pasos de dichos métodos descritos anteriormente. Los dispositivos de almacenamiento de programas pueden ser, por ejemplo, memorias digitales, medios de almacenamiento magnéticos tales como discos y cintas magnéticas, discos duros o medios de almacenamiento de datos digitales de lectura óptica. Se pretende que las realizaciones también abarquen ordenadores programados para realizar dichos pasos de los métodos descritos anteriormente.
- 25 Las funciones de los distintos elementos que se muestran en las figuras, incluyendo los bloques funcionales etiquetados como "módulos", pueden proporcionarse mediante el uso de hardware dedicado, así como de hardware capaz de ejecutar un software en asociación con el software apropiado. Cuando se proporcionan por un procesador, las funciones pueden proporcionarse por un solo procesador dedicado, por un solo procesador compartido o por una pluralidad de procesadores individuales, algunos de los cuales pueden ser compartidos. Además, no debe considerarse que el uso explícito del término "módulo" se refiera exclusivamente al hardware capaz de ejecutar software, y puede incluir implícitamente, sin limitación, hardware del procesador de señales digitales (DSP), procesador de red, circuito integrado de aplicación específica (ASIC), matriz de puertas lógicas programable en campo (FPGA), memoria de solo lectura (ROM) para almacenar el software, memoria de acceso aleatorio (RAM) y almacenamiento no volátil. También se puede incluir otro hardware, convencional y/o personalizado.
- 35 Aunque que los principios de la invención han sido establecidos anteriormente en conexión con realizaciones específicas, debe entenderse que esta descripción se hace simplemente a modo de ejemplo y no como una limitación del alcance de protección que se determina por las reivindicaciones adjuntas.

**REIVINDICACIONES**

1. Un método para estimar el género de un feto de una hembra embarazada, comprendiendo dicho método:

- 5 - medir presencias alélicas ( $D_X$ ) para una primera pluralidad de marcadores genéticos del cromosoma X y presencias alélicas ( $D_R$ ) para una segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia, diferente de los cromosomas X e Y, en una muestra de ADN libre de células de una hembra embarazada; representando cada presencia alélica la presencia en un marcador genético de al menos uno de:
  - 10 - un alelo de referencia de origen materno o fetal, y un alelo alternativo de origen materno o fetal;
  - en base a dichas presencias alélicas medidas para dicha primera pluralidad, determinar una primera fracción ( $F_X$ ) de las mismas que se asocia a marcadores genéticos puramente homocigotos;
  - en base a dicha medida de las presencias alélicas para dicha segunda pluralidad, determinar una segunda fracción ( $F_R$ ) de las mismas que se asocia a marcadores genéticos puramente homocigotos; y
  - 15 - estimar el género de dicho feto en base a dicha primera y segunda fracción, en donde una primera fracción más alta en comparación con la segunda fracción es indicativa de un feto del género masculino y donde una primera fracción aproximadamente igual en comparación con la segunda fracción es indicativa de un feto del género femenino.

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

- 20 - en base a dichas presencias alélicas medidas en la primera pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha primera pluralidad;
- determinar como la primera fracción la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado;

25 y/o  
en donde la determinación de la segunda fracción comprende:

- 30 - en base a dichas presencias alélicas medidas para la segunda pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha segunda pluralidad;
- determinar como segunda fracción la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado.

3. El método de la reivindicación 1 o 2, en donde la estimación comprende:

35 estimar un primer estimador de género utilizando una proporción entre la primera fracción ( $F_X$ ) y la segunda fracción ( $F_R$ ).

4. El método de una cualquiera de las reivindicaciones anteriores, en donde los pasos de medición y determinación se realizan para un lote que comprende una pluralidad de muestras; en donde para cada muestra del lote, se calculan la primera y la segunda fracción; y en donde el paso de estimación del género se basa además en dicha primera y segunda fracción de cada muestra.

