(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 622 088**

(51) Int. CI.:

***C12Q 1/68*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **18.11.2011 E 14195468 (5)**

(97) Fecha y número de publicación de la concesión europea: **11.01.2017 EP 2902500**

(54) Título: **Métodos no invasivos para la determinación del estado de ploidía prenatal**

(30) Prioridad:

**09.02.2011 US 201161462972 P**
**02.03.2011 US 201161448547 P**
**12.04.2011 US 201161516996 P**
**18.05.2011 US 201113110685**
**23.06.2011 US 201161571248 P**
**03.10.2011 US 201161542508 P**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**05.07.2017**

(73) Titular/es:

**NATERA, INC. (100.0%)**
**201 Industrial Road Suite 410**
**San Carlos, CA 94070, US**

(72) Inventor/es:

**RABINOWITZ, MATTHEW;**
**GEMELOS, GEORGE;**
**BANJEVIC, MILENA;**
**RYAN, ALLISON;**
**DEMKO, ZACHARY;**
**HILL, MATTHEW;**
**ZIMMERMANN, BERNHARD y**
**BANER, JOHAN**

(74) Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

## DESCRIPCIÓN

Métodos no invasivos para la determinación del estado de ploidía prenatal

**Solicitudes relacionadas**

La presente solicitud es una continuación en parte de la Solicitud de Patente USA con el número de serie 13/110 685, presentada el 18 de mayo de 2011, que reivindica las ventajas de la Solicitud de Patente provisional USA con el número de serie 61/395 850, presentada el 18 de mayo de 2010; la Solicitud de Patente provisional USA con el número de serie 61/398 159, presentada el 21 de junio de 2010; la Solicitud de Patente provisional USA con el número de serie 61/462 972, presentada el 9 de febrero de 2011; la Solicitud de Patente provisional USA con el número de serie 61/448 547, presentada el 2 de marzo de 2011; y la Solicitud de Patente provisional USA con el número de serie 61/516 996, presentada el 12 de abril de 2011; y la presente solicitud reivindica asimismo las ventajas de la Solicitud de Patente provisional USA con el número de serie 61/571 248, presentada el 23 de junio de 2011.

**Campo**

La presente divulgación se refiere en general a métodos no invasivos para la determinación del estado de ploidía prenatal.

**Antecedentes**

Los métodos de diagnóstico prenatal actuales pueden alertar a los médicos y progenitores de anomalías en los fetos en gestación. Sin un diagnóstico prenatal, uno de cada 50 bebés nacería con un problema físico o mental grave, y hasta uno de cada 30 tendría algún tipo de malformación congénita. Lamentablemente, los métodos estándar ofrecen poca precisión o implican un procedimiento invasivo que conlleva un riesgo de aborto. Los métodos basados en los niveles hormonales en la sangre materna o en las mediciones por ultrasonidos no son invasivos; sin embargo, su nivel de precisión es muy escaso. Los métodos como la amniocentesis, la biopsia coriónica y la toma de muestras de sangre fetal ofrecen una alta precisión, pero son invasivos y conllevan importantes riesgos.

La amniocentesis se realiza aproximadamente en el 3% de todos los embarazos en los Estados Unidos, aunque su frecuencia de uso ha descendido en los últimos 15 años.

Recientemente se ha descubierto que el ADN fetal libre de células y las células fetales intactas pueden entrar en la circulación de la sangre materna. Por consiguiente, el análisis de este material genético puede permitir el diagnóstico genético prenatal no invasivo (NPD).

Los humanos normales tienen dos series de 23 cromosomas en cada célula diploide sana, con una copia procedente de cada progenitor. Se cree que la aneuploidía, una condición de una célula nuclear en la que la célula contiene demasiados y/o insuficientes cromosomas, es responsable de un gran porcentaje de implantaciones fracasadas, abortos y enfermedades genéticas. La detección de anomalías cromosómicas puede identificar a los individuos o embriones con condiciones como el síndrome de Down, el síndrome de Klinefelter y el síndrome de Turner, entre otros, además de aumentar las probabilidades de que un embarazo tenga éxito. Las pruebas de anomalías cromosómicas son especialmente importantes debido a la edad de la madre: se calcula que entre los 35 y los 40 años al menos el 40% de los embriones presentan anomalías y a partir de los 40 años más de la mitad de los embriones presentan anomalías.

*Algunas pruebas utilizadas para el análisis prenatal*

Los bajos niveles de proteína A plasmática asociada al embarazo (PAPP-A) medida en el suero materno durante el primer trimestre pueden estar relacionados con anomalías cromosómicas fetales incluyendo las trisomías 13, 18 y 21. Por otra parte, unos bajos niveles de PAPP-A en el primer trimestre pueden predecir un resultado adverso del embarazo, incluyendo un bebé pequeño para su edad gestacional (SGA) o muerte fetal. Las mujeres embarazadas a menudo se someten al análisis del suero del primer trimestre, que normalmente implica la comprobación de los niveles en sangre de las hormonas PAPP-A y gonadrotropina coriónica humana beta (beta-hCG). En algunos casos las mujeres también se someten a ultrasonidos para buscar posibles defectos fisiológicos. En concreto, la medición de la translucencia nucal (NT) puede indicar un riesgo de aneuploidía en un feto. En muchas regiones, el estándar de tratamiento para el análisis prenatal incluye el análisis del suero del primer trimestre combinado con una prueba de NT.

La triple prueba, también denominada triple screening, la prueba de Kettering o la prueba de Bart, es una investigación realizada durante el embarazo en el segundo trimestre para clasificar a una paciente como de alto riesgo o bajo riesgo de anomalías cromosómicas (y defectos del tubo neural). En ocasiones también se utiliza el término «análisis de marcadores múltiples». El término «triple prueba» puede abarcar los términos «prueba doble», «triple prueba», «prueba cuádruple» y «prueba quíntuple».

La triple prueba mide los niveles en suero de alfafetoproteína (AFP), estriol no conjugado (UE2) gonadotropina coriónica humana beta (beta-hCG), antígeno trofoblástico invasivo (ITA) y/o inhibina. Una prueba positiva significa tener un elevado riesgo de anomalías cromosómicas (y defectos del tubo neural) y estas pacientes son posteriormente remitidas a procedimientos más específicos y sensibles, con el fin de obtener un diagnóstico definitivo, en su mayoría procedimientos invasivos como la amniocentesis. La triple prueba se puede utilizar para

detectar diversas condiciones, incluyendo la trisomía 21 (síndrome de Down). Además del síndrome de Down, las pruebas triple y cuádruple detectan la trisomía fetal 18, también conocida como síndrome de Edward, defectos del cierre del tubo neural, y también pueden detectar un mayor riesgo de síndrome de Turner, tripoidía, trisomía 16, mosaicismo, muerte fetal, síndrome de Smith-Lemli-Opitz, y deficiencia de sulfatasa esteroidea.

Shen Zhiyong et al. 2010, BMC Bioinformatics, 11(1): 143 divulga un programa llamado MPprimer: un programa para el diseño de cebadores fiables para PCR multiplexada, una herramienta para diseñar combinaciones de conjuntos de cebadores no dimerizantes específicos con una dimensión de amplicones limitada para los ensayos PCR multiplexada.

**Resumen**

La invención se define en las reivindicaciones adjuntas. En el presente documento se divulgan métodos para determinar el estado de ploidía de un cromosoma en un feto gestante. De acuerdo con aspectos de la divulgación ilustrados en el presente documento, en una realización un método para determinar un estado de ploidía de un cromosoma en un feto gestante incluye la obtención de una primera muestra de ADN que comprende ADN materno de la madre del feto y ADN fetal del feto, la preparación de la primera muestra aislando el ADN con el fin de obtener una muestra preparada, la medición del ADN en la muestra preparada en diversos loci polimórficos del cromosoma, el cálculo por ordenador de los recuentos de alelos en la pluralidad de loci polimórficos de las mediciones de ADN realizadas en la muestra preparada, la creación por ordenador de una pluralidad de hipótesis de ploidía correspondiente cada una de ellas a un posible estado de ploidía diferente del cromosoma, la creación por ordenador de un modelo de distribución conjunto para los recuentos de alelos previstos en la pluralidad de loci polimórficos del cromosoma para cada una de las hipótesis de ploidía, la determinación por ordenador de una probabilidad relativa de cada una de las hipótesis de ploidía utilizando el modelo de distribución conjunto y los recuentos de alelos medidos en la muestra preparada, y la determinación del estado de ploidía del feto seleccionando el estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada.

En algunas realizaciones de la divulgación, el ADN de la primera muestra procede del plasma materno. En algunas realizaciones, la preparación de la primera muestra comprende también la amplificación del ADN. En algunas realizaciones, la preparación de la primera muestra comprende también el enriquecimiento preferente del ADN de la primera muestra en diversos loci polimórficos.

En algunas realizaciones de la divulgación, el enriquecimiento preferente del ADN de la primera muestra en la pluralidad de loci polimórficos incluye la obtención de una pluralidad de sondas precircularizadas donde cada sonda está dirigida a uno de los loci polimórficos, y donde el extremo 3' y 5' de las sondas están diseñados para hibridarse con una región de ADN que está separada del punto polimórfico del locus por un pequeño número de bases, donde el pequeño número es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 a 25, 26 a 30, 31 a 60, o una combinación de estos, la hibridación de las sondas precircularizadas con el ADN de la muestra, el relleno del hueco entre los extremos de la sonda hibridada utilizando ADN polimerasa, la circularización de la sonda precircularizada, y la amplificación de la sonda circularizada.

En algunas realizaciones de la divulgación, el enriquecimiento preferente del ADN en la pluralidad de loci polimórficos incluye la obtención de una pluralidad de sondas para PCR mediadas por unión donde cada sonda para PCR está dirigida a uno de los loci polimórficos, y donde las sondas para PCR corriente ascendente y descendente están diseñadas para hibridarse con una región de ADN, en una cadena de ADN, que está separada del punto polimórfico del locus por un pequeño número de bases, donde el pequeño número es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 a 25, 26 a 30, 31 a 60, o una combinación de estos, la hibridación de las sondas para PCR mediadas por unión con el ADN de la primera muestra, el relleno del hueco entre los extremos de la sonda para PCR mediada por unión utilizando ADN polimerasa, la unión de las sondas para PCR mediadas por unión, y la amplificación de las sondas para PCR mediadas por unión.

En algunas realizaciones de la divulgación, el enriquecimiento preferente del ADN en la pluralidad de loci polimórficos incluye la obtención de una pluralidad de sondas de captura híbrida dirigidas a los loci polimórficos, la hibridación de las sondas de captura híbrida con el ADN de la primera muestra y la eliminación física de parte o la totalidad del ADN no hibridado de la primera muestra de ADN.

En algunas realizaciones de la divulgación, las sondas de captura híbridas están diseñadas para hibridarse con una región que flanquea pero no solapa el punto polimórfico. En algunas realizaciones, las sondas de captura híbridas están diseñadas para hibridarse con una región que flanquea pero no solapa el punto polimórfico, y donde la longitud de la sonda de captura de flanqueo se puede seleccionar del grupo compuesto por menos de unas 120 bases, menos de unas 110 bases, menos de unas 100 bases, menos de unas 90 bases, menos de unas 80 bases, menos de unas 70 bases, menos de unas 60 bases, menos de unas 50 bases, menos de unas 40 bases, menos de unas 30 bases y menos de unas 25 bases. En algunas realizaciones, las sondas de captura híbridas están diseñadas para hibridarse con una región que solapa el punto polimórfico y donde la pluralidad de sondas de captura híbridas comprenden al menos dos sondas de captura híbridas para cada loci polimórfico, y donde cada sonda de captura híbrida está diseñada para ser complementaria de un alelo diferente en ese locus polimórfico.

En algunas realizaciones de la divulgación, el enriquecimiento preferente del ADN en diversos loci polimórficos incluye la obtención de diversos cebadores directos internos donde cada cebador se dirige a uno de los loci polimórficos, y donde el extremo 3' de los cebadores directos internos está diseñado para hibridarse con una región

de ADN en sentido ascendente del punto polimórfico, y separada del punto polimórfico por un pequeño número de bases, donde el pequeño número de bases se selecciona del grupo compuesto por 1, 2, 3, 4, 5, 6 a 10, 11 a 15, 16 a 20, 21 a 25, 26 a 30, o 31 a 60 pares de bases, opcionalmente obteniendo diversos cebadores inversos internos donde cada cebador se dirige a uno de los loci polimórficos, y donde el extremo 3' de los cebadores inversos internos está diseñado para hibridarse con una región de ADN en sentido ascendente del punto polimórfico, y separada del punto polimórfico por un pequeño número de bases, donde el pequeño número de bases se selecciona del grupo compuesto por 1, 2, 3, 4, 5, 6 a 10, 11 a 15, 16 a 20, 21 a 25, 26 a 30, o 31 a 60 pares de bases, la hibridación de los cebadores internos con el ADN, y la amplificación del ADN utilizando la reacción en cadena de la polimerasa para formar amplicones.

En algunas realizaciones de la divulgación, el método también incluye la obtención de diversos cebadores directos externos donde cada cebador se dirige a uno de los loci polimórficos, y donde los cebadores directos externos están diseñados para hibridarse con una región de ADN en sentido ascendente del cebador directo interno, opcionalmente obteniendo diversos cebadores inversos externos donde cada cebador se dirige a uno de los loci polimórficos, y donde los cebadores inversos externos están diseñados para hibridarse con la región de ADN inmediatamente en sentido descendente del cebador inverso interno, la hibridación de los primeros cebadores con el ADN, y la amplificación del ADN utilizando la reacción en cadena de la polimerasa.

En algunas realizaciones de la divulgación, el método también incluye la obtención de diversos cebadores inversos externos donde cada cebador se dirige a uno de los loci polimórficos, y donde los cebadores inversos externos están diseñados para hibridarse con la región de ADN inmediatamente en sentido descendente del cebador inverso interno, opcionalmente obteniendo diversos cebadores directos externos donde cada cebador se dirige a uno de los loci polimórficos, y donde los cebadores directos externos están diseñados para hibridarse con la región de ADN en sentido ascendente del cebador directo interno, la hibridación de los primeros cebadores con el ADN, y la amplificación del ADN utilizando la reacción en cadena de la polimerasa.

En algunas realizaciones de la divulgación, la preparación de la primera muestra incluye también la adición de adaptadores universales al ADN en la primera muestra y la amplificación del ADN en la primera muestra utilizando la reacción en cadena de la polimerasa. En algunas realizaciones, al menos una fracción de los amplicones que son amplificados tienen menos de 100 pb, menos de 90 pb, menos de 80 pb, menos de 70 pb, menos de 65 pb, menos de 60 pb, menos de 55 pb, menos de 50 pb, o menos de 45 pb, y donde la fracción es 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, o 99%.

En algunas realizaciones de la divulgación, la amplificación del ADN se realiza en una de una pluralidad de volúmenes de reacción individuales, y donde cada volumen de reacción individual contiene más de 100 pares de cebadores directos e inversos diferentes, más de 200 pares de cebadores directos e inversos diferentes, más de 500 pares de cebadores directos e inversos diferentes, más de 1000 pares de cebadores directos e inversos diferentes, más de 2000 pares de cebadores directos e inversos diferentes, más de 5000 pares de cebadores directos e inversos diferentes, más de 10 000 pares de cebadores directos e inversos diferentes, más de 20 000 pares de cebadores directos e inversos diferentes, más de 50 000 pares de cebadores directos e inversos diferentes, o más de 100 000 pares de cebadores directos e inversos diferentes.

En algunas realizaciones de la divulgación, la preparación de la primera muestra consiste asimismo en dividir la primera muestra en diversas porciones, y donde el ADN de cada porción se enriquece preferentemente en un subconjunto de la pluralidad de loci polimórficos. En algunas realizaciones, los cebadores internos se seleccionan identificando pares de cebadores que es probable que formen dúplex de cebadores no deseados y eliminando de la pluralidad de cebadores al menos uno de los pares de cebadores identificados por su probabilidad de formar dúplex de cebadores no deseados. En algunas realizaciones, los cebadores internos contienen una región que está diseñada para hibridarse en sentido ascendente o descendente del locus polimórfico diana, y opcionalmente contienen una secuencia de cebado universal diseñada para permitir la amplificación por PCR. En algunas realizaciones, al menos algunos de los cebadores contienen adicionalmente una región aleatoria que difiere para cada molécula del cebador individual. En algunas realizaciones, al menos algunos de los cebadores contienen además un código de barras molecular.

En algunas realizaciones de la divulgación, el método incluye también la obtención de datos genotípicos de uno o los dos progenitores del feto. En algunas realizaciones, la obtención de datos genotípicos de uno o los dos progenitores del feto incluye la preparación del ADN de los padres, donde la preparación comprende el enriquecimiento preferente del ADN en la pluralidad de loci polimórficos para obtener ADN parental preparado, opcionalmente la amplificación del ADN parental preparado, y la medición del ADN parental en la muestra preparada en la pluralidad de loci polimórficos.

En algunas realizaciones de la divulgación, la creación de un modelo de distribución conjunto para las probabilidades previstas del recuento de alelos de la pluralidad de loci polimórficos en el cromosoma se realiza utilizando los datos genéticos obtenidos de uno o los dos progenitores. En algunas realizaciones, la primera muestra ha sido aislada del plasma materno y la obtención de datos genotípicos de la madre se realiza estimando los datos genotípicos maternos de las mediciones del ADN realizadas en la muestra preparada.

En algunas realizaciones de la realización, el enriquecimiento preferente resulta en un grado medio de sesgo alélico entre la muestra preparada y la primera muestra de un factor seleccionado del grupo compuesto por no más de un

factor de 2, no más de un factor de 1,5, no más de un factor de 1,2, no más de un factor de 1,1, no más de un factor de 1,05, no más de un factor de 1,02, no más de un factor de 1,01, no más de un factor de 1,005, no más de un factor de 1,002, no más de un factor de 1,001 y no más de un factor de 1,0001. En algunas realizaciones, la pluralidad de loci polimórficos son polimorfismos de un solo nucleótido (SNP). En algunas realizaciones, la medición del ADN en la muestra preparada se realiza mediante secuenciación.

En algunas realizaciones de la divulgación, se divulga un cuadro de diagnóstico para ayudar a determinar un estado de ploidía de un cromosoma en un feto en gestación, donde el cuadro de diagnóstico es capaz de ejecutar los pasos de preparación y medición del método de la reivindicación 1.

En algunas realizaciones de la divulgación, los recuentos de alelos son probabilísticos en lugar de binarios. En algunas realizaciones, las mediciones del ADN en la muestra preparada en la pluralidad de loci polimórficos también se utilizan para determinar si el feto ha heredado o no una o una pluralidad de haplotipos vinculados con enfermedades.

En algunas realizaciones de la divulgación, la creación de un modelo de distribución conjunto para las probabilidades del recuento de alelos se realiza utilizando datos sobre la probabilidad de que los cromosomas se entrecrucen en diferentes lugares de un cromosoma conforme a la dependencia del modelo entre alelos polimórficos del cromosoma. En algunas realizaciones, la creación de un modelo de distribución conjunto para los recuentos de alelos y el paso de determinación de la probabilidad relativa de cada hipótesis se realiza utilizando un método que no requiere el uso de un cromosoma de referencia.