5. El método de la reivindicación 4, en donde el paso de estimar un género comprende:

- 45 determinar un primer estimador de género ( $E_{H1}$ ) para cada muestra en base a dicha primera y segunda fracción para dicha muestra; para cada muestra para la cual el primer estimador de género indica que el género del feto es femenino, determinar una proporción ( $N_{xi}$ ) entre las presencias alélicas medidas de la primera pluralidad y las presencias alélicas medidas de la segunda pluralidad de marcadores genéticos, y determinar los parámetros de distribución estadística de dicha pluralidad de proporciones para dichas muestras asociadas a un primer
- 50 estimador de género que indica que el feto es femenino;
- utilizando dicha distribución estadística para determinar un segundo estimador de género, opcionalmente en donde el paso de estimar un género comprende además: determinar un estimador de género combinado utilizando dicho primer y segundo estimador de género.

6. El método de una cualquiera de las reivindicaciones anteriores, donde el paso de estimar un género comprende el cálculo de un primer estimador de género  $E_{H1}$  que se espera que sea aproximadamente 0 para un feto del género femenino y aproximadamente 1 para un feto del género masculino, como sigue:

$$E_{H1} = (E_{H0} - 1) / (E_{He} - 1),$$

60 con

$$E_{H0} = F_X / F_R;$$

65 siendo  $F_X$  la primera fracción;  
siendo  $F_R$  la segunda fracción; y

siendo  $E_{He}$  un valor predictivo para  $E_{H0}$  en caso de un feto del género masculino,

opcionalmente, en donde el paso de estimar un género comprende el cálculo de un segundo estimador de género  $E_{H2}$  para la muestra como

5

$$E_{H2} = Z_{O_i} / Z_{M_i},$$

siendo  $Z_{O_i}$  un valor-z calculado para la muestra  $i$  de un lote de análisis como:

$$Z_{O_i} = \frac{N_{X_i} - \mu}{\sigma}$$

10

siendo  $Z_{M_i}$  un valor esperado de valor-z bajo la suposición de que la muestra  $i$  es "masculina":

$$Z_{M_i} = -\frac{\mu}{\sigma} \times \frac{FF}{2}$$

siendo  $FF$  una fracción fetal calculada de la muestra  $i$ ,

15

siendo  $\mu$  un valor medio y  $\sigma$  una desviación estándar calculada para los valores  $N_{X_i}$  para una pluralidad de muestras de un lote que se determinaron "femeninas" utilizando el primer estimador de género,

siendo  $N_{X_i}$  una relación entre las presencias alélicas medidas de una primera pluralidad de marcadores genéticos del cromosoma X y las presencias alélicas medidas de una segunda pluralidad de marcadores genéticos de un cromosoma de referencia, para una muestra  $i$ .

20

7. El método de una cualquiera de las reivindicaciones anteriores:

a) en donde la muestra es sangre materna, plasma, orina, líquido cefalorraquídeo, suero, saliva o es fluido de lavado transcervical;

25

b) en donde dicha etapa de medición comprende al menos una de los 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; y/o

(c) comprendiendo además el suministro de la muestra *in vitro*.

30

8. Un sistema para estimar el género de un feto de una hembra embarazada, comprendiendo dicho sistema:

- un dispositivo de medición configurado para medir presencias alélicas ( $D_X$ ) para una primera pluralidad de marcadores genéticos del cromosoma X y presencias alélicas ( $D_R$ ) para una segunda pluralidad de marcadores genéticos de al menos un cromosoma de referencia, diferente de los cromosomas X e Y, en una muestra de ADN libre de células de una hembra embarazada; representando cada presencia alélica la presencia en un marcador genético de al menos uno de: un alelo de referencia de origen materno o fetal, y un alelo alternativo de origen materno o fetal;

35

- un módulo de determinación configurado para determinar una primera fracción ( $F_X$ ) de dichas presencias alélicas medidas para dicha primera pluralidad que se asocia a marcadores genéticos puramente homocigotos, y una segunda fracción ( $F_R$ ) de dichas presencias alélicas medidas para dicha segunda pluralidad que se asocia a marcadores genéticos puramente homocigotos;

40

- un módulo de estimación configurado para estimar el género de dicho feto a partir de dicha primera y segunda fracción, en donde una primera fracción más alta en comparación con la segunda fracción es indicativa de un feto del género masculino y en el que una primera fracción aproximadamente igual en comparación con la segunda fracción es indicativa de un feto del género femenino

45

9. El sistema de la reivindicación 8, en donde el módulo de determinación esta configurado para:

- en base a dichas presencias alélicas medidas en la primera pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha primera pluralidad;

50

- determinar como la primera fracción la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado;

y/o para:

55

- en base a dichas presencias alélicas medidas para la segunda pluralidad de marcadores genéticos, calcular un número correspondiente de frecuencias alélicas para dicha segunda pluralidad;

- determinar como segunda fracción la fracción de dichas presencias alélicas medidas para las cuales la frecuencia alélica es 0 o 1 dentro de un margen de error predeterminado.

60

10. El sistema de la reivindicación 8 o 9, en donde el módulo de estimación está configurado para:

estimar un primer estimador de género utilizando una proporción entre la primera fracción ( $F_X$ ) y la segunda fracción ( $F_R$ ).

5 11. El sistema de una cualquiera de las reivindicaciones 8-10, en donde el dispositivo de medición y el módulo de determinación están configurados para realizar la medición y la determinación para un lote que comprende una pluralidad de muestras, en donde para cada muestra del lote, se calculan la primera y la segunda fracción; y en donde el módulo de estimación se configura para estimar el género en base a dicha primera y segunda fracción de cada muestra.

10 12. El sistema de la reivindicación 11, en donde el módulo de estimación está configurado para estimar un género mediante:

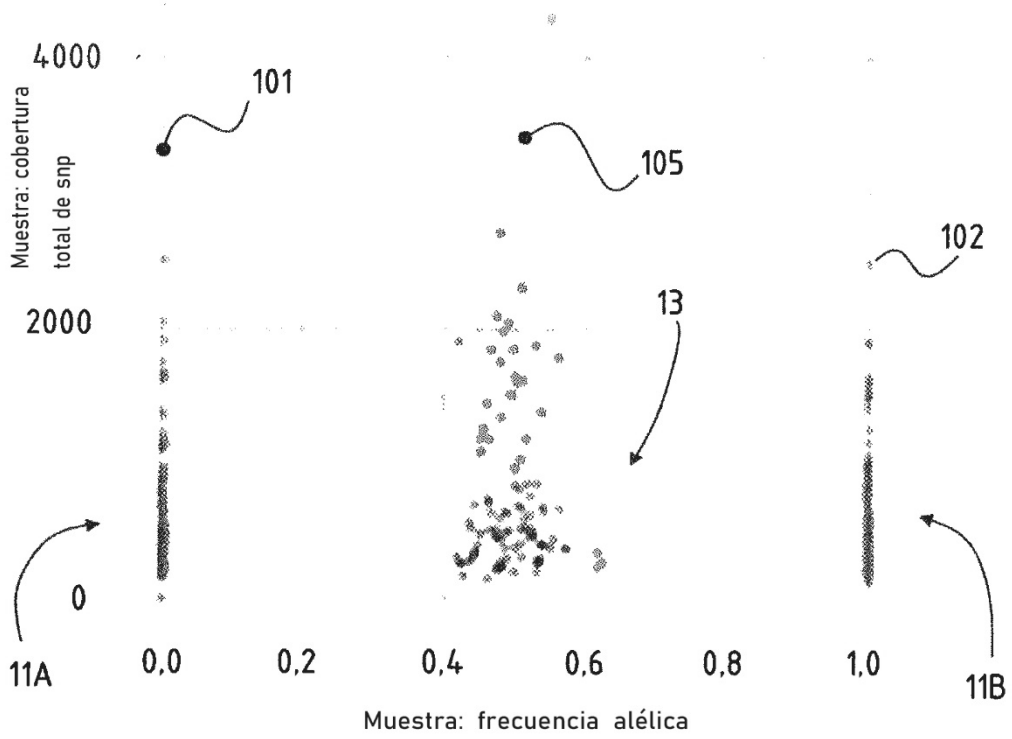
15 la determinación de un primer estimador de género ( $E_{H1}$ ) para cada muestra en base a dicha primera y segunda fracción para dicha muestra; para cada muestra para la cual el primer estimador de género indica que el género del feto es femenino, la determinación de una proporción ( $N_{xi}$ ) entre las presencias alélicas medidas de la primera pluralidad y las presencias alélicas medidas de la segunda pluralidad de marcadores genéticos, y la determinación de los parámetros de distribución estadística de dicha pluralidad de proporciones para dichas muestras asociadas a un primer estimador de género que indica que el feto es femenino;  
20 utilizando dicha distribución estadística para determinar un segundo estimador de género; y, opcionalmente, la determinación de un estimador de género combinado utilizando dicho primer y segundo estimador de género.