En algunas realizaciones de la divulgación, la determinación de la probabilidad relativa de cada hipótesis hace uso de una fracción estimada de ADN fetal en la muestra preparada. En algunas realizaciones, las mediciones de ADN de la muestra preparada utilizada para calcular las probabilidades del recuento de alelos y determinar la probabilidad relativa de cada hipótesis comprenden los datos genéticos primarios. En algunas realizaciones, la selección del estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada se realiza utilizando estimaciones máximas de probabilidad o estimaciones máximas a posteriori.

En algunas realizaciones de la divulgación, la determinación del estado de ploidía del feto incluye también la combinación de las probabilidades relativas de cada una de las hipótesis de ploidía determinadas utilizando el modelo de distribución conjunto y las probabilidades del recuento de alelos con probabilidades relativas de cada una de las hipótesis de ploidía que se calculan utilizando técnicas estadísticas tomadas de un grupo compuesto por un análisis del recuento leído, la comparación de las tasas de heterocigosidad, una estadística que solamente está disponible cuando se utiliza información genética parental, la probabilidad de señales genotípicas normalizadas para determinados contextos parentales, una estadística que se calcula utilizando una fracción fetal estimada de la primera muestra o de la muestra preparada, y combinaciones de estas.

En algunas realizaciones de la divulgación, se calcula una estimación de certeza para el estado de ploidía determinado. En algunas realizaciones, el método también incluye emprender una acción clínica basada en el estado de ploidía determinado del feto, donde la acción clínica se selecciona entre una de terminar el embarazo o mantener el embarazo.

En algunas realizaciones de la divulgación, el método se puede realizar para fetos de entre 4 y 5 semanas de gestación; entre 5 y 6 semanas de gestación; entre 6 y 7 semanas de gestación; entre 7 y 8 semanas de gestación; entre 8 y 9 semanas de gestación; entre 9 y 10 semanas de gestación; entre 10 y 12 semanas de gestación; entre 12 y 14 semanas de gestación; entre 14 y 20 semanas de gestación; entre 20 y 40 semanas de gestación; en el primer trimestre; en el segundo trimestre; en el tercer trimestre; o combinaciones de estos.

En algunas realizaciones de la divulgación, se genera un informe que contiene un estado de ploidía determinado de un cromosoma de un feto en gestación utilizando el método. En algunas realizaciones, se divulga un kit para determinar un estado de ploidía de un cromosoma diana de un feto en gestación diseñado para ser utilizado con el método de la reivindicación 9, donde el kit incluye una pluralidad de cebadores directos internos y opcionalmente la pluralidad de cebadores inversos internos, donde cada uno de los cebadores está diseñado para hibridarse con la región de ADN inmediatamente en sentido ascendente y/o descendente de uno de los puntos polimórficos en el cromosoma diana, y opcionalmente cromosomas adicionales, donde la región de hibridación está separada del punto polimórfico por un pequeño número de bases, donde el pequeño número se selecciona del grupo compuesto por 1, 2, 3, 4, 5, 6 a 10, 11 a 15, 16 a 20, 21 a 25, 26 a 30, 31 a 60, y combinaciones de estos.

En algunas realizaciones de la divulgación, se divulga un método para determinar la presencia o ausencia de aneuploidía fetal en una muestra de tejido materno que comprende ADN genómico materno y fetal, donde el método incluye: (a) la obtención de una mezcla de ADN genómico materno y fetal de dicha muestra de tejido materno; (b) la realización de una secuenciación de ADN paralela masiva de los fragmentos de ADN aleatoriamente seleccionados de la mezcla de ADN genómico materno y fetal del paso (a) con el fin de determinar la secuencia de los mencionados fragmentos de ADN; (c) la identificación de cromosomas a los que pertenecen las secuencias obtenidas en el paso (b); (d) la utilización de los datos del paso (c) para determinar una cantidad de al menos un primer cromosoma en dicha mezcla de ADN genómico materno y fetal, donde se presume que este al menos primer cromosoma es euploide en el feto; (e) la utilización de los datos del paso (c) para determinar a cantidad de un segundo cromosoma en dicha mezcla de ADN genómico materno y fetal, donde se sospecha que dicho segundo cromosoma es aneuploide en el feto; (f) calcular la fracción de ADN fetal en la mezcla de ADN materno y fetal; (g)

calcular una distribución prevista de la cantidad del segundo cromosoma diana si el segundo cromosoma diana es euploide, utilizando la cifra del paso (d); (h) calcular una distribución prevista de la cantidad del segundo cromosoma diana si el segundo cromosoma diana es aneuploide, utilizando la primera cifra del paso (d) y la fracción calculada de ADN fetal de la mezcla de ADN materno y fetal del paso (f); e (i) utilizar el método de la probabilidad máxima o máxima a posteriori para determinar si la cantidad del segundo cromosoma determinada en el paso (e) es más probable que forme parte de la distribución calculada en el paso (g) o la distribución calculada en el paso (h); indicando de este modo la presencia o ausencia de aneuploidía fetal.

*Breve descripción de los gráficos*

Las realizaciones divulgadas en el presente documento se explicarán asimismo por referencia a los gráficos adjuntos, donde las estructuras similares se designan mediante numerales similares en todas las vistas. Los esquemas mostrados no están necesariamente a escala, sino que en general se hace hincapié en ilustrar los principios de las realizaciones divulgadas en el presente documento.

Figura 1: Representación gráfica del método de mini-PCR multiplexada directa.

Figura 2: Representación gráfica del método de mini-PCR semi-anidada.

Figura 3: Representación gráfica del método de mini-PCR totalmente anidada.

Figura 4: Representación gráfica del método de mini-PCR hemi-anidada.

Figura 5: Representación gráfica del método de mini-PCR triplemente hemi-anidada.

Figura 6: Representación gráfica del método de mini-PCR anidada unilateral.

Figura 7: Representación gráfica del método de mini-PCR unilateral.

Figura 8: Representación gráfica del método de mini-PCR semi-anidada inversa.

Figura 9: Algunos flujos de trabajo posibles de los métodos semi-anidados.

Figura 10: Representación gráfica de adaptadores de unión en bucle.

Figura 11: Representación gráfica de cebadores etiquetados internamente.

Figura 12: Un ejemplo de algunos cebadores con etiquetas internas.

Figura 13: Representación gráfica de un método utilizando cebadores con una región de unión al adaptador de unión.

Figura 14: Precisiones simuladas de la determinación del estado de ploidía para el método del recuento con dos técnicas de análisis diferentes.

Figura 15: Ratio de dos alelos para una pluralidad de SNP de una línea celular en el Experimento 4.

Figura 16: Ratio de dos alelos para una pluralidad de SNP de una línea celular en el Experimento 4 clasificadas por cromosoma.

Figura 17: Ratio de dos alelos para una pluralidad de SNP en cuatro muestras de plasma de mujeres embarazadas, clasificadas por cromosoma.

Figura 18: Fracción de datos que se puede explicar por la varianza binomial antes y después de la corrección de datos.

Figura 19: Gráfico que muestra el enriquecimiento relativo del ADN fetal de las muestras siguiendo un breve protocolo de preparación de bibliotecas.

Figura 20: Profundidad del gráfico de lectura que compara los métodos de PCR directa y semi-anidada.

Figura 21: Comparación de la profundidad de lectura para la PCR directa de tres muestras genómicas.

Figura 22: Comparación de la profundidad de lectura para la mini-PCR semi-anidada de tres muestras.

Figura 23: Comparación de la profundidad de lectura para reacciones 1200-plex y 9600-plex.

Figura 24: Ratios del recuento de lectura para seis células de tres cromosomas.

Figura 25: Ratios de alelos para dos reacciones de tres células y una tercera reacción realizada con 1 ng de ADN genómico en tres cromosomas.

Figura 26: Ratios de alelos para dos reacciones de una única célula en tres cromosomas.

A pesar de que los dibujos anteriormente identificados exponen las realizaciones divulgadas en el presente documento, también se contemplan otras realizaciones, tal y como se hace constar en la exposición. La presente divulgación presente realizaciones ilustrativas a título representativo y sin carácter limitador.

**Descripción detallada**

La invención se define en las reivindicaciones adjuntas. En una realización, la presente divulgación proporciona métodos ex vivo para determinar el estado de ploidía en un cromosoma de un feto en gestación a partir de los datos genotípicos medidos de una muestra combinada de ADN (es decir, ADN de la madre del feto y ADN del feto) y opcionalmente de los datos genotípicos medidos de una muestra de material genético de la madre y posiblemente también del padre, donde la determinación se realiza utilizando un modelo de distribución conjunto para crear un conjunto de distribuciones de alelos previstas para diferentes estados de ploidía fetal posibles dados los datos genotípicos parentales, y comparando las distribuciones alélicas previstas con las distribuciones alélicas reales medidas en la muestra combinada, y seleccionando el estado de ploidía cuyo patrón de distribución alélica prevista se aproxime más al patrón de distribución alélica observada. En una realización, la muestra combinada se obtiene de sangre materna, plasma o suero materno. En una realización, la muestra combinada de ADN puede ser enriquecida preferentemente en diversos loci polimórficos. En una realización, el enriquecimiento preferente se realiza de forma que se minimiza el sesgo alélico. En una realización, la presente divulgación se refiere a una composición de ADN que ha sido preferentemente enriquecido en diversos loci de forma que el sesgo alélico es reducido. En una realización, la distribución o distribuciones alélicas se miden secuenciando el ADN de la muestra combinada. En una realización, el modelo de distribución conjunto asume que los alelos se distribuirán de forma binomial. En una realización, el conjunto de distribuciones alélicas conjuntas previstas se crean para loci genéticamente vinculados, al tiempo que se tienen en cuenta las frecuencias de recombinación existentes de diversas fuentes, por ejemplo, utilizando datos del International HapMap Consortium.

En una realización, la presente divulgación proporciona métodos para el diagnóstico prenatal no invasivo (NPD), en particular, determinando el estado de aneuploidía de un feto mediante la observación de las mediciones alélicas en diversos loci polimórficos de los datos genotípicos medidos en las mezclas de ADN, donde determinadas mediciones alélicas son indicativas de un feto aneuploide, mientras que otras mediciones alélicas son indicativas de un feto euploide. En una realización, los datos genotípicos se miden secuenciando mezclas de ADN obtenidas de plasma materno. En una realización, la muestra de ADN puede ser preferiblemente enriquecida con moléculas de ADN que corresponden a la pluralidad de loci cuyas distribuciones alélicas se están calculando. En una realización, una muestra de ADN que comprende solo o prácticamente solo material genético de la madre y posiblemente también una muestra de ADN que comprende solo o prácticamente solo material genético del padre se someten a medición. En una realización, las mediciones genéticas de uno o los dos progenitores junto con la fracción fetal estimada se utilizan para crear una pluralidad de distribuciones alélicas previstas correspondientes a diferentes estados genéticos subyacentes posibles del feto; las distribuciones alélicas previstas se pueden denominar hipótesis. En una realización, los datos genéticos maternos no se determinan midiendo el material genético que pertenece exclusiva o casi exclusivamente a la madre, sino que se estiman a partir de las mediciones genéticas realizadas con el plasma materno que comprende una mezcla de ADN materno y fetal. En algunas realizaciones, las hipótesis pueden comprender la ploidía del feto en uno o más cromosomas, donde los segmentos de dichos cromosomas del feto fueron heredados de los progenitores, y combinaciones de estos. En algunas realizaciones, el estado de ploidía del feto se determina comparando las mediciones alélicas observadas con las diferentes hipótesis, donde al menos algunas de las hipótesis corresponden a diferentes estados de ploidía, y seleccionando el estado de ploidía que corresponde a la hipótesis que es más probable que sea cierta dadas las mediciones alélicas observadas. En una realización, este método implica el uso de datos de mediciones alélicas de algunos o todos los SNP medidos, con independencia de que los loci sean homocigotos o heterocigotos, y por tanto no implica el uso de alelos en loci que son únicamente heterocigotos. Este método puede no resultar apropiado para situaciones en las que los datos genéticos pertenecen solo a un locus polimórfico. Este método resulta particularmente ventajoso cuando los datos genéticos comprenden datos para más de 10 loci polimórficos de un cromosoma diana o más de 20 loci polimórficos. Este método resulta especialmente ventajoso cuando los datos genéticos comprenden datos para más de 50 loci polimórficos para un cromosoma diana, más de 100 loci polimórficos o más de 200 loci polimórficos para un cromosoma diana. En algunas realizaciones, los datos genéticos pueden comprender datos para más de 500 loci polimórficos de un cromosoma diana, más de1000 loci polimórficos, más de 2000 loci polimórficos o más de 5000 loci polimórficos para un cromosoma diana.

En una realización, un método divulgado en el presente documento utiliza técnicas de enriquecimiento selectivas que preservan las frecuencias alélicas relativas que se encuentran presentes en la muestra original de ADN en cada locus polimórfico de un conjunto de loci polimórficos. En algunas realizaciones, la técnica de amplificación y/o enriquecimiento selectivo puede implicar una PCR, como una PCR mediada por unión, captura de un fragmento mediante hibridación, sondas de inversión molecular u otras sondas de circularización. En algunas realizaciones, los métodos para la amplificación o el enriquecimiento selectivo pueden implicar el uso de sondas, donde, tras la correcta hibridación con la secuencia diana, el extremo 3' o el extremo 5' de una sonda de nucleótidos está separado del punto polimórfico del alelo por un pequeño número de nucleótidos. Esta separación reduce la amplificación preferente de un alelo, lo que se denomina sesgo alélico. Esto supone una mejora respecto de los métodos que implican el uso de sondas en las que el extremo '3 y el extremo 5' de una sonda correctamente hibridada se encuentran directamente adyacentes o muy cerca del punto polimórfico de un alelo. En una realización, las sondas en las que la región de hibridación puede contener o contiene un punto polimórfico son excluidas. Los puntos polimórficos del punto de hibridación pueden causar una hibridación desigual o inhibir directamente la hibridación en algunos alelos, lo que resulta en una amplificación preferente de determinados alelos. Estas realizaciones implican mejoras con respecto a otros métodos que conllevan una amplificación focalizada y/o un enriquecimiento selectivo,

en el sentido de que preservan mejor las frecuencias alélicas originales de la muestra en cada locus polimórfico, cuando la muestra es una muestra genómica pura de un único individuo o una mezcla de individuos.

Una realización del método de la invención utiliza la PCR focalizada altamente multiplexada y altamente eficiente para amplificar ADN, seguida de una secuenciación de alto rendimiento para determinar las frecuencias alélicas en cada locus diana. La capacidad para multiplexar más de unos 50 o 100 cebadores para PCR en una reacción de forma que la mayoría de la secuencia resultante represente una lectura del mapa de los loci diana es algo novedoso y no obvio. Una técnica que permite realizar una PCR focalizada altamente multiplexada de forma altamente eficiente implica el diseño de cebadores que es poco probable que se hibriden entre sí. Las sondas para PCR, típicamente denominadas cebadores, se seleccionan creando un modelo termodinámico de interacciones potencialmente adversas entre al menos 500, al menos 1000, al menos 5000, al menos 10 000, al menos 20 000, al menos 50 000 o al menos 100 000 pares de cebadores potenciales, o interacciones imprevistas entre los cebadores y el ADN de la muestra, y después utilizando el modelo para eliminar diseños que son incompatibles con otros diseños del conjunto. Otra técnica que permite realizar una PCR focalizada altamente multiplexada de forma altamente eficiente consiste en utilizar un método de anidado parcial o total en la PCR focalizada. El uso de uno o de una combinación de estos métodos permiten el multiplexado de al menos 300, al menos 800, al menos 1200, al menos 5000 o al menos 10 000 cebadores en un único conjunto con el ADN amplificado resultante que comprende una mayoría de moléculas de ADN que, cuando se secuencia, representará un mapa de los loci diana. El uso de un o una combinación de estos métodos permite el multiplexado de un gran número de cebadores en un único conjunto con el ADN amplificado resultante que comprende más de un 50%, más de un 80%, más de un 90%, más de un 95%, más de un 98% o más de un 99% de moléculas de ADN que representan un mapa de los loci diana.

En una realización, un método divulgado en el presente documento produce una medición cuantitativa del número de observaciones independientes de cada alelo en un locus polimórfico. Esto contrasta con la mayoría de los métodos, como la PCR cualitativa o de microarrays, que proporcionan información sobre el ratio de dos alelos, pero no cuantifican el número de observaciones independientes de ninguno de los alelos. Con los métodos que proporcionan información cuantitativa sobre el número de observaciones independientes, en los cálculos de la ploidía solo se utiliza el ratio, mientras que la información cuantitativa en sí no resulta útil. Para ilustrar la importancia de retener información sobre el número de observaciones independientes, tendremos en cuenta el locus de la muestra con dos alelos, A y B. En un primer experimento se observan 20 alelos A y 20 alelos B, y en un segundo experimento se observan 200 alelos A y 200 alelos B. En ambos experimentos el ratio (A/(A+B)) es igual a 0,5; sin embargo, el segundo experimento proporciona más información que el primero acerca de la certidumbre de la frecuencia del alelo A o B. Algunos métodos conocidos en la técnica implican determinar una media o sumar los ratios de los alelos (ratios del canal) (es decir, x/y) de los alelos individuales y analizar este ratio, sea comparándolo con un cromosoma de referencia o utilizando una regla sobre cómo se espera que este ratio se comporte en situaciones concretas. Estos métodos conocidos en la técnica no implican ninguna ponderación de los alelos y asumen que se puede garantizar aproximadamente la misma cantidad de producto de la PCR para cada alelo y que todos los alelos se deberían comportar del mismo modo. Este método presenta una serie de desventajas y, lo que es más importante, impide el uso de una serie de mejoras que se describen en la presente divulgación.

En una realización, un método divulgado en el presente documento establece un modelo explícitamente de las distribuciones de la frecuencia alélica prevista en la disomia, así como una pluralidad de distribuciones de la frecuencia alélica que se pueden esperar en caso de trisomía resultante de la ausencia de disyunción durante la meiosis I, la ausencia de disyunción durante la meiosis II, y/o la ausencia de disyunción durante la mitosis temprana en el desarrollo fetal. Para ilustrar por qué esto es importante, ponemos por ejemplo, un caso en el que no se produjeron cruces: la ausencia de disyunción durante la meiosis I resultaría en una trisomía en la que dos homólogos diferentes se heredaron de un progenitor; por el contrario, la ausencia de disyunción durante la meiosis II o la mitosis temprana en el desarrollo fetal resultaría en dos copias del mismo homólogo de un progenitor. Cada escenario daría como resultado unas frecuencias alélicas previstas diferentes en cada locus polimórfico y también en todos los loci considerados conjuntamente, debido a la unión genética. Los cruces, que provocan el intercambio de material genético entre homólogos, hacen que el patrón de herencia sea más complejo; en una realización, el método instantáneo tiene esto en cuenta utilizando información sobre la tasa de recombinación además de la distancia física entre loci. En una realización, para permitir una distinción mejorada entre la ausencia de disyunción de la meiosis I y la ausencia de disyunción de la meiosis II o la mitosis, el método instantáneo incorpora en el modelo una probabilidad creciente de cruce dado que la distancia desde el centrómero aumenta. La ausencia de disyunción de la meiosis II y la mitosis se puede distinguir por el hecho de que la ausencia de disyunción mitótica típicamente resulta en copias idénticas o prácticamente idénticas de un homólogo, mientras que los dos homólogos presentes tras un evento de ausencia de disyunción de la meiosis II a menudo difieren debido a uno o más cruces durante la gametogénesis.

En algunas realizaciones, un método divulgado en el presente documento implica la comparación de las mediciones alélicas observadas con las hipótesis teóricas correspondientes a una posible aneuploidía genética fetal, y no implica un paso de cuantificación del ratio de alelos en un locus heterocigoto. Cuando el número de loci es inferior a unos 20, la determinación de la ploidía realizada utilizando un método que comprende la cuantificación de un ratio de alelos en un locus heterocigoto y una determinación de la ploidía realizada utilizando un método consistente en comparar las mediciones alélicas observadas con las hipótesis de distribución alélica teóricas correspondientes a posibles estados genéticos fetales pueden dar un resultado similar. Sin embargo, cuando el número de loci es

superior a 50, es probable que estos dos métodos proporcionen resultados muy diferentes; cuando el número de loci es superior a 400, superior a 1000, o superior a 2000, es muy probable que estos dos métodos den resultados con unas diferencias cada vez más significativas. Estas diferencias se deben al hecho de que un método que comprende la cuantificación de un ratio de alelos en un locus heterocigoto sin medir la magnitud de cada alelo independientemente y agregar o determinar la media de los ratios impide el uso de técnicas como el uso de un modelo de distribución conjunto, realizar un análisis de enlaces, utilizar un modelo de distribución binomial, y/u otras técnicas estadísticas avanzadas, mientras que el uso de un método que comprende la comparación de las mediciones alélicas observadas con hipótesis de distribución alélica teóricas correspondientes a posibles estados genéticos fetales permite utilizar estas técnicas que pueden aumentar de forma significativa la precisión de la determinación.

En una realización, un método divulgado en el presente documento implica la determinación de si la distribución de mediciones alélicas observadas es indicativa de un feto euploide o aneuploide utilizando un modelo de distribución conjunto. El uso de un modelo de distribución conjunto es diferente y representa una mejora significativa respecto de los métodos que determinan las tasas de heterocigosidad tratando los loci polimórficos de forma independiente, y las determinaciones resultantes son significativamente más precisas. Sin intención de vincularse a ninguna teoría concreta, se cree que una razón por la que tienen una mayor precisión es que el modelo de distribución conjunto tiene en cuenta el enlace entre SNP, y la probabilidad de que se hayan producido cruces durante la meiosis que dio lugar a los gametos que formaron el embrión que creció hasta convertirse en feto. El propósito de utilizar el concepto de enlace a la hora de crear la distribución prevista de las mediciones alélicas para una o más hipótesis es que permite la creación de distribuciones de las mediciones alélicas previstas que se corresponden con la realidad considerablemente mejor que cuando no se utiliza el enlace. Por ejemplo, imaginemos que hay dos SNP, 1 y 2, cercanos entre sí, y la madre es A en el SNP 1 y A en el SNP 2 de un homólogo, y B en SN1 y B en SNP 1 del homólogo dos. Si el padre es A en los dos SNP de los dos homólogos, y se mide una B para el SNP1 del feto, esto indica que el homólogo dos ha sido heredado por el feto y, por tanto, tiene muchas más probabilidades de que B esté presente en el feto en SNP 2. Un modelo que tiene en cuenta el enlace prevería esto, mientras que un modelo que no lo tiene en cuenta no. Alternativamente, si una madre era AB en SNAP 1 y AB en SNP 2 cercano, entonces se podrían utilizar dos hipótesis correspondientes a la trisomía materna en ese lugar: una que implicaría un error en la copia coincidente (ausencia de disyunción en la meiosis II o mitosis en el desarrollo fetal temprano) y una que implicaría un error en la copia no coincidente (ausencia de disyunción en la meiosis I). En el caso de la trisomía por error en la copia coincidente, si el feto heredase AA de la madre en SNP 1, entonces el feto tendría muchas más posibilidades de heredar AA o BB de la madre en SNP 2, pero no AB. En el caso del error en la copia no coincidente, el feto heredaría AB de la madre en los dos SNP. Las hipótesis de distribución alélica realizadas con un método de determinación del estado de ploidía que tiene en cuenta el enlace realizarían estas predicciones y, por tanto, se corresponden con las mediciones alélicas reales en una medida mucho mayor que en el caso de un método de determinación del estado de ploidía que no tiene en cuenta el enlace. Cabe señalar que el método del enlace no resulta posible cuando se utiliza un método que depende del cálculo de los ratios de alelos y en la suma de dichos ratios de alelos.

Una razón por la que se cree que las determinaciones del estado de ploidía que utilizan un método que comprende la comparación de las mediciones alélicas observadas con las hipótesis teóricas correspondientes a posibles estados genéticos fetales ofrecen una mayor precisión es que cuando se utiliza la secuenciación para medir los alelos, este método puede averiguar más información que otros métodos de los datos de los alelos cuando el número total de lecturas es bajo; un método que depende del cálculo y la suma de los ratios de alelos produciría un ruido estocástico desproporcionadamente ponderado. Por ejemplo, imaginemos un caso que implicase la medición de los alelos utilizando secuenciación y en el que hubiese un conjunto de loci en el que solo se detectasen cinco lecturas de secuencias para cada locus. En una realización, para cada uno de los alelos, los datos se pueden comparar con las hipótesis de distribución alélica, y ponderarse en función del número de lecturas de secuencias; por tanto, los datos de estas mediciones serían convenientemente ponderados e incorporados a la determinación total. Esto contrasta con un método que implica la cuantificación de un ratio de alelos en un locus heterocigoto, dado que este método solo podría calcular ratios de 0%, 20%, 40%, 60%, 80% o 100% como ratios de alelos posibles; ninguno de estos puede acercarse a los ratios de alelos previstos. En este último caso, los ratios de alelos calculados tendrían que ser descartados debido a lecturas insuficientes o tendrían una ponderación desproporcionada e introducirían un ruido estocástico en la determinación, reduciendo así su precisión. En una realización, las mediciones alélicas individuales pueden ser tratadas como mediciones independientes, donde la relación entre las mediciones realizadas en alelos del mismo locus no es diferente de la relación entre las mediciones realizadas en alelos de diferentes loci.

En una realización, un método divulgado en el presente documento implica la determinación de si la distribución de las mediciones alélicas observadas es indicativa de un feto euploide o aneuploide sin comparar ninguna métrica con las mediciones alélicas observadas en un cromosoma de referencia que se espera que sea disómico (denominado método RC). Esto supone una mejora significativa respecto de otros métodos, como los métodos que utilizan una secuenciación por fuerza bruta que detectan la aneuploidía evaluando la proporción de fragmentos aleatoriamente secuenciados de un cromosoma sospechoso con respecto a uno o más cromosomas de referencia que se suponen disómicos. El método RC produce resultados incorrectos si el cromosoma de referencia supuestamente disómico no lo es en realidad. Esto puede ocurrir en casos en los que la aneuploidía es más sustancial de la trisomía de un único cromosoma o cuando el feto es triploide y todos los autosomas son trisómicos. En el caso de feto triploide femenino

(69, XXX) en efecto no existen cromosomas disómicos en absoluto. El método descrito en el presente documento no requiere un cromosoma de referencia y podría identificar correctamente cromosomas trisómicos en fetos triploides femeninos. Para cada cromosoma, hipótesis, fracción del niño y nivel de ruido, se puede adaptar un modelo de distribución conjunto, sin datos de un cromosoma de referencia, sin una estimación de la fracción del niño y sin una hipótesis de referencia establecida.

En una realización, un método divulgado en el presente documento demuestra cómo la observación de distribuciones alélicas en loci polimórficos se puede utilizar para determinar el estado de ploidía de un feto con una mayor precisión que los métodos existentes en la técnica. En una realización, el método utiliza la secuenciación focalizada para obtener genotipos materno-fetales mixtos y opcionalmente genotipos de la madre y/o del padre en diversos SNP para establecer primero las diversas distribuciones de la frecuencia alélica previstas bajo las diferentes hipótesis y, a continuación, observar la información alélica cuantitativa obtenida sobre la mezcla materno-fetal y evaluar qué hipótesis se ajusta mejor a los datos, donde el estado genético correspondiente a la hipótesis que más se ajusta a los datos se denomina el estado genético correcto. En una realización, un método divulgado en el presente documento también utiliza el grado de adecuación para generar una certeza en que el estado genético determinado es el estado genético correcto.

En una realización, un método divulgado en el presente documento implica el uso de algoritmos que analizan la distribución de alelos encontrados para los loci que tienen distintos contextos parentales y la comparación de las distribuciones alélicas observadas con las distribuciones alélicas previstas para los diferentes estados de ploidía para los diferentes contextos parentales (diferentes patrones genotípicos parentales). Esto difiere y supone una mejora respecto de los métodos que no permiten la estimación del número de casos independientes de cada alelo en cada locus en una muestra combinada materno-fetal. En una realización, un método divulgado en el presente documento implica la determinación de si la distribución de mediciones alélicas observadas es indicativa de un feto euploide o aneuploide utilizando distribuciones alélicas observadas en loci donde la madre es heterocigótica. Esto difiere y supone una mejora respecto de los métodos que no utilizan las distribuciones alélicas observadas en loci donde la madre es heterocigótica porque, en los casos en los que el ADN no es preferentemente enriquecido o es preferentemente enriquecido para loci que se sabe que son altamente informativos para ese individuo diana concreto, permite el uso de aproximadamente el doble de datos de medición genética de un conjunto de datos de la secuencia en la determinación del estado de ploidía, lo que se traduce en una determinación más precisa.

En una realización, un método divulgado en el presente documento utiliza un modelo de distribución conjunto que asume que las frecuencias alélicas en cada locus son de naturaleza multinomial (y por tanto binomiales cuando los SNP son bialélicos). En algunas realizaciones el modelo de distribución conjunto utiliza distribuciones beta-binomiales. Cuando el uso de una técnica de medición, como la secuenciación, proporciona una medida cuantitativa para cada alelo presente en cada locus, el modelo binomial se puede aplicar a cada locus y se puede averiguar el grado de las frecuencias alélicas subyacentes y la certeza en dicha frecuencia. Con los métodos conocidos en la técnica que generan determinaciones del estado de ploidía a partir de ratios de alelos o los métodos en los que la información cuantitativa de los alelos es desechada, no resulta posible averiguar la certidumbre del ratio observado. El método instantáneo es diferente y supone una mejora respecto de los métodos que calculan ratios de alelos y suman estos ratios para realizar una determinación del estado de ploidía, puesto que cualquier método que implica el cálculo de un ratio de alelos en un locus completo y la suma de dichos ratios necesariamente asume que las intensidades o los recuentos medidos que son indicativos de la cantidad de ADN de cualquier alelo o locus concreto se distribuirán de forma gaussiana. El método divulgado en el presente documento no implica el cálculo de ratios de alelos. En algunas realizaciones, un método divulgado en el presente documento puede implicar la incorporación del número de observaciones de cada alelo de una pluralidad de loci a un modelo. En algunas realizaciones, un método divulgado en el presente documento puede implicar el cálculo de las propias distribuciones previstas, permitiendo el uso de un modelo de distribución binomial que puede ser más preciso que cualquier modelo que asuma una distribución gaussiana de las mediciones alélicas. La probabilidad de que el modelo de distribución binomial sea significativamente más preciso que la distribución gaussiana se incrementa a medida que aumenta el número de loci. Por ejemplo, cuando se analizan menos de 20 loci, la probabilidad de que el modelo de distribución binomial sea significativamente mejor es baja. Sin embargo, cuando se utilizan más de 100, o especialmente más de 400, o especialmente más de 1000, o especialmente más de 2000 loci, el modelo de distribución binominal tiene una probabilidad muy elevada de ser significativamente más preciso que el modelo de distribución gaussiana, lo que resulta en una determinación del estado de ploidía más preciso. La probabilidad de que el modelo de distribución binomial sea significativamente más preciso que la distribución gaussiana también se incrementa a medida que aumenta el número de observaciones en cada locus. Por ejemplo, cuando se observan menos de 10 secuencias distintas en cada locus, la probabilidad de que el modelo de distribución binomial sea significativamente mejor es baja. Sin embargo, cuando se utilizan más de 50 lecturas de secuencias, o especialmente más de 100 lecturas de secuencias, o especialmente más de 200 lecturas de secuencias, o especialmente más de 300 lecturas de secuencias, el modelo de distribución binominal tiene una probabilidad muy elevada de ser significativamente más preciso que el modelo de distribución gaussiana, lo que resulta en una determinación del estado de ploidía más preciso.

En una realización, un método divulgado en el presente documento utiliza la secuenciación para medir el número de casos de cada alelo en cada locus de una muestra de ADN. Cada lectura de secuenciación se puede correlacionar con un locus concreto y tratarse como una lectura de secuencia binaria; alternativamente, la probabilidad de la

identidad de la lectura y/o correlación se puede incorporar como parte de la lectura de la secuencia, resultando en una lectura de la secuencia probabilística que es el número entero o la fracción probable de lecturas de la secuencia que se correlaciona con unos determinados loci. Utilizando recuentos binarios o probabilidad de recuentos se puede utilizar una distribución binomial para cada conjunto de mediciones, lo que permite calcular un intervalo de certeza alrededor del número de recuentos. Esta capacidad de utilizar la distribución binomial permite realizar estimaciones de la ploidía más precisas y calcular intervalos de certeza más precisos. Esto difiere y supone una mejora respecto de otros métodos que utilizan intensidades para medir la cantidad de un alelo presente, como los métodos que utilizan microarrays o los métodos que realizan mediciones utilizando lectores de fluorescencia para medir la intensidad de ADN con etiquetas fluorescentes en bandas electroforéticas.

En una realización, un método divulgado en el presente documento utiliza aspectos del conjunto de datos presentes para determinar parámetros para la distribución de la frecuencia alélica estimada para ese conjunto de datos. Esto supone una mejora respecto de los métodos que utilizan conjuntos de datos de formación o conjuntos de datos previos para establecer parámetros para determinar las distribuciones de la frecuencia alélica prevista presente o posiblemente los ratios de alelos previstos. Esto se debe a que hay diferentes series de condiciones implicadas en la recogida y medición de cada muestra genética y, por tanto, un método que utiliza datos del conjunto de datos instantáneo para determinar los parámetros para el modelo de distribución conjunto que se utilizará en la determinación del estado de ploidía para esa muestra tenderá a ser más preciso.

En una realización, un método divulgado en el presente documento implica la determinación de si la distribución de mediciones alélicas observadas es indicativa de un feto euploide o aneuploide utilizando una técnica de probabilidad máxima. El uso de la técnica de la probabilidad máxima es diferente y supone una mejora significativa respecto de los métodos que utilizan la técnica de rechazo por hipótesis única en el sentido de que las determinaciones resultantes se realizarán con una precisión significativamente mayor. Un motivo de ello es que las técnicas de rechazo por hipótesis única establecen umbrales de corte basados en una única distribución de la medición y no en dos, lo que significa que normalmente los umbrales no son óptimos. Otra razón es que la técnica de la probabilidad máxima permite la optimización del umbral de corte para cada muestra individual en lugar determinar un umbral de corte que se utilizará para todas las muestras con independencia de las características concretas de cada muestra individual. Otra razón es que el uso de la técnica de la probabilidad máxima permite el cálculo de una certeza para cada determinación del estado de ploidía. La capacidad para realizar un cálculo de certeza para cada determinación permite a un médico saber qué determinaciones resultan precisas y cuáles es más probable que sean erróneas. En algunas realizaciones, se puede combinar una amplia variedad de métodos con una técnica de estimación de la probabilidad máxima para mejorar la precisión de las determinaciones del estado de ploidía. En una realización, la técnica de la probabilidad máxima se puede utilizar en combinación con el método descrito en la Patente USA 7 888 017. En una realización, la técnica de la probabilidad máxima se puede utilizar en combinación con el método de utilizar la amplificación por PCR focalizada para amplificar el ADN en la muestra combinada seguida por la secuenciación y el análisis utilizando un método de recuento de lecturas como el utilizado por TANDEM DIAGNOSCS, presentado en el Congreso Internacional de Genética Humana celebrado en Montreal en octubre de 2011. En una realización, un método divulgado en el presente documento implica la estimación de la fracción fetal de ADN en la muestra combinada y la utilización de dicha estimación para calcular tanto el estado de ploidía como la certeza del estado de ploidía. Cabe señalar que esto es diferente y distinto de los métodos que utilizan la fracción fetal estimada como filtro para determinar la fracción fetal suficiente, seguida de una determinación de la ploidía realizada utilizando la técnica de rechazo por hipótesis única que no tiene en cuenta la fracción fetal ni produce un cálculo de la certeza para la determinación.

En una realización, un método divulgado en el presente documento tiene en cuenta la tendencia de los datos a ser ruidosos y a contener errores asignando una probabilidad a cada medición. El uso de técnicas de probabilidad máxima para seleccionar la hipótesis correcta del conjunto de hipótesis que se realizaron utilizando los datos de medición con estimaciones probabilísticas asignadas hace que sea más probable que se descuenten las mediciones incorrectas y que se utilicen las mediciones correctas en los cálculos que conducen a la determinación de la ploidía. Para ser más precisos, este método reduce de forma sistemática la influencia de los datos que son incorrectamente medidos sobre la determinación del estado de ploidía. Esto representa una mejora respecto de los métodos en los que se asume que todos los datos son igualmente correctos o de los métodos en los que los datos periféricos son excluidos de forma arbitraria de los cálculos que conducen a la determinación del estado de ploidía. Los métodos existentes que utilizan mediciones del ratio del canal afirman ampliar el método a los SNP múltiples estableciendo una media de los ratios del canal de los SNP individuales. El hecho de no ponderar los SNP individuales por la varianza de la medición prevista basada en la calidad del SNP y en la profundidad observada de la lectura reduce la precisión de la estadística resultante, lo que provoca una reducción significativa de la precisión de la determinación del estado de ploidía, en particular en los casos dudosos.

En una realización, un método divulgado en el presente documento no presupone el conocimiento de los SNP u otros loci polimórficos que son heterocigotos en el feto. Este método permite realizar una determinación del estado de ploidía en los casos en los que no se dispone de información genotípica paterna. Esto supone una mejora respecto de los métodos en los que es necesario conocer cuáles de los SNP son heterocigotos de antemano para seleccionar convenientemente los loci diana o para interpretar las mediciones genéticas realizadas sobre la muestra de ADN fetal/materno mezclada.