25 13. El sistema de una cualquiera de las reivindicaciones 8-12, en donde el módulo de medición está configurado 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.

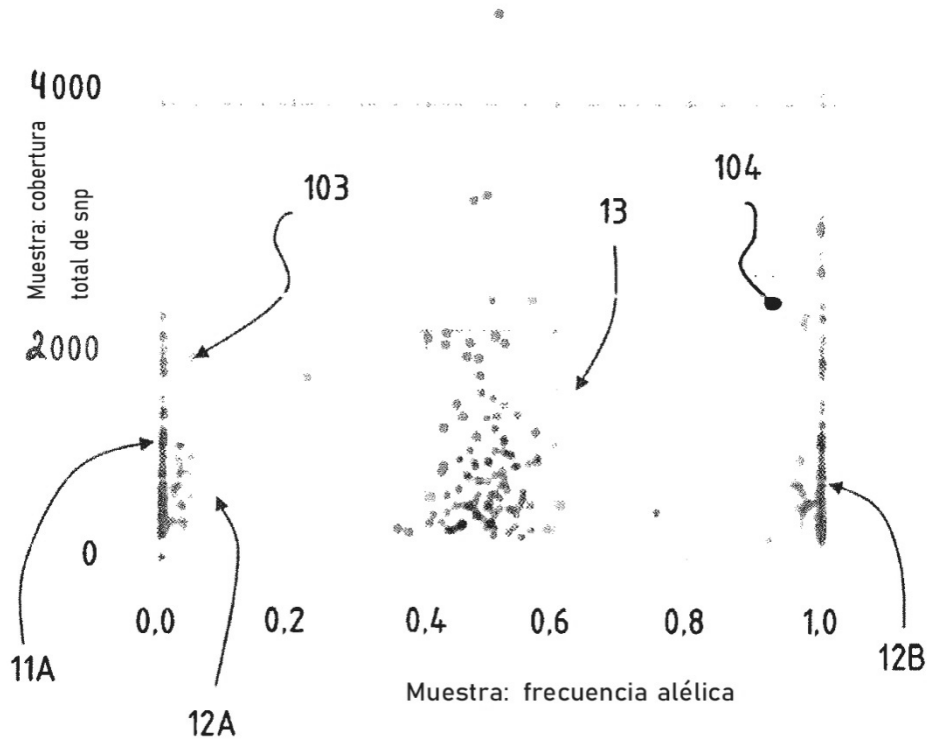
30 14. Un programa informático que comprende instrucciones ejecutables por ordenador para realizar, cuando el programa se ejecuta en un ordenador, al menos el paso de estimar el género del método de una cualquiera de las reivindicaciones 1-7.

15. Un medio de almacenamiento de datos digitales que codifica un programa de instrucciones ejecutables por una máquina para realizar al menos el paso de estimación del género del método de una cualquiera de las reivindicaciones 1-7.