Los métodos descritos en el presente documento resultan particularmente ventajosos cuando se utilizan con muestras en las que hay una pequeña cantidad de ADN disponible o cuando el porcentaje de ADN fetal es bajo. Esto se debe a la correspondiente tasa de pérdida de alelos más elevada que se produce cuando solo hay una pequeña cantidad de ADN disponible y/o la correspondiente tasa de pérdida de alelos fetales más elevada que se produce cuando el porcentaje de ADN fetal es bajo en una muestra combinada de ADN fetal y materno. Una tasa de pérdida de alelos elevada, lo que significa que un importante porcentaje de los alelos no se han medido para el individuo diana, resulta en un mal cálculo de las fracciones fetales precisas y en una mala determinación del estado de ploidía preciso. Dado que los métodos divulgados en el presente documento pueden utilizar un modelo de distribución conjunto que tiene en cuenta el enlace en los patrones de herencia entre SNP, se pueden realizar determinaciones del estado de ploidía notablemente más precisas. Los métodos descritos en el presente documento permiten realizar una determinación precisa del estado de ploidía cuando el porcentaje de moléculas de ADN que son fetales en la mezcla es inferior al 40%, inferior al 30%, inferior al 20%, inferior al 10%, inferior al 8% e incluso inferior al 6%.

En una realización de la divulgación, resulta posible determinar el estado de ploidía de un individuo basándose en las mediciones, cuando el ADN de dicho individuo está mezclado con ADN de un individuo relacionado. En una realización, la mezcla de ADN es el ADN flotante libre que se encuentra en el plasma materno, que puede incluir ADN de la madre, con un cariotipo conocido y un genotipo conocido, y que puede estar mezclado con ADN del feto, con cariotipo conocido y genotipo conocido. Resulta posible utilizar la información genotípica conocida de uno o los dos progenitores para predecir una pluralidad de potenciales estados genéticos del ADN de la muestra combinada para diferentes estados de ploidía, diferentes contribuciones cromosómicas de cada uno de los progenitores al feto y, opcionalmente, diferentes fracciones de ADN fetal en la mezcla. Cada composición potencial se puede referir como una hipótesis. El estado de ploidía del feto se puede determinar a continuación analizando las mediciones reales y determinando qué potenciales composiciones resultan más probables a tenor de los datos observados.

En algunas realizaciones, se podría utilizar un método divulgado en el presente documento en situaciones en las que hay una cantidad muy reducida de ADN presente, como en la fertilización in vitro o en situaciones forenses, cuando se dispone de una o muy pocas células (normalmente menos de 10 células, menos de 20 células o menos de 40 células). En estas realizaciones, un método divulgado en el presente documento sirve para realizar determinaciones del estado de ploidía a partir de una pequeña cantidad de ADN que no está contaminado con otro ADN, pero en las que la determinación del estado de ploidía resulta muy difícil dada la reducida cantidad de ADN. En algunas realizaciones, un método divulgado en el presente documento se podría utilizar en situaciones en las que el ADN diana está contaminado con ADN de otro individuo, por ejemplo, en la sangre materna en el contexto del diagnóstico prenatal, pruebas de paternidad o productos de pruebas de concepción. Otras situaciones en las que estos métodos resultarían particularmente ventajosos serían en las pruebas del cáncer cuando solamente hay una o un pequeño número de células presente entre una mayor cantidad de células normales. Las mediciones genéticas utilizadas como parte de estos métodos se podrían realizar en cualquier muestra que comprenda ADN o ARN, por ejemplo, a título meramente enunciativo: sangre, plasma, fluidos corporales, orina, cabello, lágrimas, saliva, tejido, piel, uñas, blastómeros, embriones, líquido amniótico, muestras de vello coriónico, heces, bilis, linfa, moco cervical, semen u otras células o materiales que comprenden ácidos nucleicos. En una realización, un método divulgado en el presente documento se podría realizar con métodos de detección de ácido nucleico, como la secuenciación, microarrays, qPCR, PCR digital u otros métodos utilizados para medir ácidos nucleicos. Si por alguna razón se considerase deseable, los ratios de las probabilidades del recuento de alelos en un locus se podrían calcular y los ratios de alelos se podrían utilizar para determinar el estado de ploidía en combinación con algunos de los métodos descritos en el presente documento, siempre que los métodos sean compatibles. En algunas realizaciones, un método divulgado en el presente documento implica el cálculo, por ordenador, de ratios de alelos en la pluralidad de loci polimórficos de las mediciones de ADN realizadas con las muestras procesadas. En algunas realizaciones, un método divulgado en el presente documento implica el cálculo, por ordenador, de ratios de alelos en la pluralidad de loci polimórficos de las mediciones de ADN realizadas con las muestras procesadas junto con cualquier combinación de otras mejoras descritas en esta divulgación.

En otra parte de este documento se recoge una exposición más detallada de los anteriores puntos.

*Diagnóstico prenatal no invasivo (NPD)*

El proceso del diagnóstico prenatal no invasivo de conformidad con la presente divulgación implica una serie de pasos. Algunos de los pasos pueden incluir: (1) obtención de material genético del feto; (2) enriquecimiento del material genético del feto que puede estar en una muestra combinada, ex vivo; (3) amplificación del material genético, ex vivo; (4) enriquecimiento preferente de loci específicos en el material genético, ex vivo; (5) medición del material genético, ex vivo; y (6) análisis de los datos genotípicos, por ordenador y ex vivo. Los métodos para reducir la práctica de estos seis pasos y otros relevantes se describen en el presente documento. Al menos algunos de los pasos del método no se aplican directamente sobre el organismo. En una realización, la presente divulgación se refiere a métodos de tratamiento y diagnóstico aplicados al tejido y otros materiales biológicos aislados y separados del organismo. Al menos algunos de los pasos del método se ejecutan por ordenador.

Algunas realizaciones de la presente divulgación permiten a un médico determinar el estado genético de un feto que se está gestando en una madre de forma no invasiva, de forma que la salud del bebé no se pone en peligro por la recogida del material genético del feto y que la madre no se tiene que someter a un procedimiento invasivo. Por otra

parte, en determinados aspectos, la presente divulgación permite determinar el estado genético fetal con una elevada precisión, una precisión significativamente mayor que, por ejemplo, las pruebas basadas en el analito del suero materno no invasivas, como la triple prueba, que se utiliza generalmente en la atención prenatal.

La alta precisión de los métodos divulgados en el presente documento es el resultado de un método informático para el análisis de los datos del genotipo, tal y como se describe en el presente documento. Los avances tecnológicos recientes han permitido medir grandes cantidades de información genética a partir de una muestra genética utilizando métodos como la secuenciación de alto rendimiento y los arrays para la determinación del genotipo. Los métodos divulgados en el presente documento permiten a un médico sacar mayor partido de las grandes cantidades de datos disponibles y realizar un diagnóstico más preciso del estado genético fetal. Los detalles de una serie de realizaciones se recogen más abajo. Las diferentes realizaciones pueden implicar diferentes combinaciones de los mencionados pasos. Se pueden utilizar de forma intercambiable diversas combinaciones de las diferentes realizaciones de los distintos pasos.

En una realización de la divulgación, se toma una muestra de sangre de una madre embarazada, y el ADN flotante libre en el plasma de la sangre de la madre, que contiene una mezcla de ADN de origen materno y ADN de origen fetal, que es aislado y utilizado para determinar el estado de ploidía del feto. En una realización, un método divulgado en el presente documento implica el enriquecimiento preferente de las secuencias de ADN en una mezcla de ADN que corresponden a los alelos polimórficos de forma que los ratios de alelos y/o las distribuciones de alelos se mantienen fundamentalmente constantes tras el enriquecimiento. En una realización, un método divulgado en el presente documento implica la amplificación basada en la PCR focalizada de alta eficiencia de forma que un porcentaje muy elevado de las moléculas resultantes corresponden a los loci diana. En una realización, un método divulgado en el presente documento implica la secuenciación de una mezcla de ADN que contiene tanto ADN de origen materno como ADN de origen fetal. En una realización, un método divulgado en el presente documento implica el uso de distribuciones de alelos medidas para determinar el estado de ploidía de un feto que se está gestando en una madre. En una realización, un método divulgado en el presente documento implica la comunicación del estado de ploidía determinado a un médico. En una realización, un método divulgado en el presente documento implica emprender una acción clínica, por ejemplo, realizar pruebas de seguimiento invasivas como una amniocentesis o muestreo de vello coriónico, preparación para el nacimiento de un individuo trisómico o la terminación electiva de un feto trisómico.

Esta solicitud hace referencia a la Solicitud de Patente USA con el número de serie 11/603 406, presentada el 28 de noviembre de 2006 (Publicación USA n.º 20070184467); la Solicitud de Patente USA con el número de serie 12/076 348, presentada el 17 de marzo de 2008 (Publicación USA n.º 20080243398); la Solicitud de Patente PCT con el número de serie PCT/US09/52730, presentada el 4 de agosto de 2009 (Publicación PCT n.º W0/2010/017214); la Solicitud de Patente PCT con el número de serie PCT/US10/050824, presentada el 30 de septiembre de 2010 (Publicación PCT n.º W0/2011/041485), y la Solicitud de Patente USA con el número de serie 13/110 685, presentada el 18 de mayo de 2011. Parte del vocabulario utilizado en este documento puede tener sus antecedentes en estas referencias. Parte de los conceptos descritos en el presente documento se pueden entender mejor a la luz de los conceptos que se encuentran en estas referencias.

*Análisis de sangre materna que contiene ADN fetal flotante libre*

Los métodos empleados en el presente documento se pueden utilizar para ayudar a determinar el genotipo de un niño, feto u otro individuo diana cuando el material genético de la diana se encuentra en presencia de una cantidad de otro material genético. En algunas realizaciones, el genotipo se puede referir al estado de ploidía de uno o una pluralidad de cromosomas, se puede referir a uno o una pluralidad de alelos vinculados a una enfermedad o alguna combinación de estos. En la presente divulgación, el debate se centra en determinar el estado genético de un feto cuando el ADN fetal se encuentra en la sangre materna, pero este ejemplo no pretende limitar los posibles contextos en los que se puede aplicar este método. Por otra parte, el método puede resultar aplicable en casos en los que la cantidad de ADN diana se encuentra en cualquier proporción con el ADN no diana; por ejemplo, el ADN diana podría representar cualquier proporción entre el 0,000001 y el 99,999999% del ADN presente. Asimismo, el ADN no diana no tiene por qué ser necesariamente de un individuo, o incluso de un individuo relacionado, siempre que los datos genéticos de alguno o todos los individuos no diana relevantes sean conocidos. En una realización, un método divulgado en el presente documento se puede utilizar para determinar los datos genotípicos de un feto a partir de la sangre materna que contiene ADN fetal. También se puede utilizar en un caso en el que hay múltiples fetos en el útero de una mujer embarazada o cuando pueda haber otro ADN contaminante en la muestra, por ejemplo, de otros hermanos ya nacidos.

Esta técnica puede hacer uso del fenómeno de las células sanguíneas fetales consiguiendo acceder a la circulación materna a través del vello coriónico. Normalmente solo un pequeño número de células fetales entra en la circulación materna de esta forma (no suficiente para producir una prueba de Kleihauer-Betke para la hemorragia fetal-materna). Las células fetales se pueden clasificar y analizar a través de diversas técnicas para buscar secuencias de ADN concretas, pero sin los riesgos que implican inherentemente los procedimientos invasivos. Esta técnica también puede hacer uso del fenómeno del ADN fetal flotante libre obteniendo acceso a la circulación materna a través de la liberación de ADN posterior a la apoptosis de tejido placentario cuando el tejido placentario en cuestión contiene ADN del mismo genotipo que el feto. Se ha demostrado que el ADN flotante libre que se encuentra en el plasma materno contiene ADN fetal en proporciones de hasta el 30-40% de ADN fetal.

En una realización de la divulgación, se puede extraer sangre a una mujer embarazada. La investigación ha demostrado que la sangre materna puede contener una pequeña cantidad de ADN flotante libre del feto, además de ADN flotante libre de origen materno. Por otra parte, también puede haber células sanguíneas fetales enucleadas que contienen ADN de origen fetal, además de múltiples células sanguíneas de origen materno, que típicamente no contienen ADN nuclear. Existen métodos conocidos en la técnica para aislar ADN fetal o crear fracciones enriquecidas en ADN fetal. Por ejemplo, se ha demostrado que la cromatografía crea determinadas fracciones enriquecidas en ADN fetal.

Una vez que se dispone de la muestra de sangre, plasma u otro fluido materno, extraído de manera relativamente no invasiva y que contiene una cantidad de ADN fetal, sea celular o flotante libre, enriquecido en su proporción respecto del ADN materno o en su proporción original, se puede determinar el genotipo del ADN encontrado en dicha muestra. En algunas realizaciones, la sangre se puede extraer utilizando una aguja para extraer sangre de una vena, por ejemplo, la vena basílica. El método descrito en el presente documento se puede utilizar para determinar los datos genotípicos del feto. Por ejemplo, se pueden utilizar para determinar el estado de ploidía de uno o más cromosomas, se pueden utilizar para determinar la identidad de uno o un conjunto de SNP, incluyendo inserciones, deleciones y translocaciones. Se pueden utilizar para determinar uno o más haplotipos, incluyendo el progenitor de origen de una o más características genotípicas.

Cabe señalar que este método funcionará con cualquiera de los ácidos nucleicos que se pueden utilizar para cualesquiera métodos de secuenciación o determinación del genotipo, como la plataforma ILLUMINA INFINIUM ARRAY, AFFYMETRIX GENECHIP, ILLUMINA GENOME ANALYZER, o LIFE TECHNOLGIES' SOLID SYSTEM. Esto incluye ADN flotante libre extraído del plasma o amplificaciones (por ejemplo, amplificación del genoma completo, PCR) de este; ADN genómico de otros tipos de células (por ejemplo, linfocitos humanos de sangre completa) o amplificaciones de este. Para la preparación del ADN, también funcionará cualquier extracción o método de purificación que genere ADN genómico adecuado para una de estas plataformas. Este método podría funcionar igualmente bien con muestras de ARN. En una realización, el almacenamiento de muestras se puede realizar de forma que se minimice la degradación (por ejemplo, mediante congelación a unos -20ºC o a una temperatura inferior).

*Parental Support*

Algunas realizaciones de la divulgación se pueden utilizar en combinación con el método PARENTAL SUPPORT™ (PS), cuyas realizaciones se describen en la Solicitud USA n.º 11/603 406 (Publicación USA n.º 20070184467), Solicitud USA n.º 12/076 348 (Publicación USA n.º: 20080243398), Solicitud USA 13/110 685, Solicitud PCT PCT/US09/52730 (Publicación PCT n.º W0/2010/017214), y Solicitud PCT n.º PCT/US10/050824 (Publicación PCT n.º W0/2011/041485) que se incorporan al presente por referencia en su totalidad. PARENTAL SUPPORT™ es un método basado en la informática que se puede utilizar para analizar datos genéticos. En algunas realizaciones, los métodos divulgados en el presente documento se pueden considerar parte del método PARENTAL SUPPORT™. En algunas realizaciones, el método PARENTAL SUPPORT ™ es un conjunto de métodos que pueden ser utilizados para determinar los datos genéticos del individuo diana, con alta precisión, de una célula o un número reducido de células de ese individuo, o una mezcla de ADN compuesto por ADN del individuo diana y ADN de uno o una pluralidad de otros individuos, específicamente para determinar alelos vinculados a enfermedad, otros alelos de interés, y/o el estado de ploidía de uno o una pluralidad de cromosomas en el individuo diana. PARENTAL SUPPORT™ se puede referir a cualquiera de estos métodos. PARENTAL SUPPORT™ es un ejemplo de método basado en la informática.

El método PARENTAL SUPPORT™ utiliza datos genéticos parentales conocidos; es decir, datos genéticos haplotípicos y/o diploides de la madre y/o el padre, junto con el conocimiento del mecanismo de la meiosis y la medición imperfecta del ADN diana, y posiblemente de uno o más individuos relacionados, junto con frecuencias de cruce basadas en la población, para reconstruir, in silico, el genotipo de una pluralidad de alelos, y/o el estado de ploidía de un embrión o de cualquier célula o células diana, y el ADN diana en la localización de loci clave con un alto grado de certeza. El método PARENTAL SUPPORT™ puede reconstruir no solamente polimorfismos de un solo nucleótido (SNP) que fueran medidos deficientemente, sino también inserciones y supresiones, y SNP o regiones completas de ADN que no fueron medidas en absoluto. Además, el método PARENTAL SUPPORT™ puede medir múltiples loci vinculados a enfermedad, y también realizar un cribado de aneuploidía, partiendo de una sola célula. En algunas realizaciones, el método PARENTAL SUPPORT™ puede ser utilizado para caracterizar una o más células de embriones biopsiados durante un ciclo de IVF para determinar la condición genética de una o más células.

El método PARENTALSUPPORT™ permite limpiar datos genéticos ruidosos. Esto puede hacerse deduciendo los correctos alelos genéticos en el genoma diana (embrión) utilizando el genotipo de individuos relacionados (padres) como referencia. El PARENTAL SUPPORT™ puede ser especialmente pertinente cuando se dispone solamente de una pequeña cantidad de material genético (ej. PGD) y donde las mediciones directas de los genotipos son inherentemente ruidosas, debido a las limitadas cantidades de material genético. El PARENTAL SUPPORT™ puede ser especialmente pertinente cuando solamente una pequeña fracción del material genético disponible procede del individuo diana (ej. NPD) y donde las mediciones directas de los genotipos son inherentemente ruidosas, debido a la señal del ADN contaminante de otro individuo. El método PARENTAL SUPPORT™ puede reconstruir secuencias de alelos diploides ordenadas con alta precisión en el embrión, junto con el número de copias de segmentos de

cromosomas, aunque las mediciones diploides convencionales no ordenadas pueden caracterizarse por elevadas tasas de pérdidas de alelos, inclusiones, sesgos de amplificación variable y otros errores. El método puede emplear un modelo genético subyacente, y un modelo subyacente de medición de error. El modelo genético puede determinar probabilidades de alelos en cada SNP y probabilidades de cruce entre SNP. Las probabilidades de los alelos pueden ser modeladas en cada SNP en base a datos obtenidos de los padres, y modelar probabilidades de cruce entre SNP en base a datos obtenidos de la base de datos HapMap, desarrollada por el International HapMap Project. Con el modelo genético subyacente y el modelo de error de medición, puede utilizarse una estimación *maximum a posteriori* (MAP), con modificaciones para la eficiencia informática, para calcular los valores de alelos ordenados correctos en cada SNP en el embrión.

En algunos casos las técnicas anteriormente mencionadas pueden determinar el genotipo de un individuo con una cantidad muy pequeña de ADN procedente de ese individuo. Este podría ser ADN de una o de un pequeño número de células o podría proceder de la pequeña cantidad de ADN fetal que se encuentra en la sangre materna.

*Definiciones*

*Polimorfismo de un solo nucleótido (SNP)* se refiere a un solo nucleótido que puede ser distinto entre los genomas de dos miembros de la misma especie. El empleo del término no debe implicar ninguna limitación en la frecuencia con que se produce cada variante.

Secuencia puede referirse a una secuencia de ADN o una secuencia genética. Puede referirse a la estructura primaria, física de la molécula o cadena de ADN en un individuo. Puede referirse a la secuencia de nucleótidos que se encuentra en esa molécula de ADN o a la cadena complementaria de la molécula de ADN. Puede referirse a la información contenida en la molécula de ADN como su representación in silico.

Locus se refiere a una región determinada de interés en el ADN de un individuo, que puede referirse a un SNP, el punto de una posible inserción o supresión, o el punto de otra variación genética relevante. Los SNP vinculados a enfermedad pueden referirse también a loci vinculados a enfermedad. Alelo polimórfico, también "locus polimórfico, se refiere a un alelo o locus en el que el genotipo varía entre individuos de una determinada especie. Algunos ejemplos de alelos polimórficos incluyen polimorfismos de un solo nucleótido, repeticiones en tándem cortas, deleciones, duplicaciones e inversiones.

*Sitio polimórfico* se refiere a nucleótidos específicos encontrados en una región polimórfica que varía entre individuos.

*Alelo* se refiere a los genes que ocupan un locus determinado.

*Datos genéticos*, también "datos genotípicos", se refiere a los datos que describen aspecto del genoma de uno o más individuos. Pueden referirse a uno o un conjunto de loci, secuencias parciales o completas, cromosomas parciales o completos o al genoma completo. Puede referirse a la identidad de uno de una pluralidad de nucleótidos; puede referirse a un conjunto de nucleótidos secuenciales o nucleótidos de diferentes ubicaciones en el genoma, o una combinación de estos.

Los datos genotípicos se utilizan típicamente in silico; sin embargo, también es posible considerar nucleótidos físicos en una secuencia como datos genéticos químicamente codificados. Se puede decir que los datos genotípicos son "sobre", "de" o "procedentes del" individuo o los individuos. Los datos genotípicos se pueden referir a medidas resultantes de una plataforma de determinación genotípica cuyas mediciones se realizan con material genético.

*Material genético,* también "muestra genética" se refiere a materia física, como tejido o sangre, de uno o más individuos que comprenden Datos genéticos ruidosos de ADN o ARN, se refiere a datos genéticos con cualquiera de los siguientes elementos: pérdidas de alelos, mediciones inciertas de pares de bases, mediciones incorrectas de pares de bases, mediciones ausentes de pares de bases, mediciones inciertas de inserciones o deleciones, mediciones inciertas de número de copias de segmentos de cromosomas, señales falsas, mediciones ausentes, otros errores, o combinaciones de estos.

*Certeza* se refiere a la probabilidad estadística que el SNP, el alelo, el conjunto de alelos determinados, o el número determinado de copias de segmento de cromosoma represente correctamente el estado genético real del individuo.

*Determinación de ploidía*, también “determinación del número de copias de cromosoma”, o “determinación del número de copias (CNC), puede ser la acción de determinar la cantidad e identidad cromosómica de uno o más cromosomas presentes en una célula.

Aneuploidía se refiere al estado en que están presentes en una célula un número erróneo de cromosomas. En el caso de una célula somática humana, puede referirse al caso de que una célula no contenga 22 pares de cromosomas autosómicos y un par de cromosomas sexuales. En un gameto humano, puede referirse al caso de que una célula no contenga uno de cada uno de los 23 cromosomas. En caso de un único tipo de cromosoma, se puede referir al caso en el que hay presentes más o menos de dos cromosomas homólogos pero no idénticos, o en el que hay dos copias de cromosomas presentes procedentes del mismo progenitor.

*Estado de ploidía* se refiere a la cantidad e identidad cromosómica de uno o más tipos de cromosomas en una célula.

*Cromosoma* se puede referir a una única copia del cromosoma, lo que significa una única molécula de ADN de las que hay 46 en una célula somática normal; un ejemplo es el "cromosoma de origen materno 18". Cromosoma también se puede referir al tipo de cromosoma, de lo que hay 23 en una célula somática humana normal; un ejemplo es "cromosoma 18".

*Identidad cromosómica* puede referirse al número de cromosomas referente, es decir el tipo de cromosoma. Los humanos normales tienen 22 tipos de cromosomas autosómicos numerados, y dos tipos de cromosomas sexuales. Puede referirse también al origen parental del cromosoma. Puede también referirse a un cromosoma específico heredado del progenitor. También puede referirse a otras características identificadoras de un cromosoma.

*El estado del material genético* o simplemente “estado genético” puede referirse a la identidad de un conjunto de SNP en el ADN, puede referirse a los haplotipos por fases del material genético, y puede referirse a la secuencia del ADN, incluyendo inserciones, supresiones, repeticiones y mutaciones. También puede referirse al estado de ploidía de uno o más cromosomas, segmentos cromosómicos o conjunto de segmentos cromosómicos.

*Datos alélicos* se refiere a un conjunto de datos genotípicos respecto a un conjunto de uno o más alelos. Puede referirse a los datos haplotípicos por fases. Puede referirse a identidades de SNP, y puede referirse a los datos de secuencia del ADN, incluyendo inserciones, supresiones, repeticiones y mutaciones. Puede incluir el origen parental de cada alelo.

*Estado alélico* se refiere al estado real de los genes en un conjunto de uno o más alelos. Puede referirse al estado real de los genes descrito por los datos alélicos.

*Ratio alélico*, o ratio de alelos, se refiere al ratio entre la cantidad de cada alelo en un locus que se encuentra presente en una muestra o en un individuo. Cuando la muestra se ha medido mediante secuenciación, el ratio alélico se puede referir al ratio de lecturas de secuencias que corresponden a cada alelo del locus. Cuando la muestra se ha medido por un método de medición basado en la intensidad, el ratio del alelo se puede referir al ratio de las cantidades de cada alelo presente en ese locus estimado por el método de medición.

*Recuento de alelos* se refiere al número de secuencias que corresponden a un determinado locus y, si ese locus es polimórfico, se refiere al número de secuencias que corresponden a cada uno de los alelos. Si cada alelo se somete a recuento de forma binaria, entonces el recuento de alelos será el número entero. Si los alelos se someten a recuento probabilístico, entonces el recuento de alelos podrá ser una fracción.

*Probabilidad del recuento de alelos* se refiere al número de secuencias que es probable que correspondan a un locus determinado o a un conjunto de alelos en un locus polimórfico, combinado con la probabilidad de la correspondencia. Cabe señalar que los recuentos de alelos son equivalentes a las probabilidades del recuento de alelos, donde la probabilidad de correspondencia para cada secuencia recontada es binaria (cero o uno). En algunas realizaciones, las probabilidades del recuento de alelos pueden ser binarias. En algunas realizaciones, las probabilidades del recuento de alelos se pueden establecer para que sean iguales a las mediciones de ADN.

*Distribución alélica*, o "distribución del recuento de alelos", se refiere a la cantidad relativa de cada alelo que se encuentra presente para cada locus de un conjunto de loci. Una distribución alélica se puede referir a un individuo, a una muestra, o a un conjunto de mediciones realizadas con la muestra. En el contexto de la secuenciación, la distribución alélica se refiere al número o número probable de lecturas que corresponden a un alelo determinado para cada alelo de un conjunto de loci polimórficos. Las mediciones de alelos se pueden tratar probabilísticamente, es decir, la probabilidad de que un determinado alelo se encuentre presente para una determinada lectura de secuencia es una fracción entre 0 y 1, o se pueden tratar de forma binaria, es decir, cada lectura dada se considera exactamente cero o una copia de un determinado alelo.

*Patrón de distribución alélica* se refiere a un conjunto de distribuciones alélicas diferentes para distintos contextos parentales. Determinados patrones de distribución alélica pueden ser indicativos de determinados estados de ploidía.

*Sesgo alélico* se refiere al grado en el que el ratio medido de alelos de un locus heterocigoto difiere del ratio que estaba presente en la muestra original de ADN. El grado de sesgo alélico en un determinado locus es igual al ratio alélico observado en ese locus, medido, dividido por el ratio de alelos de la muestra de ADN original en ese locus. El sesgo alélico se puede definir para que sea superior a uno, de forma que si el cálculo del grado de sesgo alélico da un valor, x, que es inferior a 1, entonces el grado de sesgo alélico se puede reajustar como 1/x. El sesgo alélico se puede deber a un sesgo de amplificación, un sesgo de purificación o algún otro fenómeno que afecta a distintos alelos de forma diferente.

*Cebador*, también "sonda para PCR", se refiere a una única molécula de ADN (un oligómero de ADN) o un conjunto de moléculas de ADN (oligómeros de ADN), donde las moléculas de ADN son idénticas, o prácticamente idénticas, y donde el cebador contiene una región que está diseñada para hibridarse con un locus polimórfico diana, y m contiene una secuencia de cebado diseñada para permitir la amplificación de PCR. Un cebador también puede contener un código de barras molecular. Un cebador puede contener una región aleatoria que difiere para cada molécula individual.

*Sonda de captura híbrida* se refiere a cualquier secuencia de ácido nucleico, posiblemente modificada, que es generada a través de diversos métodos, como PCR o síntesis directa, y diseñada para ser complementaria de una

cadena de una secuencia de ADN diana concreta en una muestra. Las sondas de captura híbridas exógenas se pueden añadir a una muestra preparada e hibridada a través de un proceso de desnaturalización-rehibridación para formar dúplex de fragmentos exógenos-endógenos. Posteriormente, estos dúplex se pueden separar físicamente de la muestra a través de diversos medios.

*Lectura de secuencia* se refiere a los datos que representan una secuencia de bases de nucleótidos que han sido medidas utilizando un método de secuenciación clonal. La secuenciación clonal puede producir datos de la secuencia que representan una única molécula o clones o agrupaciones de una molécula de ADN original. Una lectura de secuencia también puede tener asociada una puntuación de calidad en la posición de cada base de la secuencia que indica la probabilidad de que el nucleótido se ha determinado correctamente.

*Establecer la correspondencia de una lectura de secuencia* es el proceso de determinar la ubicación de origen de la lectura de secuencia en la secuencia del genoma de un organismo concreto. La ubicación del origen de las lecturas de secuencia se basa en la similitud de la secuencia de nucleótidos de la lectura y la secuencia del genoma.

*Error de copia emparejada*, también “aneuploidía de cromosoma emparejado”, o “MCA”, se refiere a un estado de aneuploidía donde una célula contiene dos cromosomas idénticos o casi idénticos. Este tipo de aneuploidía puede surgir durante la formación de los gametos en la mitosis, y puede ser denominada error de no disyunción mitótica. Este tipo de error se puede producir en la mitosis. La trisomía emparejada se puede referir al caso en el que tres copias de un determinado cromosoma se encuentran presentes en un individuo y dos de las copias son idénticas.

*Error de copia no emparejada*, también “Aneuploidía de Cromosoma Único” o “UCA”, se refiere a un estado de aneuploidía en el que una célula contiene dos cromosomas que proceden del mismo progenitor, y que pueden ser homólogos pero no idénticos. Este tipo de aneuploidía puede surgir durante la meiosis, y puede ser denominado error meiótico. La trisomía no emparejada se puede referir al caso en el que tres copias de un determinado cromosoma se encuentran presentes en un individuo y dos de las copias son del mismo progenitor, y son homólogas, pero no idénticas. Cabe señalar que trisomía no emparejada se puede referir al caso en el que dos cromosomas homólogos de un progenitor se encuentran presentes y donde algunos segmentos de los cromosomas son idénticos mientras que otros segmentos son simplemente homólogos.

*Cromosomas homólogos* se refiere a copias de cromosomas que contienen el mismo conjunto de genes que normalmente se emparejan durante la meiosis.

*Cromosomas idénticos* se refiere a cromosomas que contienen el mismo conjunto de genes, y para cada gen tienen el mismo conjunto de alelos que son idénticos, o casi idénticos.

*Pérdida de alelos* (Allele Drop Out, ADO) se refiere a la situación en la que no se detecta uno de los pares de bases en un conjunto de pares de bases de cromosomas homólogos en un alelo determinado.

*Pérdida de locus* (Locus Drop Out, LDO) se refiere a la situación en la que no se detectan ambos pares de bases en un conjunto de pares de bases de cromosomas homólogos en un alelo determinado.

*Homocigoto* se refiere a tener alelos similares en loci cromosómicos correspondientes.

*Heterocigoto puede referirse a tener alelos distintos en loci cromosómicos correspondie*ntes.

*La tasa de heterocigosidad* se refiere a la tasa de individuos de la población que tienen alelos heterocigotos en un determinado locus. La tasa de heterocigosidad también se puede referir al ratio previsto o medido de alelos, en un determinado locus de un individuo o una muestra de ADN.

*Polimorfismo de un solo nucleótido altamente informativo* (HISNP) se refiere a un SNP donde el feto tiene un alelo que no se encuentra presente en el genotipo de la madre.

*Región cromosómica* se refiere a un segmento de cromosoma o un cromosoma completo.

*Segmento de un cromosoma* puede referirse a una sección de un cromosoma que en cuanto a tamaño puede ir de un par de bases al cromosoma entero.

*Cromosoma* se refiere a un cromosoma completo, o también a un segmento o sección de un cromosoma.

*Copias* se refiere al número de copias de un segmento de cromosoma. Puede referirse a copias idénticas, o puede referirse a copias no idénticas homólogas de un segmento de cromosoma, donde las distintas copias del segmento de cromosoma contienen un conjunto de loci sustancialmente similar, y donde uno o más de los alelos son distintos. Hay que advertir que en algunos casos de aneuploidía, tales como el error de copia M2, es posible tener algunas copias del segmento de cromosoma determinado que sean idénticas, así como algunas copias del mismo segmento de cromosoma que no sean idénticas

*Haplotipo* es una combinación de alelos en múltiples loci que son típicamente heredados juntos en el mismo cromosoma. El haplotipo puede referirse a solamente dos loci o a un cromosoma completo, dependiendo del número de eventos de recombinación que se han producido entre un conjunto de loci determinado. El haplotipo puede referirse también a un conjunto de polimorfismos de un solo nucleótido (SNP) en una sola cromátida que están asociados estadísticamente.

*Datos haplotípicos* llamados también “datos por fase” o “datos genéticos ordenados”; puede referirse a datos de un solo cromosoma en un genoma diploide o poliploide; es decir, la copia materna o paterna aislada de un cromosoma en un genoma diploide.

*Ajuste por fases* puede referirse a la acción de determinar los datos genéticos haplotípicos de un individuo concreto no ordenados, datos genéticos diploides (o poliploides). Puede referirse a la acción de determinar cuál de dos genes en un alelo, para un conjunto de alelos hallado en un cromosoma, está asociado con cada uno de los dos cromosomas homólogos en un individuo.

*Datos por fases* puede referirse a los datos genéticos donde se han determinado uno o más haplotipos.

*Hipótesis* se refiere a un posible estado de ploidía en un conjunto de cromosomas determinado, o un conjunto de estados alélicos posibles en un conjunto de loci determinado. El conjunto de posibilidades puede contener uno o más elementos.

*Hipótesis de número de copias*, también “hipótesis de estado de ploidía”, se refiere a una hipótesis sobre cuántas copias de un cromosoma determinado hay en un individuo. Puede referirse también a una hipótesis sobre la identidad de cada uno de los cromosomas, incluyendo el progenitor de origen de cada cromosoma, y cuál de los dos cromosomas del progenitor está presente en el individuo. También puede referirse a una hipótesis sobre qué cromosomas, o segmentos de cromosomas, de haberlos, de un individuo relacionado, se corresponden genéticamente con un cromosoma determinado de un individuo

*Individuo diana* se refiere al individuo cuyo estado genético está siendo determinado. En algunas realizaciones, está disponible solamente una cantidad limitada de ADN del individuo diana. En algunas realizaciones, el individuo diana es un feto. En algunas realizaciones, puede haber más de un individuo diana. En algunas realizaciones, cada feto derivado de un par de progenitores puede ser considerado individuo diana. En algunas realizaciones, los datos genéticos que se están determinando son uno o un conjunto de determinaciones de alelos. En algunas realizaciones, los datos genéticos que se están determinando son la determinación de un estado de ploidía.

*Individuo relacionado* se refiere a cualquier individuo que esté relacionado genéticamente y comparta por tanto bloques de haplotipos con el individuo diana. En un contexto, el individuo diana puede ser un progenitor genético del individuo diana, o cualquier material genético derivado de un progenitor, como esperma, un cuerpo polar, un embrión, un feto o un niño. Puede referirse también a un hermano, progenitor o abuelo.

*Hermano* se refiere a cualquier individuo cuyos padres genéticos sean los mismos que los del individuo en cuestión. En algunas realizaciones, puede referirse a un niño ya nacido, a un embrión, o a un feto, o una o más células procedentes de un niño ya nacido, de un embrión o de un feto. Un hermano puede referirse también a un individuo haploide procedente de uno de los progenitores, como esperma, un cuerpo polar, o cualquier otro conjunto de materia genética haplotípica. Un individuo puede ser considerado hermano de sí mismo.

*Fetal* se refiere al feto o a la región de la placenta que es genéticamente similar al feto. En una mujer embarazada, alguna porción de la placenta es genéticamente similar al feto, y el ADN fetal flotante libre que se encuentra en la sangre materna puede haberse originado en la porción de la placenta con un genotipo que coincide con el del feto. Cabe señalar que la información genética de la mitad de los cromosomas de un feto se hereda de la madre del feto. En algunas realizaciones, el ADN de estos cromosomas heredados de la madre procedente de una célula fetal se considera "de origen fetal" no "de origen materno".

*ADN de origen fetal* se refiere al ADN que era originalmente parte de la célula cuyo genotipo era esencialmente equivalente al del feto.

*ADN de origen materno* se refiere al ADN que era originalmente parte de la célula cuyo genotipo era esencialmente equivalente al de la madre.

*Niño* se refiere a un embrión, un blastómero o un feto. Cabe señalar que, en las realizaciones divulgadas en el presente documento, los conceptos descritos se aplican igualmente bien a los individuos que son un niño ya nacido, un embrión o un conjunto de células de estos. El uso del término niño puede simplemente indicar que el individuo designado como el niño es la descendencia genética de los padres.

*Progenitor* se refiere a la madre o el padre genéticos de un individuo. Un individuo tiene típicamente dos progenitores, una madre y un padre, aunque esto no tiene que ser necesariamente así, por ejemplo, en el quimerismo cromosómico o genético. Un progenitor puede ser considerado un individuo.

*Contexto parental* se refiere al estado genético de un SNP determinado, en cada uno de los dos cromosomas relevantes para cada uno de los dos progenitores de la diana.

*Desarrollo según lo deseado*, también “desarrollo normal”, se refiere a un embrión viable implantado en un útero y que resulte en un embarazo y/o al embarazo que continúa y que resulta en un nacimiento vivo y/o a que el niño nacido carece de anomalías cromosómicas y/o a que el niño nacido carece de otros estados genéticos no deseados, tales como genes vinculados a enfermedad. El término “desarrollo según deseado” comprende todo aquello que puedan desear los padres o el personal sanitario. En algunos casos “desarrollo según deseado” puede referirse a un embrión viable o no viable que resulte útil para la investigación médica u otros fines.

*Inserción en un útero* se refiere al proceso de transferencia de un embrión en la cavidad uterina, en el contexto de la fertilización in vitro.

*Plasma materno* se refiere a la porción de plasma de la sangre de una mujer que está embarazada.

*Decisión clínica* se refiere a toda decisión de emprender o no una acción, que tenga un resultado que afecte a la salud o la supervivencia de un individuo. En el contexto del diagnóstico prenatal, una decisión clínica puede referirse a la decisión de abortar o no un feto. Una decisión clínica puede referirse también a la decisión de realizar más pruebas, de emprender acciones para evitar un fenotipo no deseado o de emprender acciones para prepararse para el nacimiento de un niño con anomalías.