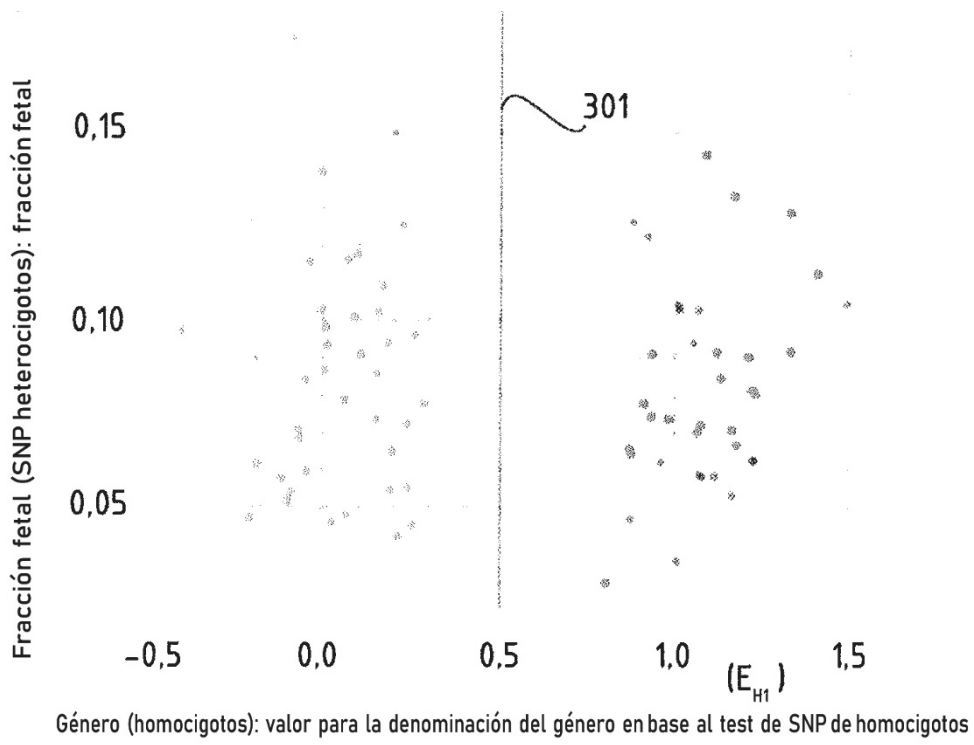




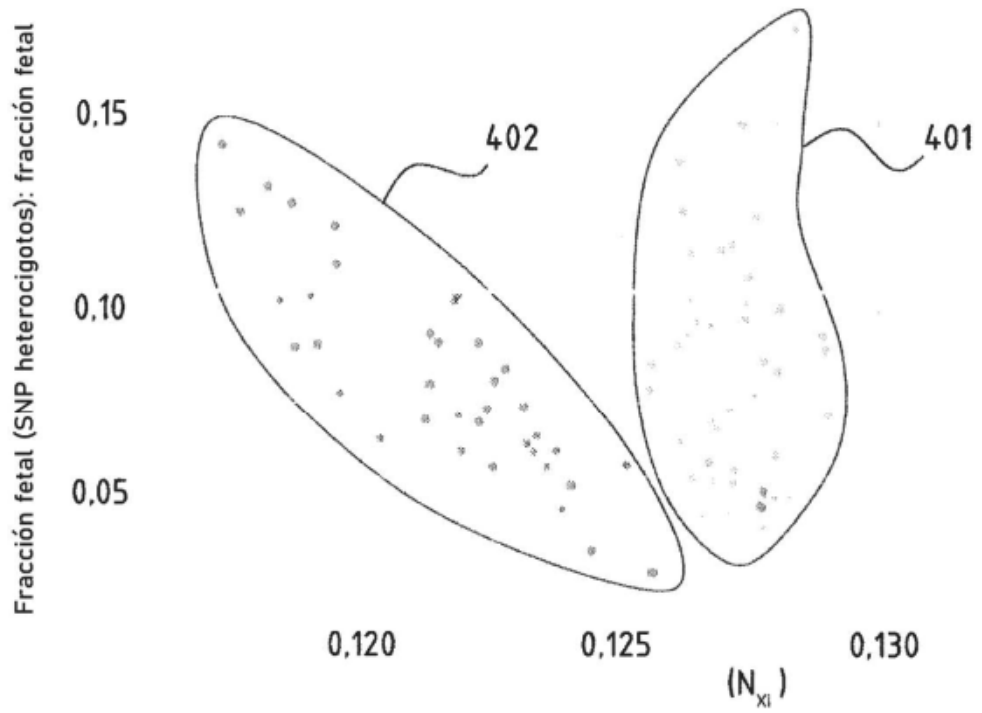
**FIG. 1**



**FIG. 2**

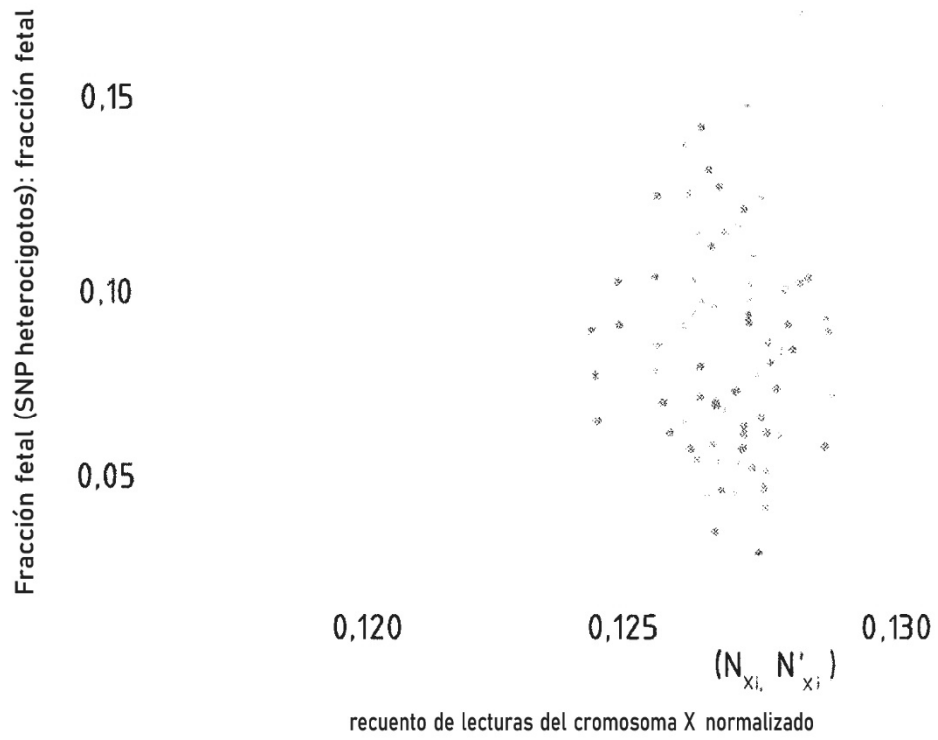