*Caja de diagnóstico* se refiere a una o una combinación de máquinas diseñadas para realizar uno o una pluralidad de aspectos de los métodos divulgados en el presente documento. En una realización, la caja de diagnóstico se puede colocar en un punto del cuidado del paciente. En una realización, la caja de diagnóstico puede realizar una amplificación focalizada seguida por una secuenciación. En una realización, la caja de diagnóstico puede funcionar sola o con la ayuda de un técnico.

*Método basado en informática* se refiere a un método que confía sustancialmente en la estadística para dar sentido a una gran cantidad de datos. En el contexto del diagnóstico prenatal, se refiere a un método diseñado para determinar el estado de ploidía de uno o más cromosomas, o el estado alélico de uno o más alelos, deduciendo estadísticamente el estado más probable, en lugar de medir directamente de forma física el estado, partiendo de una gran cantidad de datos genéticos, por ejemplo, una secuenciación o array molecular. En una realización de la presente divulgación, la técnica basada en informática puede ser una divulgada en esta patente. En una realización de la presente divulgación puede ser PARENTAL SUPPORT™.

*Datos genéticos primarios* se refiere a las señales de intensidad análogos que son producidos por una plataforma de determinación del genotipo. En el contexto de los arrays de SNP, los datos genéticos primarios se refieren a las señales de intensidad antes de que se haya realizado ninguna determinación del genotipo. En el contexto de la secuenciación, los datos genéticos primarios se refieren a las mediciones análogas, análogas al cromatograma, que proceden del secuenciador antes de que se haya determinado la identidad de cualesquiera pares de bases y antes de que la secuencia se haya correlacionado con el genoma.

*Datos genéticos secundarios* se refiere a los datos genéticos procesados que son producidos por una plataforma de determinación del genotipo. En el contexto de un array del SNP, los datos genéticos secundarios se refieren a las determinaciones del alelo realizadas por el software asociado con el lector de arrays del SNP, donde el software ha realizado una determinación con independencia de que un determinado alelo se encuentre presente o no en la muestra. En el contexto de la secuenciación, los datos genéticos secundarios se refieren a las identidades del par de bases de las secuencias que se han determinado y posiblemente también donde las secuencias se han correlacionado con el genoma.

*Diagnóstico prenatal no invasivo (NPD)*, o también "Análisis prenatal no invasivo" (NPS), se refiere a un método para determinar el estado genético de un feto que se está gestando en una madre, utilizando material genético que se encuentra en la sangre de la madre, donde el material genético se obtiene extrayendo sangre intravenosa de la madre.

*Enriquecimiento preferente de ADN* que corresponde a un locus, o enriquecimiento preferente de ADN en un locus, se refiere a cualquier método que resulta en un aumento del porcentaje de moléculas de ADN en la mezcla de ADN que corresponde al locus tras el enriquecimiento respecto del porcentaje de moléculas de ADN existente en la mezcla de ADN que corresponde al locus antes del enriquecimiento. El método puede implicar la amplificación selectiva de moléculas de ADN que corresponden a un locus. El método puede implicar la eliminación de moléculas de ADN que no corresponden al locus. El método puede implicar una combinación de métodos. El grado de enriquecimiento se define como el porcentaje de moléculas de ADN tras el enriquecimiento de la mezcla que corresponde al locus dividido por el porcentaje de moléculas de ADN antes del enriquecimiento de la mezcla que corresponde al locus. El enriquecimiento preferente se puede realizar en diversos loci. En algunas realizaciones de la presente divulgación, el grado de enriquecimiento es mayor que 20. En algunas realizaciones de la presente divulgación, el grado de enriquecimiento es mayor que 200. En algunas realizaciones de la presente divulgación, el grado de enriquecimiento es mayor que 2000. Cuando se realiza un enriquecimiento preferente en diversos loci, el grado de enriquecimiento se puede referir al grado medio de enriquecimiento de todos los loci que componen el conjunto de loci.

*Amplificación* se refiere a un método que aumenta el número de copias de una molécula de ADN.

*Amplificación selectiva* se puede referir a un método que aumenta el número de copias de una molécula concreta de ADN o moléculas de ADN que corresponden a una región concreta del ADN. También se puede referir a un método que aumenta el número de copias de una molécula diana concreta de ADN o de una región diana de ADN más de lo que aumenta en las moléculas o regiones de ADN no diana. La amplificación selectiva puede ser un método de enriquecimiento preferente.

*Secuencia de cebado universal* se refiere a una secuencia de ADN que se puede unir a una población de moléculas de ADN diana, por ejemplo, mediante enlace, PCR o PCR mediada por enlace. Una vez añadida a la población de

moléculas diana, se pueden utilizar cebadores específicos para las secuencias de cebado universal con el fin de amplificar la población diana utilizando un único par de cebadores de amplificación. Las secuencias de cebado universal típicamente no están relacionadas con las secuencias diana.

*Adaptadores universales*, o "adaptadores de unión" o "etiquetas de biblioteca" son moléculas de ADN que contienen una secuencia de cebado universal que se puede unir mediante enlace covalente al extremo 5' y 3' de una población de moléculas de ADN de doble cadena diana. La adición de los adaptadores proporciona secuencias de cebado universal a los extremos 5' y 3' de la población diana con la que se puede producir la amplificación por PCR, amplificando todas las moléculas de la población diana con un único par de cebadores de amplificación.

*Focalización* se refiere a un método utilizado para amplificar de forma selectiva o enriquecer de forma preferente las moléculas de ADN que corresponden a un conjunto de loci en una mezcla de ADN.

*Modelo de distribución conjunto* se refiere a un modelo que define la probabilidad de eventos definidos en términos de múltiples variables aleatorias, dada una pluralidad de variables aleatorias definidas en el mismo espacio de probabilidad, donde las probabilidades de la variable están vinculadas. En algunas realizaciones, se puede utilizar el caso degenerado cuando las probabilidades de las variables no están vinculadas.

*Hipótesis*

En el contexto de esta divulgación, una hipótesis se refiere a un posible estado genético. Puede referirse a un posible estado de ploidía. Puede referirse a un posible estado alélico. Un conjunto de hipótesis se puede referir a un conjunto de posibles estados genéticos, un conjunto de posibles estados de ploidía o combinaciones de estos. En algunas realizaciones, un conjunto de hipótesis puede ser diseñado de forma que una de las hipótesis del conjunto corresponda al estado genético real de un individuo determinado. En algunas realizaciones, un conjunto de hipótesis puede estar diseñado de forma que todo posible estado genético pueda ser descrito por lo menos por una hipótesis del conjunto. En algunas realizaciones de la presente divulgación, un aspecto del método consiste en determinar qué hipótesis corresponde al estado genético real del individuo en cuestión.

En otra realización de la presente divulgación, un paso incluye la creación de una hipótesis. En algunas realizaciones puede ser una hipótesis del número de copias. En algunas realizaciones puede incluir una hipótesis sobre qué segmentos de un cromosoma de cada uno de los individuos relacionados corresponde genéticamente a qué segmentos, de haberlos, de los otros individuos relacionados. Crear una hipótesis puede referirse al hecho de establecer los límites de las variables, de forma que la totalidad del conjunto de posibles estados genéticos que están siendo considerados estén comprendidos en esas variables.

Una “hipótesis de número de copias”, denominada también una “hipótesis de ploidía”, o una “hipótesis de estado de ploidía”, puede referirse a una hipótesis relacionada con un posible estado de ploidía para un cromosoma determinado, un tipo de cromosoma, o sección de un cromosoma, en el individuo diana. Puede referirse también al estado de ploidía en más de uno de los tipos de cromosomas del individuo. Un conjunto de hipótesis de número de copias puede referirse a un conjunto de hipótesis donde cada hipótesis corresponde a un posible estado de ploidía distinto en un individuo. Un conjunto de hipótesis se puede referir a un conjunto de posibles estados de ploidía, un conjunto de posibles contribuciones de haplotipos parentales, un conjunto de posibles porcentajes de ADN fetal en la muestra combinada o combinaciones de estos.

Un individuo normal contiene uno de cada tipo de cromosoma de cada progenitor. No obstante, debido a errores en meiosis y mitosis, es posible que un individuo tenga 0, 1,2, o más de un tipo de cromosoma determinado de cada progenitor. En la práctica, es poco frecuente ver más de dos de un cromosoma determinado de un progenitor. En esta divulgación, algunas realizaciones solo consideran las hipótesis posibles en las que 0, 1, o 2 copias de un cromosoma determinado proceden de un progenitor; resulta una extensión trivial considerar más o menos copias posibles procedentes de un progenitor. En algunas realizaciones, para un cromosoma determinado hay nueve posibles hipótesis: las tres hipótesis posibles referentes a 0, 1, o 2 cromosomas de origen materno, multiplicado por las tres hipótesis posibles sobre 0, 1, o 2 cromosomas de origen paterno. Consideremos que (m, f) se refiere a la hipótesis en la que m es el número de un cromosoma determinado heredado de la madre, y f es el número de un cromosoma determinado heredado del padre. En consecuencia, las nueve hipótesis son (0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), y (2,2). Esto también se puede escribir como H00, H01, H02, H10, H12, H20, H21, y H22. Las distintas hipótesis corresponden a diferentes estados de ploidía. Por ejemplo, (1,1) se refiere a un cromosoma disómico normal; (2,1) se refiere a una trisomía materna, y (0,1) se refiere a una monosomía paterna. En algunas realizaciones, el caso en el que dos cromosomas son heredados de un progenitor y un cromosoma del otro puede diferenciarse además en dos casos: uno en el que los dos cromosomas son idénticos (error de copias emparejadas), y uno en el que los dos cromosomas son homólogos pero no idénticos (error de copias no emparejadas). En estas realizaciones, hay dieciséis hipótesis posibles. Cabe señalar que se pueden utilizar otros conjuntos de hipótesis y un número diferente de hipótesis.

En algunas realizaciones de la presente divulgación, la hipótesis de la ploidía puede referirse a una hipótesis sobre qué cromosoma de otros individuos relacionados corresponde a un cromosoma hallado en el genoma del individuo diana. En algunas realizaciones, una clave del método es el hecho de que cabe esperar que individuos relacionados compartan bloques haplotípicos, y utilizando datos genéticos medidos de individuos relacionados, junto con el conocimiento de qué bloques haplotípicos coinciden entre el individuo diana y el individuo relacionado, es posible inferir los datos genéticos correctos de un individuo diana con mayor certeza que utilizando solamente las

mediciones genéticas del individuo diana. Como tal, en algunas realizaciones la hipótesis de ploidía puede referirse no solamente al número de cromosomas, sino también a qué cromosomas en individuos relacionados son idénticos, o casi idénticos, a uno o más cromosomas del individuo diana.

Una vez se ha definido el conjunto de hipótesis, cuando los algoritmos operan sobre los datos genéticos de entrada, pueden dar como resultado una probabilidad estadística determinada para cada una de las hipótesis consideradas. Las probabilidades de las diversas hipótesis pueden determinarse calculando matemáticamente, para cada una de las distintas hipótesis, el valor de probabilidad, como lo indican una o más de las técnicas de experto, los algoritmos, y/o los métodos descritos en otra parte de esta divulgación, utilizando como entrada los datos genéticos pertinentes.

Una vez calculadas las probabilidades de las distintas hipótesis, como se haya determinado por diversas técnicas, se pueden combinar. Esto puede implicar multiplicar para cada hipótesis las probabilidades determinadas mediante cada técnica. El producto de las probabilidades de las hipótesis puede ser normalizado. Hay que advertir que una hipótesis de ploidía se refiere a un posible estado de ploidía de un cromosoma.

El proceso de “combinación de probabilidades”, denominado también “hipótesis combinadas”, o combinar los resultados de técnicas de experto, es un concepto que debe resultar familiar a los expertos en la técnica del álgebra lineal. Una posible forma de combinar probabilidades es como sigue: cuando se utiliza una técnica de experto para evaluar un conjunto de hipótesis en un conjunto determinado de datos genéticos, el resultado del método es un conjunto de probabilidades asociadas, de forma uno-a-uno, a cada hipótesis del conjunto de hipótesis. Cuando un conjunto de probabilidades que han sido determinadas por una primera técnica de experto, cada una de las cuales está asociada a una de las hipótesis del conjunto, se combina con un conjunto de probabilidades determinadas por una segunda técnica de experto, cada una de las cuales va asociada con el mismo conjunto de hipótesis, los dos conjuntos de probabilidades se multiplican. Esto significa que, para cada hipótesis del conjunto, las dos probabilidades asociadas a esa hipótesis, determinada por los dos métodos de experto, se multiplican juntas, y el producto correspondiente es el resultado de probabilidades. Este proceso puede ampliarse a cualquier número de técnicas de experto. Si se utiliza solamente una técnica de experto, las probabilidades de salida son las mismas que las de entrada. Si se utilizan más de dos técnicas de experto, las probabilidades pertinentes pueden multiplicarse al mismo tiempo. Los productos pueden normalizarse de forma que las probabilidades de las hipótesis en el conjunto de hipótesis sumen 100%.

En algunas realizaciones de la divulgación, si las probabilidades combinadas de una hipótesis determinada son mayores que las probabilidades combinadas de cualquiera de las otras hipótesis, puede considerarse que esa hipótesis se determina como la más probable. En algunas realizaciones, se puede determinar una hipótesis como la más probable, y el estado de ploidía, u otro estado genético puede ser determinado si la probabilidad normalizada es superior a un umbral. En una realización, esto puede significar que el número y la identidad de los cromosomas asociados a esa hipótesis pueden ser determinados como el estado de ploidía. En una realización, esto puede significar que la identidad de los alelos asociados a esa hipótesis puede ser determinada como el estado alélico. En algunas realizaciones el umbral puede situarse entre el 50% y aproximadamente el 80%. En algunas realizaciones el umbral puede situarse entre el 80% y aproximadamente el 90%. En algunas realizaciones el umbral puede situarse entre el 90% y aproximadamente el 95%. En algunas realizaciones el umbral puede situarse entre el 95% y aproximadamente el 99%. En algunas realizaciones el umbral puede situarse entre el 99% y aproximadamente el 99,9%. En algunas realizaciones el umbral puede situarse por encima de aproximadamente el 99,9%.

*Contextos parentales*

El contexto parental se refiere al estado genético de un alelo determinado, en cada uno de los dos cromosomas relevantes para cada uno de los dos progenitores de la diana. Hay que advertir que en una realización, el contexto parental no se refiere al estado alélico de la diana, sino al estado alélico de los padres. El contexto parental de un SNP determinado puede consistir en cuatro pares de bases, dos paternos y dos maternos; pueden ser iguales o distintos entre sí. Esto viene expresado típicamente como “mim2|fif2”, donde m1 y m2 son el estado genético del SNP concreto en los dos cromosomas maternos, y f1 y f2 son el estado genético de dicho SNP en los dos cromosomas paternos. En algunas realizaciones, el contexto parental puede venir expresado como “f1f2|m1m2’. Hay que señalar que los subíndices “1” y “2” se refieren al genotipo, en ese alelo determinado, del primer y el segundo cromosoma; ver también que la elección de qué cromosoma se etiqueta como “1” y cuál como “2” es arbitraria.

Hay que señalar que en esta divulgación, A y B se utilizan frecuentemente para representar de forma genérica identidades de pares de bases; A o B podrían representar igualmente bien a C (citosina), G (guanina), A (adenina) o T (timina). Por ejemplo, si en un alelo basado en SNP, el genotipo materno fue T en ese SNP de un cromosoma, y G en ese SNP del cromosoma homólogo, y el genotipo paterno en ese alelo es G en ese SNP en ambos cromosomas homólogos, se podría decir que el alelo del individuo diana tiene el contexto parental de AB|BB; también se podría decir que el alelo tiene el contexto parental de AB|AA. Ver que, en teoría, cualquiera de los cuatro nucleótidos posibles podría darse en un alelo determinado, y así es posible, por ejemplo, que la madre tenga un genotipo de AT, y el padre tenga un genotipo de GC en un alelo determinado. No obstante, datos empíricos indican que en la mayoría de los casos solo dos de los cuatro posibles pares de bases se observan en un alelo determinado. Resulta posible, por ejemplo, cuando se utilizan repeticiones en tándem únicas, tener más de dos, más de cuatro e incluso más de diez contextos parentales. En esta divulgación, en la discusión se supone que se observarán solamente dos

posibles pares de bases en un alelo determinado, aunque las realizaciones divulgadas en el presente documento se podrían modificar para tener en cuenta los casos en los que este supuesto no se sostiene.

Un “contexto parental” puede referirse a un conjunto o subconjunto de SNP diana que tienen el mismo contexto parental. Por ejemplo, si hubiera que medir 1000 alelos en un cromosoma determinado en un individuo diana, el contexto AA|BB podría referirse al conjunto de todos los alelos en el grupo de 1000 alelos donde el genotipo de la madre de la diana era homocigoto, y el genotipo del padre de la diana es homocigoto, pero donde el genotipo materno y el genotipo paterno son distintos en ese locus. Si los datos parentales no están por fases, y por tanto AB = BA, hay nueve contextos parentales posibles: AA|AA, AA|AB, AA|BB, AB|AA, AB|AB, AB|BB, BB|AA, BB|AB, y BB|BB. Si los datos parentales están ajustados por fases, y por tanto AB ≠ BA, hay dieciséis contextos parentales distintos posibles: AA|AA, AA|AB, AA|BA, AA|BB, AB|AA, AB|AB, AB|BA, AB|BB, BA|AA, BA|AB, BA|BA, BA|BB, BB|AA, BB|AB, BB|BA, y BB[BB. Cada alelo SNP de un cromosoma, excluyendo algunos SNP en los cromosomas sexuales, tiene uno de esos contextos parentales. El conjunto de SNP donde el contexto parental en un progenitor es heterocigoto puede ser denominado el contexto heterocigoto.

*Uso de los contextos parentales en NPD*

El diagnóstico prenatal no invasivo es una técnica importante que se puede utilizar para determinar el estado genético de un feto a partir del material genético que se obtiene de manera no invasiva, por ejemplo, de la sangre extraída a la madre embarazada. La sangre se podría separar y el plasma aislarse, para después aislar el ADN del plasma. La selección del tamaño se podría utilizar para aislar el ADN de la longitud apropiada. El ADN se puede enriquecer preferentemente en un conjunto de loci. Este ADN se puede entonces medir a través de varios medios, como mediante hibridación con un array para la determinación del genotipo y midiendo la fluorescencia o secuenciación en un secuenciador de alto rendimiento.