**FIG. 3**

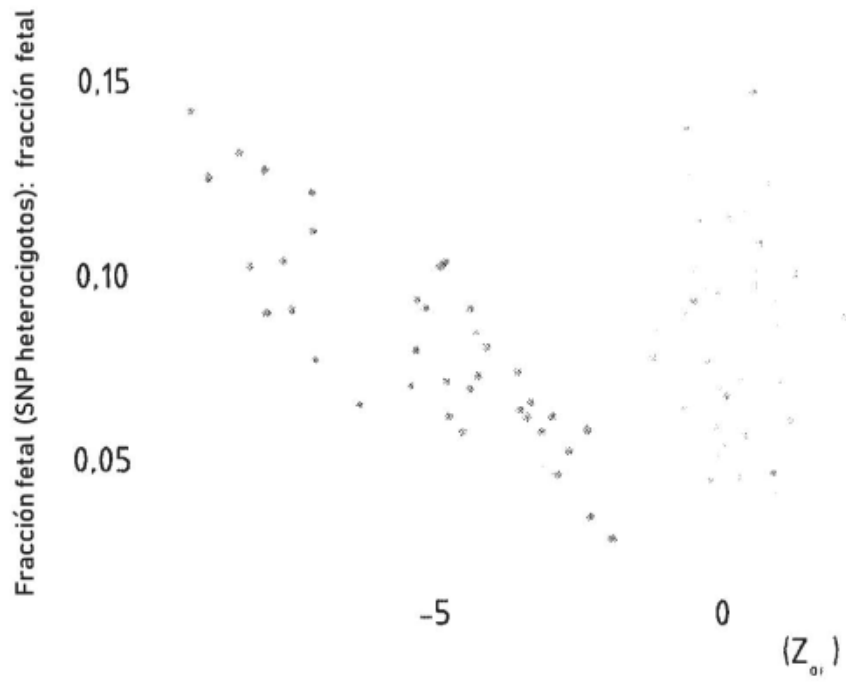


Género (homocigotos): recuento de lecturas del cromosoma X normalizado

**FIG. 4**