Cuando se utiliza la secuenciación para determinar el estado de ploidía de un feto en el contexto del diagnóstico prenatal no invasivo, existen varias formas de utilizar los datos de la secuencia. La forma más habitual en que se podrían utilizar los datos de la secuencia consiste simplemente en contar en número de lecturas que corresponden a un determinado cromosoma. Por ejemplo, imagine que está intentando determinar el estado de ploidía del cromosoma 21 del feto. Imagine también que el ADN de la muestra se compone de un 10% de ADN de origen fetal y un 90% de ADN de origen materno. En este caso, se podría analizar el número medio de lecturas en un cromosoma que cabe esperar que sea disómico, por ejemplo, el cromosoma 3, y compararlo con el número de lecturas del cromosoma 21, donde las lecturas se ajustan para el número de pares de bases de ese cromosoma que forman parte de una secuencia única. Si el feto fuese euploide, cabría esperar que la cantidad de ADN por unidad de genoma fuese aproximadamente la misma en todas las ubicaciones (con sujeción a variaciones estocásticas). Por otra parte, si el feto fuese trisómico en el cromosoma 21, entonces cabría esperar que hubiese algo más de ADN por unidad genética del cromosoma 21 que en las demás ubicaciones del genoma. Concretamente cabría esperar que hubiese aproximadamente un 5% más de ADN del cromosoma 21 en la mezcla. Cuando se utiliza la secuenciación para medir el ADN, cabría esperar aproximadamente un 5% más de lecturas únicamente susceptibles de correlación con el cromosoma 21 por segmento único que con los demás cromosomas. Se podría utilizar la observación de una cantidad de ADN de un cromosoma concreto que supere un determinado umbral, cuando está ajustada para el número de secuencias que son únicamente susceptibles de correlación con ese cromosoma, como base para el diagnóstico de la aneuploidía. Otro método que se podría utilizar para detectar la aneuploidía es similar al anterior, salvo por el hecho de que se podrían tener en cuenta los contextos parentales.

A la hora de plantearse los alelos a focalizar, se podría considerar la probabilidad de que algunos contextos parentales ofrezcan más información que otros. Por ejemplo, AA|BB y el contexto simétrico BB|AA son los contextos más informativos, porque se sabe que el feto porta un alelo que es diferente de la madre. Por razones de simetría, tanto los contextos AA|BB como BB|AA se pueden denominar AA|BB. Otro conjunto de contextos parentales informativos son AA|AB y BB|AB, porque en estos casos el feto tiene un 50% de posibilidades de portar un alelo que la madre no tiene. Por razones de simetría, tanto los contextos AA|AB como BB|AB se pueden denominar AA|AB. Un tercer conjunto de contextos parentales informativos son AB|AA y AB|BB, porque en estos casos el feto porta un alelo paterno conocido, y el alelo también se encuentra presente en el genoma materno. Por razones de simetría, tanto los contextos AB|AA como AB|BB se pueden denominar AB|AA. Un cuarto contexto parental es AB|AB cuando el feto tiene un estado alélico desconocido y, con independencia del estado alélico, es uno en el que la madre tiene los mismos alelos. El quinto contexto parental es AA|AA donde la madre y el padre son heterocigotos.

*Diferentes implementaciones de las realizaciones divulgadas en el presente documento*

En el presente documento se divulgan métodos para determinar el estado de ploidía de un individuo diana. El individuo diana puede ser un blastómero, un embrión o un feto. En algunas realizaciones de la presente divulgación, un método para determinar el estado de ploidía de uno o más cromosomas en un individuo diana puede incluir cualquiera de los pasos descritos en este documento y combinaciones de estos.

En algunas realizaciones de la divulgación, la fuente del material genético que se va a utilizar para determinar el estado de ploidía del feto puede ser células fetales, como glóbulos rojos fetales nucleados, aislados de la sangre materna. El método puede implicar la obtención de una muestra de sangre de la madre embarazada. El método puede implicar el aislamiento de un glóbulo rojo fetal utilizando técnicas visuales, basadas en la idea de que una determinada combinación de colores está exclusivamente asociada con un glóbulo rojo nucleado, y una combinación

de colores similar no está asociada a ninguna otra célula presente en la sangre materna. La combinación de colores asociada con los glóbulos rojos nucleados puede incluir el color rojo de la hemoglobina alrededor del núcleo, color que se puede diferenciar mediante tinción, y el color del material nuclear que puede ser teñido, por ejemplo, de azul. Al aislar las células sanguíneas materna y extenderlas sobre un portaobjetos para después identificar aquellos puntos en los que se aprecia tanto el rojo (de la hemoglobina) como de la azul (del material nuclear), resulta posible identificar la ubicación de los glóbulos rojos nucleados. A continuación, se pueden extraer estos glóbulos rojos nucleados utilizando un micromanipulador, utilizando técnicas de determinación del genotipo y/o secuenciación para medir aspectos del genotipo del material genético que contienen esas células.

En una realización de la divulgación, se pueden teñir los glóbulos rojos nucleados con un colorante que solo emite fluorescencia en presencia de hemoglobina fetal y no de hemoglobina materna, para de este modo eliminar la ambigüedad entre los glóbulos rojos nucleados obtenidos de la madre y del feto. Algunas realizaciones de la presente divulgación pueden implicar la tinción u otro tipo de marcado del material nuclear. Algunas realizaciones de la presente divulgación pueden implicar específicamente el marcado del material nuclear fetal utilizando anticuerpos específicos para células fetales.

Hay muchas otras maneras de aislar células fetales de la sangre materna o el ADN fetal de la sangre materna, o para enriquecer muestras de material genético fetal en presencia de material genético materno. Algunos de los métodos se recogen aquí, aunque no se deberá considerar una lista exhaustiva. Algunas técnicas apropiadas se enumeran en el presente documento por conveniencia: uso de anticuerpos etiquetados con fluorescencia u otro marcado, cromatografía de exclusión por tamaño, etiquetas de afinidad etiquetadas magnéticamente o de otro modo, diferencias epigenéticas, como metilación diferencial entre las células maternas y fetales en alelos específicos, centrifugación en gradiente de densidad seguida por reducción de CD45/14 y selección CD71-positiva de las células CD45/14 negativas, gradientes de Percoll simples o dobles con diferentes osmolaridades, o método de lectina específico de galactosa.

En una realización de la presente divulgación, el individuo diana es un feto y se realizan las diferentes mediciones del genotipo con una pluralidad de muestras de ADN fetal. En algunas realizaciones de la presente divulgación, las muestras de ADN fetal proceden de células fetales aisladas donde las células fetales pueden estar mezcladas con células maternas. En algunas realizaciones de la presente divulgación, las muestras de ADN fetal proceden de ADN fetal flotante libre, donde el ADN fetal puede estar mezclado con ADN materno flotante libre.

En algunas realizaciones, las muestras de ADN fetal se pueden obtener de plasma materno o sangre materna que contiene una mezcla de ADN materno y ADN fetal. En algunas realizaciones, el ADN fetal puede estar mezclado con ADN materno en ratios materno:fetal que oscilan entre 99,9:0,1% a 99:1%; 99:1 % a 90:10%; 90:10% a 80:20%; 80:20% a 70:30%; 70:30% a 50:50%; 50:50% a 10:90%; o 10:90% a 1:99%; 1:99% a 0,1:99,9%.

En algunas realizaciones de la divulgación, la muestra genética puede ser preparada y/o purificada.

Hay una serie de procedimientos estándar conocidos en la técnica para llegar a tal fin. En algunas realizaciones, la muestra puede ser centrifugada para separar diversas capas. En algunas realizaciones, el ADN puede ser aislado utilizando filtración. En algunas realizaciones, la preparación de ADN puede implicar amplificación, separación, purificación mediante cromatografía, separación líquido-líquido, aislamiento, enriquecimiento preferente, amplificación preferente, amplificación focalizada o cualquiera de una serie de otras técnicas conocidas en la técnica o descritas en el presente documento.

En algunas realizaciones de la divulgación, un método de la presente divulgación puede implicar la amplificación de ADN. La amplificación del ADN, un proceso que transforma una pequeña cantidad de material genético en una cantidad mayor de material genético que comprende un conjunto similar da datos genéticos, se puede realizar mediante diversos métodos, incluyendo, entre otros, la reacción en cadena de la polimerasa (PCR). Un método de amplificación de ADN es la amplificación del genoma completo (WGA). Existen una serie de métodos disponibles para la WGA: PCR mediada por unión (LM-PCR), PCR de cebadores oligonucleótidos degenerados (DOP-PCR) y amplificación por desplazamiento múltiple (MDA). En la LM-PCR, unas secuencias de ADN cortas llamadas adaptadores se unen a los extremos romos del ADN. Estos adaptadores contienen secuencias de amplificación universales, que se utilizan para amplificar el ADN mediante PCR. En la DOP-PCR, se utilizan cebadores aleatorios que también contienen secuencias de amplificación universales en una primera ronda de hibridación y PCR. A continuación, se utiliza una segunda ronda de PCR para continuar amplificando las secuencias con secuencias de cebadores universales. La MDA utiliza la polimerasa phi-29, que es una enzima no específica y altamente procesiva que replica el ADN y se ha utilizado para el análisis de células únicas. Las principales limitaciones para la amplificación del material a partir de una sola célula son: (1) necesidad de utilizar concentraciones de ADN extremadamente diluidas o un volumen extremadamente reducido de una mezcla de reacción; y (2) dificultad de disociar de forma fiable el ADN de las proteínas del conjunto del genoma. Sin embargo, la amplificación del genoma completo de una sola célula se ha utilizado con éxito para diversas aplicaciones durante varios años. Existen métodos para amplificar el ADN de una muestra de ADN. La amplificación del ADN transforma la muestra inicial de ADN en una muestra de ADN que es similar en el conjunto de secuencias, pero en una cantidad mucho mayor. En algunos casos, la amplificación puede no resultar necesaria.

En algunas realizaciones, el ADN puede ser amplificado utilizando una amplificación universal, como WGA o MDA. En algunas realizaciones de la invención, el ADN se puede amplificar mediante amplificación focalizada, por

ejemplo, utilizando una PCR focalizada o sondas de circularización. En algunas realizaciones, el ADN se puede enriquecer preferentemente utilizando un método de amplificación focalizado, o un método que resulte en la separación completa o parcial del ADN deseado del no deseado, como la captura mediante métodos de hibridación. En algunas realizaciones, el ADN se puede amplificar utilizando una combinación de un método de amplificación universal y un método de enriquecimiento preferente. Una descripción más completa de algunos de estos métodos se puede encontrar en otro apartado de este documento.

Los datos genéticos del individuo diana y/o del individuo relacionado se pueden transformar de un estado molecular a un estado electrónico, midiendo el material genético adecuado utilizando herramientas o técnicas seleccionadas de un grupo, incluyendo, entre otras, microarrays para la determinación del genotipo y secuenciación de alto rendimiento. Algunos métodos de secuenciación de alto rendimiento incluyen la secuenciación de ADN de Sanger, la pirosecuenciación, la plataforma ILLUMINA SOLEXA, GENOME ANALYZER de ILLUMINA, o la plataforma de secuenciación 454 de APPLIED BIOSYSTEM, la plataforma TRUE SINGLE MOLECULE SEQUENCING de HELICOS, el método de secuenciación por microscopio de electrones de HALCYON MOLECULAR, o cualquier otro método de secuenciación. Todos estos métodos transforman físicamente los datos genéticos almacenados en una muestra de ADN en un conjunto de datos genéticos que se almacenan típicamente en un dispositivo de memoria a procesar.

Los datos genéticos de un individuo relevante se pueden medir analizando las sustancias tomadas de un grupo, incluyendo, entre otras: el tejido diploide bruto del individuo, una o más células diploides del individuo, una o más células haploides del individuo, uno o más blastómeros del individuo diana, material genético extracelular encontrado en el individuo, material genético extracelular del individuo encontrado en la sangre materna, células del individuo encontradas en la sangre materna, uno o más embriones creados de un gameto o gametos del individuo relacionado, uno o más blastómeros tomados de este embrión, material genético extracelular encontrado sobre el individuo relacionado, material genético que se sabe que procede del individuo relacionado y combinaciones de estos.

En algunas realizaciones de la divulgación, se puede crear un conjunto de al menos una hipótesis del estado de ploidía para cada uno de los tipos de cromosomas de interés del individuo diana. Cada una de las hipótesis del estado de ploidía se puede referir a un posible estado de ploidía del cromosoma o segmento del cromosoma del individuo diana. El conjunto de hipótesis puede incluir algunos o todos los posibles estados de ploidía que cabe esperar que presente el cromosoma del individuo diana. Algunos de los posibles estados de ploidía pueden incluir nulisomía, monosomía, disomía, disomía uniparental, euploidía, trisomía, trisomía emparejada, trisomía no emparejada, trisomía materna, trisomía paterna, tetrasomía, tetrasomía equilibrada (2:2), tetrasomía no equilibrada (3:1), pentasomía, hexasomía, otra aneuploidía y combinaciones de estas. Cualquiera de estos estados de aneuploidía puede ser una aneuploidía mixta o parcial, como translocaciones no equilibradas, translocaciones equilibradas, translocaciones Robertsonianas, recombinaciones, deleciones, inserciones, cruces y combinaciones de estos.

En algunas realizaciones de la divulgación, el conocimiento del estado de ploidía determinado se puede utilizar para tomar una decisión clínica. Este conocimiento, típicamente guardado como una disposición física de la materia en un dispositivo de memoria, se puede transformar después en un informe. Posteriormente se puede actuar en función de este informe. Por ejemplo, la decisión clínica puede ser poner fin al embarazo; alternativamente, la decisión clínica puede ser continuar el embarazo. En algunas realizaciones, la decisión clínica puede implicar una intervención diseñada para reducir la gravedad de la presentación fenotípica de un trastorno genético, o la decisión de emprender las acciones pertinentes para prepararse para un niño con necesidades especiales.

En una realización de la presente divulgación, cualquiera de los métodos descritos en el presente documento se puede modificar para permitir que múltiples dianas procedan del mismo individuo diana, por ejemplo, múltiples extracciones de sangre de la misma mujer embarazada. Esto puede mejorar la precisión del modelo, dado que múltiples mediciones genéticas pueden proporcionar más datos con los que se puede determinar el genotipo diana. En una realización, un conjunto de datos genéticos diana se proporcionan como datos primarios de un informe y los demás se proporcionan como datos para una doble comprobación de los datos genéticos primarios. En una realización, una pluralidad de conjuntos de datos genéticos, cada uno de ellos medidos del material genético tomado del individuo diana, se consideran en paralelo, y así ambos conjuntos de datos genéticos diana contribuyen a determinar qué secciones de los datos genéticos parentales, medidos con una alta precisión, componen el genoma fetal.

En una realización de la divulgación, el método se puede utilizar para realizar las pruebas de paternidad. Por ejemplo, dada la información genotípica basada en el SNP de la madre y de un hombre que puede ser o no el padre genético, y la información genotípica medida de la muestra combinada, se puede determinar si la información genotípica del hombre representa de hecho la del padre genético real del feto en gestación. Una forma sencilla de hacerlo consiste simplemente en analizar los contextos donde la madre es AA y el posible padre es AB o BB. En estos casos, cabe esperar que la contribución del padre se observe la mitad (AA|AB) o la totalidad (AA|BB) del tiempo, respectivamente. Teniendo en cuenta la ADO prevista, resulta sencillo determinar si los SNP fetales observados están correlacionados o no con los del posible padre.

Una realización de la presente divulgación podría ser como sigue: una mujer embarazada quiere saber si su feto está afectado por el síndrome de Down y/o si sufrirá fibrosis quística, y no desea tener un hijo que esté afectado por

ninguna de estas condiciones. Un especialista le extrae sangre y tiñe la hemoglobina con un marcador de forma que aparezca claramente roja y tiñe el material nuclear con otro marcador para que aparezca claramente azul. Sabiendo que los glóbulos rojos maternos son típicamente anucleares, mientras que una elevada proporción de células fetales contienen un núcleo, el médico es capaz de aislar visualmente una serie de glóbulos rojos nucleados identificando las células que presentan tanto el color rojo como el azul. El médico recoge estas células del portaobjetos con un micromanipulador y las envía a un laboratorio que amplifica y determina los genotipos de diez células individuales. Utilizando las mediciones genéticas, el método PARENTAL SUPPORT™ es capaz de determinar que seis de las células son células sanguíneas maternas y cuatro de las diez células son células fetales. Si una mujer embarazada ya ha tenido un hijo, PARENTAL SUPPORT™ se puede utilizar también para determinar que las células fetales son distintas de las células del niño nacido marcando las determinaciones de alelos fiables en las células fetales y mostrando que son distintas de las del niño nacido. Cabe señalar que el concepto de este método es similar al de la realización de la prueba de paternidad de la presente divulgación. Los datos genéticos medidos de las células fetales pueden tener muy mala calidad, al comprender muchas pérdidas de alelos, debido a la dificultad de determinar el genotipo de células únicas. El médico es capaz de utilizar el ADN fetal medido junto con las mediciones de ADN fiables de los progenitores para deducir aspectos del genoma del feto con una alta precisión utilizando PARENTAL SUPPORT™, transformando así los datos genéticos contenidos en el material genético del feto en el estado genético previsto del feto, almacenado por ordenador. El médico es capaz de determinar tanto el estado de ploidía del feto como la presencia o ausencia de una pluralidad de genes asociados con la enfermedad de interés. Resulta que el feto es euploide y no es portador de la fibrosis quística y la madre decide continuar el embarazo.

En una realización de la presente divulgación, a una mujer embarazada le gustaría determinar si su feto está afectado por alguna anomalía cromosómica. Acude al médico y se le extrae una muestra de sangre, y se toman muestras a ella y su pareja de su propio ADN con hisopos bucales. Un investigador del laboratorio determina los genotipos del ADN parental utilizando el protocolo MDA para amplificar el ADN parental y arrays ILLUMINA INFINIUM para medir los datos genéticos de los progenitores en un gran número de SNP. A continuación, el investigador centrifuga la sangre, recoge el plasma y aísla una muestra de ADN flotante libre utilizando la cromatografía de exclusión por tamaño. Alternativamente, el investigador utiliza uno o más anticuerpos fluorescentes, como uno que es específico para la hemoglobina fetal, para aislar un glóbulo rojo fetal nucleado. A continuación, el investigador toma el material genético fetal aislado o enriquecido y lo amplifica utilizando una biblioteca de oligonucleótidos de 70-mer convenientemente diseñada de forma que dos extremos de cada oligonucleótido se correspondan con las secuencias que flanquean cualquier lado de un alelo diana. Tras la adición de una polimerasa, ligasa y los reactivos adecuados, los oligonucleótidos se sometieron a circularización con llenado de huecos, capturando el alelo deseado. Se añadió una exonucleasa, inactivada por calor, y los productos fueron utilizados directamente como plantilla para la amplificación por PCR. Los productos de la PCR fueron secuenciados en un ILLUMINA GENOME ANALYZER. Las lecturas de secuencia se utilizaron como datos de entrada para el método PARENTAL SUPPORT™, que posteriormente predijo el estado de ploidía del feto.