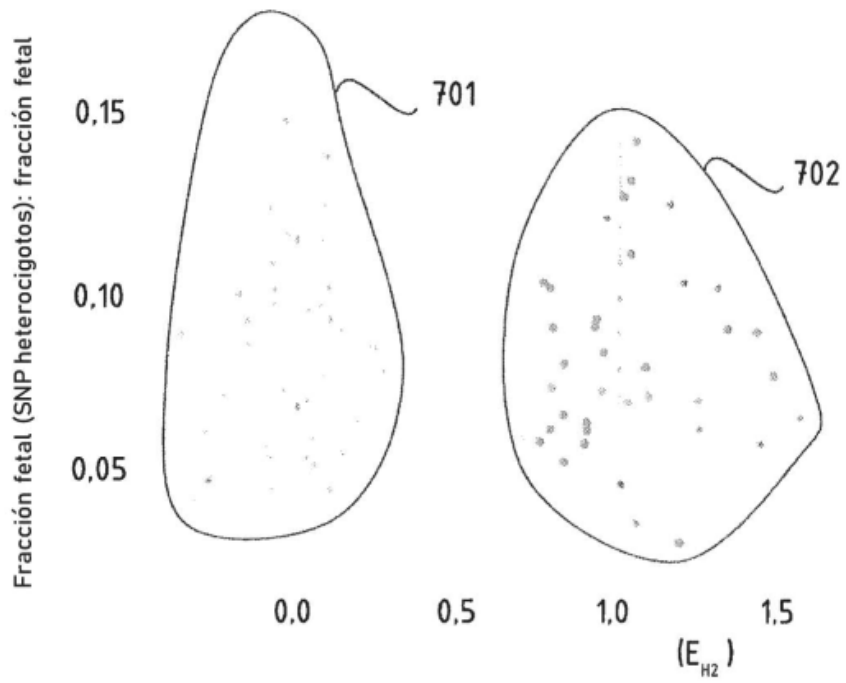


**FIG. 5**



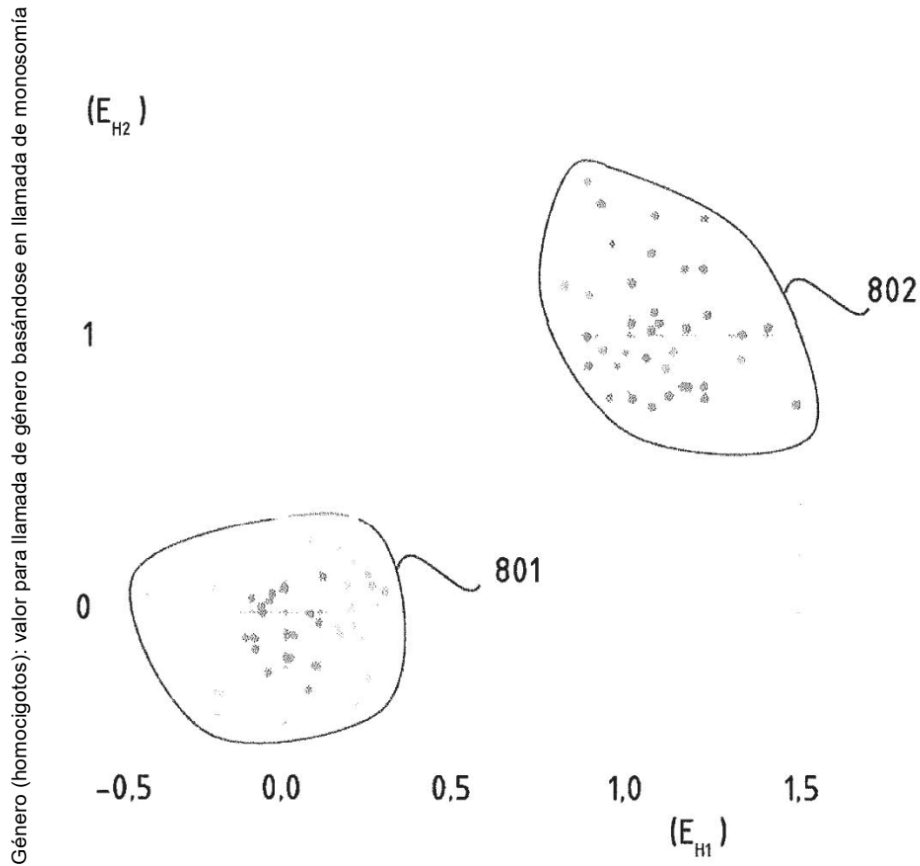
Género (homocigotos): valor-z negativo para llamada de monosomía de género

**FIG. 6**



Género (homocigotos): valor para la llamada de género basándose en llamada de monosomía

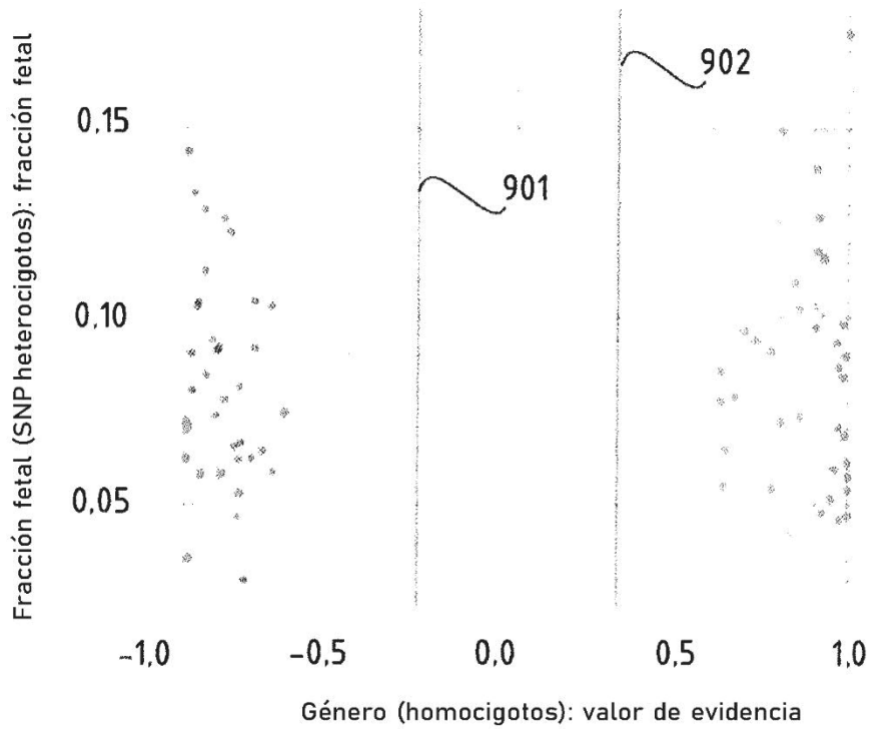
**FIG. 7**



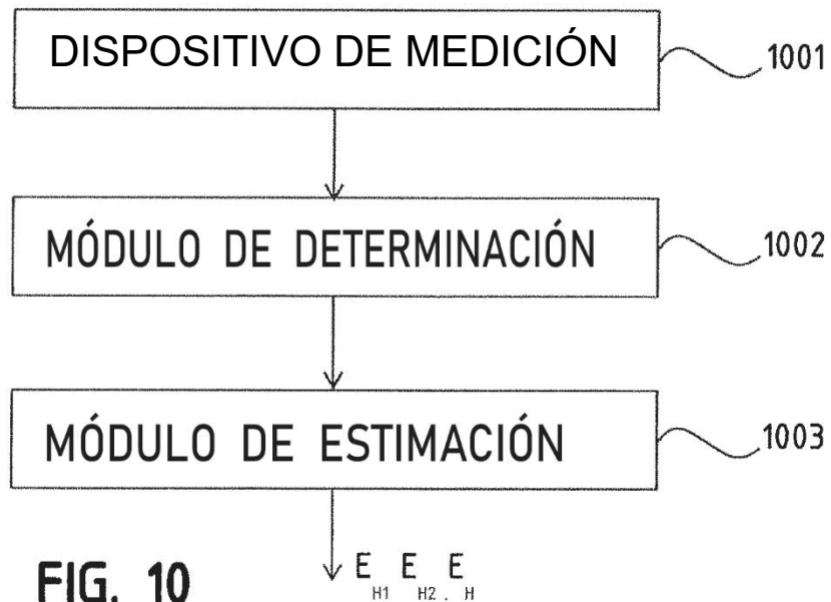
Género (homocigotos): valor para llamada de género basándose en una prueba de SNP homocigotos

**FIG. 8**





**FIG. 9**



**FIG. 10**