En otra realización de la divulgación, una pareja —en la que la madre, que está embarazada, tiene una edad materna avanzada— quiere saber si el feto en gestación padece síndrome de Down, síndrome de Turner, síndrome de Prader Willi o alguna otra anomalía cromosómica. El obstetra toma una muestra de sangre a la madre y al padre. La sangre se envía a un laboratorio, donde un técnico centrifuga la muestra materna para aislar el plasma y la capa leucocitaria. El ADN de la capa leucocitaria y la muestra de sangre paterna se transforman mediante amplificación y los datos genéticos codificados en el material genético amplificado se transforman también a partir de los datos genéticos molecularmente almacenados en datos genéticos electrónicamente almacenados procesando el material genético en un secuenciador de alto rendimiento para medir los genotipos parentales. La muestra de plasma es enriquecida preferentemente en un conjunto de loci utilizando un método de PCR focalizada hemi-anidada de 5000-plex. La mezcla de fragmentos de ADN se prepara creando una biblioteca de ADN adecuada para la secuenciación. A continuación, el ADN es secuenciado utilizando un método de secuenciación de alto rendimiento, por ejemplo, el ILLUMINA GAIIx GENOME ANALYZER. La secuenciación transforma la información que está codificada molecularmente en el ADN en información que está codificada electrónicamente en un software informático. Una técnica basada en la informática que incluye las realizaciones divulgadas en el presente documento, como PARENTAL SUPPORT™, puede ser utilizada para determinar el estado de ploidía del feto. Esto puede implicar el cálculo, por ordenador, de probabilidades del recuento de alelos en la pluralidad de loci polimórficos de las mediciones de ADN realizadas en la muestra preparada; la creación, por ordenador, de una pluralidad de hipótesis de ploidía pertenecientes cada una de ellas a un posible estado de ploidía diferente del cromosoma; la creación por ordenador, de un modelo de distribución conjunto para los recuentos de alelos previstos en la pluralidad de loci polimórficos del cromosoma para cada una de las hipótesis de ploidía; la determinación, por ordenador, de una probabilidad relativa de cada una de las hipótesis de ploidía utilizando el modelo de distribución conjunto y los recuentos de alelos medidos en la muestra preparada; y la determinación del estado de ploidía del feto seleccionando el estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada. Se determina que el feto tiene síndrome de Down. Se imprime un informe o se envía por medios electrónicos al obstetra, que transmite el diagnóstico a la mujer. La mujer, su pareja y el médico se sientan a hablar sobre sus opciones. La pareja decide poner fin al embarazo basándose en el conocimiento de que el feto está afectado por una patología trisómica.

En una realización de la divulgación, una empresa puede decidir ofrecer una tecnología de diagnóstico diseñada para detectar la aneuploidía en un feto en gestación a partir de una muestra de sangre materna. Su producto puede implicar que una madre visite a la consulta de su obstetra, donde se le extrae una muestra de sangre. El obstetra

también puede recoger una muestra genética del padre del feto. Un médico puede aislar el plasma de la sangre materna y purificar el ADN del plasma. Un médico también puede aislar la capa leucocitaria de la sangre materna y preparar el ADN de la capa leucocitaria. Un médico también puede preparar ADN de la muestra genética paterna. El médico puede utilizar las técnicas de biología molecular descritas en la presente divulgación para unir etiquetas de amplificación universales al ADN obtenido de la muestra de plasma. El clínico puede amplificar el ADN etiquetado universalmente. El médico puede enriquecer de forma preferente el ADN a través de diversas técnicas incluyendo la captura mediante hibridación y PCR focalizada. La PCR focalizada puede implicar un método de anidado, hemi-anidado o semi-anidado, o cualquier otro método para conseguir un enriquecimiento eficiente del ADN obtenido del plasma. La PCR focalizada puede ser masivamente multiplexada, por ejemplo, con 10 000 cebadores en una reacción, donde los cebadores van dirigidos a los SNP de los cromosomas 13, 18, 21, X y aquellos loci que son comunes tanto para X como para Y, y opcionalmente también para otros cromosomas. El enriquecimiento selectivo y/o la amplificación pueden implicar el etiquetado de cada molécula individual con diferentes etiquetas, códigos de barras moleculares, etiquetas para la amplificación y/o etiquetas para la secuenciación. A continuación, el médico puede secuenciar la muestra de plasma y también posiblemente el ADN materno y/o paterno preparado. Los pasos de biología molecular se pueden ejecutar parcial o totalmente por una caja de diagnóstico. Los datos de la secuencia se pueden introducir en un único ordenador u otro tipo de plataforma informática, como las que se pueden encontrar en "la nube". La plataforma informática puede calcular recuentos de alelos en los loci polimórficos focalizados de las mediciones realizadas por el secuenciador. La plataforma informática puede crear una pluralidad de hipótesis de ploidía correspondientes a la nulisomía, monosomía, disomía, trisomía emparejada, y trisomía no emparejada para cada uno de los cromosomas 13, 18, 21, X e Y. La plataforma informática puede crear un modelo de distribución conjunto para los recuentos de alelos previstos en los loci focalizados del cromosoma para cada una de las hipótesis de ploidía para cada uno de los cinco cromosomas que se van a valorar. La plataforma informática puede determinar una probabilidad de que cada una de las hipótesis de ploidía sea cierta utilizando el modelo de distribución conjunto y los recuentos de alelos medidos en el ADN enriquecido preferentemente obtenido de la muestra de plasma. La plataforma informática puede determinar el estado de ploidía del feto, para cada uno de los cromosomas 13, 18, 21, X e Y, seleccionando el estado de ploidía correspondiente a la hipótesis relevante con la probabilidad más alta. Se puede generar un informe que comprende los estados de ploidía determinados, que puede ser enviado al obstetra por medios electrónicos, desplegado en un dispositivo o se puede imprimir una copia en papel del informe que se puede entregar al obstetra. El obstetra puede informar a la paciente y opcionalmente al padre del feto, y estos pueden decidir las opciones clínicas que tienen y cuál es la más recomendable.

En otra realización de la divulgación, una mujer embarazada, en adelante denominada "la madre" puede decidir que desea saber si su feto o fetos portan o no alguna anomalía genética u otras patologías. Es posible que ella quiera asegurarse de que no hay anomalías importantes antes de decidir si quiere continuar con el embarazo. La madre puede acudir a su obstetra para que le tome una muestra de sangre. También puede tomar una muestra genética, por ejemplo, con un hisopo bucal del interior la boca. Es posible que también quiera tomar una muestra genética del padre del feto, por ejemplo, con un hisopo bucal, una muestra de esperma o una muestra de sangre. El obstetra puede enviar las muestras a un médico. El médico puede enriquecer la fracción de ADN fetal flotante libre de la muestra de sangre materna. El médico puede enriquecer la fracción de células sanguíneas fetales enucleadas que contiene la muestra de sangre materna. El médico puede utilizar diversos aspectos de los métodos descritos en el presente documento para determinar los datos genéticos del feto. Esos datos genéticos pueden incluir el estado de ploidía del feto y/o la identidad de una o una serie de alelos relacionados con enfermedades en el feto. Se puede generar un informe resumiendo los resultados del diagnóstico prenatal. El informe puede ser transmitido o enviado por correo al especialista, que podrá informar a la madre del estado genético del feto. La madre podrá decidir interrumpir el embarazo basándose en el hecho de que el feto presente una o más anomalías cromosómicas o genéticas, o condiciones no deseables. También puede decidir continuar el embarazo basándose en el hecho de que el feto no presenta ninguna anomalía cromosómica o genética grave, ni condiciones genéticas de interés.

Otro ejemplo puede implicar una mujer embarazada que se ha sometido a una inseminación artificial con esperma de un donante. Quiere minimizar el riesgo de que el feto tenga una enfermedad genética. Se va a extraer sangre a un especialista y las técnicas descritas en la presente divulgación se utilizan para aislar tres glóbulos rojos fetales nucleados, y también se toma una muestra de tejido de la madre y del padre genético. El material genético del feto y de la madre y el padre se amplifican cuando corresponde y se determina el genotipo utilizando ILLUMINA INFINIUM BEADARRAY, y los métodos descritos en el presente documento limpian y determinan la fase del genotipo parental y fetal con una elevada precisión, además de determinar el estado de ploidía del feto. Se descubre que el feto es euploide y las susceptibilidades fenotípicas se predicen a partir del genotipo fetal reconstruido, y se genera un informe que es enviado al médico de la madre para que puedan decidir cuál es la mejor decisión clínica.

En una realización de la divulgación, el material genético bruto de la madre y del padre se transforma a través de la amplificación en una cantidad de ADN que tiene una secuencia similar pero en una cantidad mayor. A continuación, a través de un método de determinación del genotipo, los datos genotípicos codificados por los ácidos nucleicos se transforman en mediciones genéticas que pueden ser almacenadas física y/o electrónicamente en un dispositivo de memoria, como los que se han descrito anteriormente. Los algoritmos relevantes que componen el algoritmo de PARENTAL SUPPORT™, cuyas partes relevantes se exponen detalladamente en el presente documento, se trasladan a un programa informático utilizando un lenguaje de programación. A continuación, ejecutando el programa informático en el hardware informático, en lugar de ser bits y bytes físicamente codificados, dispuestos en un patrón que representa datos de medición en bruto, se transforman en un patrón que representa una determinación de alta

certeza del estado de ploidía del feto. Los detalles de esta transformación se basarán en los propios datos y en el lenguaje informático y el sistema de hardware utilizados para ejecutar el método descrito en el presente documento. A continuación, los datos que están físicamente configurados para representar una determinación del estado de ploidía de alta calidad del feto se transforman en un informe que puede ser enviado a un profesional sanitario. Esta transformación se puede realizar utilizando una impresora o una pantalla de ordenador. El informe puede ser una copia impresa, en papel u otro medio adecuado, o en soporte electrónico. En el caso de un informe electrónico, puede ser transmitido, puede ser físicamente almacenado en un dispositivo de memoria en una ubicación del ordenador al que puede acceder el profesional sanitario; o también puede ser desplegado en una pantalla para que pueda ser leído. En caso de que se representen en una pantalla, los datos pueden ser transformados en un formato legible causando la transformación física de los píxeles en el dispositivo de visualización. La transformación se puede realizar activando físicamente los electrones en una pantalla fosforescente, alterando una carga eléctrica que cambia físicamente la transparencia de un conjunto específico de píxeles sobre una pantalla que puede encontrarse frente a un sustrato que emite o absorbe fotones. Esta transformación se puede realizar cambiando la orientación a nanoescala de las moléculas en un cristal líquido, por ejemplo, de una fase nemática a colestérica o esméctica, en un conjunto específico de píxeles. Esta transformación se puede conseguir por corriente eléctrica que provoca que los fotones se emitan desde un conjunto específico de píxeles producidos a partir de una pluralidad de diodos emisores de luz dispuestos en un patrón significativo. Esta transformación se puede realizar a través de cualquier otra forma utilizada para desplegar información, como la pantalla de un ordenador o algún otro dispositivo de salida o manera de transmitir información. A continuación, el profesional sanitario puede actuar en función del informe, de forma que los datos del informe se transforman en una acción. La acción puede ser continuar o interrumpir el embarazo, en cuyo caso un feto en gestación con una anomalía genética se transformará en un feto no vivo. Las transformaciones enumeradas en el presente documento se pueden agregar, de forma que, por ejemplo, se puede transformar el material genético de una madre embarazada y el padre, a través de una serie de pasos expuestos en esta divulgación, en una decisión médica consistente en abordar el feto con anomalías genéticas o continuar el embarazo. Alternativamente, se puede transformar un conjunto de mediciones genotípicas en un informe que ayuda a un médico a tratar a esta paciente embarazada.

En una realización de la presente divulgación, el método descrito en el presente documento se puede utilizar para determinar el estado de ploidía de un feto incluso cuando la madre hospedadora, es decir la mujer embarazada, no es la madre biológica del feto que porta. En una realización de la presente divulgación, el método descrito en el presente documento se puede utilizar para determinar el estado de ploidía de un feto que utiliza únicamente la muestra de sangre materna, y sin necesidad de una muestra genética paterna.

Parte de las matemáticas de las realizaciones de la presente realización realizan las hipótesis relativas a un número limitado de estados de aneuploidía. En algunos ejemplos, por ejemplo, se espera que solo cero, uno o dos cromosomas procedan de cada progenitor. En algunas realizaciones de la presente divulgación, las derivaciones matemáticas se pueden ampliar para tener en cuenta otras formas de aneuploidía, como la cuatrisomía, donde tres cromosomas proceden de un progenitor, la pentasomía, la hexasomía, etc., sin variar los conceptos fundamentales de la presente divulgación. Al mismo tiempo, es posible centrarse en un número más reducido de estados de ploidía, por ejemplo, solo trisomía y disomía. Cabe señalar que las determinaciones del estado de ploidía que indican un número no entero de cromosomas pueden indicar moisaicismo en una muestra de material genético.

En algunas realizaciones de la divulgación, la anomalía genética es un tipo de aneuploidía, como el síndrome de Down (o trisomía 21), el síndrome de Edwards (trisomía 18), síndrome de Patau (trisomía 13), síndrome de Turner (45X), síndrome de Klinefelter (un hombre con 2 cromosomas X), síndrome de Prader-Willi y síndrome de DiGeorge (UPD15). Por lo general, los trastornos congénitos, como los enumerados en la frase anterior, suelen ser indeseados, y saber que un feto está afectado por una o más anomalías fenotípicas puede proporcionar la base para la decisión de poner fin a un embarazo, para tomar las precauciones necesarias para prepararse para el nacimiento de un niño con necesidades especiales, o para adoptar algún método terapéutico orientado a aliviar la gravedad de una anomalía cromosómica.

En algunas realizaciones, los métodos descritos en el presente documento se pueden utilizar en una edad gestacional muy temprana, por ejemplo, tan solo a las cuatro semanas, tan solo a las cinco semanas, tan solo a las seis semanas, tan solo a las siete semanas, tan solo a las ocho semanas, tan solo a las nueve semanas, tan solo a las diez semanas, tan solo a las once semanas y tan solo a las doce semanas.

Cabe señalar que se ha demostrado que el ADN procedente del cáncer que vive en un huésped se puede encontrar en la sangre del huésped. Del mismo modo que se pueden realizar diagnósticos genéticos con la medición del ADN mezclado que se encuentra en la sangre materna, también se pueden realizar diagnósticos genéticos con la medición del ADN mezclado que se encuentra en la sangre del hospedador. Los diagnósticos genéticos pueden incluir estados de aneuploidía o mutaciones genéticas. Cualquier reivindicación de la divulgación instantánea relativa a la determinación del estado de ploidía o del estado genético de un feto a partir de las mediciones realizadas con la sangre materna se puede aplicar igualmente a la determinación del estado de ploidía o del estado genético de un cáncer a partir de las mediciones realizadas con la sangre del hospedador.

En algunas realizaciones, un método de la presente divulgación permite determinar el estado de ploidía de un cáncer. El método incluye la obtención de una muestra combinada que contiene material genético del hospedador y material genético del cáncer, y la medición del ADN en la muestra combinada; el cálculo de la fracción de ADN de

origen cancerígeno en la muestra combinada; y la determinación del estado de ploidía del cáncer utilizando las mediciones realizadas en la muestra combinada y la fracción calculada. En algunas realizaciones, el método puede incluir también la administración de un terapéutico para el cáncer basado en la determinación del estado de ploidía del cáncer. En algunas realizaciones, el método puede incluir también la administración de un terapéutico para el cáncer basado en la determinación del estado de ploidía del cáncer, donde el terapéutico se toma del grupo compuesto por un producto farmacéutico, un terapéutico biológico y una terapia con anticuerpos o una combinación de estos.

En algunas realizaciones, se utiliza un método divulgado en el presente documento en el contexto del diagnóstico genético previo a la implantación (PGD) para la selección del embrión durante la fertilización in vitro, donde el individuo diana es un embrión, y los datos genotípicos parentales se pueden utilizar para realizar determinaciones del estado de ploidía sobre el embrión a partir de los datos de secuencia de una biopsia de una o dos células de un embrión de tres días o una biopsia del trofectodermo de un embrión de cinco o seis días. En el contexto del PGD, solamente se mide el ADN del niño y solamente se somete a ensayo una pequeña cantidad de células, por lo general entre una y cinco, aunque pueden ser hasta diez, veinte o cincuenta. A continuación, se determina el número total de copias de inicio de los alelos A y B (en el SNP) de forma trivial a través del genotipo del niño y del número de células. En el NPD, el número de copias de inicio es muy elevado y por tanto se espera que el ratio de alelos tras la PCR refleje de forma precisa el ratio de partida. Sin embargo, el reducido número de copias de inicio en el PGD hace que la contaminación y la escasa eficiencia de la PCR tengan un efecto no trivial sobre el ratio de alelos tras la PCR. Este efecto puede ser más importante que la profundidad de la lectura en la predicción de la varianza en el ratio de alelos medido tras la secuenciación. La distribución del ratio de alelos medido dado un genotipo conocido del niño se puede crear mediante simulación de Monte Carlo del proceso de PCR basándose en la eficiencia de la sonda de la PCR y en la probabilidad de contaminación. Dada una distribución del ratio de alelos para cada posible genotipo del niño, la probabilidad de diversas hipótesis se puede calcular tal y como se describe para el NIPD.

Cualquiera de las realizaciones divulgadas en el presente documento se puede implementar en circuitos electrónicos digitales, circuidos integrados, ASIC (circuitos integrados de aplicación específica) especialmente diseñados, hardware informático, firmware, software o combinaciones de estos. El aparato de las realizaciones divulgadas en el presente documento se puede implementar en un producto de programa informático realizado de forma tangible en un dispositivo de almacenamiento legible electrónicamente para su ejecución por un procesador programable; y los pasos del método de las realizaciones divulgadas en el presente documento se pueden realizar mediante un procesador programable que ejecuta un programa de instrucciones para realizar las funciones de las realizaciones divulgadas en el presente documento operando sobre los datos de entrada y generando datos de salida. Las realizaciones divulgadas en el presente documento se pueden implementar ventajosamente en uno o más programas informáticos que son ejecutables y/o interpretables en un sistema programable, que incluirá al menos un procesador programable, que puede tener un propósito especial o general, conectado para recibir datos e instrucciones de un sistema de almacenamiento, así como para transmitir datos e instrucciones a este sistema, al menos un dispositivo de entrada, y al menos un dispositivo de salida. Cada programa informático se puede implementar en un lenguaje de programación orientado al objeto o procedural de alto nivel o en un lenguaje ensamblador o lenguaje máquina si se desea; y en cualquier caso el lenguaje puede ser un lenguaje compilado o interpretado. Se puede desplegar un programa informático en cualquier forma, incluyendo como programa independiente, o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Se puede desplegar un programa informático para que sea ejecutado o interpretado por ordenador o múltiples ordenadores en un punto, o distribuido en múltiples puntos e interconectado a través de una red de comunicación.

Los medios de almacenamiento legibles por ordenador, como los utilizados en el presente documento, se refieren a un almacenamiento físico o tangible (en oposición a las señales) e incluyen, entre otros, medios volátiles y no volátiles, extraíbles y no extraíbles, implementados en cualquier método o tecnología para el almacenamiento tangible de información como instrucciones legibles por ordenador, estructuras de datos, módulos de programas y otros datos., Los medios de almacenamiento legibles por ordenador incluyen, entre otros, RAM, ROM, EPROM, EEPROM, memoria flash y otra tecnología de memoria de estado sólido, CD-ROM, DVD, u otro almacenamiento óptico, casetes magnéticos, cintas magnéticas, almacenamiento en discos magnéticos u otros dispositivos de almacenamiento magnético, o cualquier otro medio físico o material que se pueda utilizar para almacenar de forma tangible la información deseada o los datos e instrucciones y al que se pueda acceder a través de un ordenador o procesador.

Cualquiera de los métodos descritos en el presente documento puede incluir la producción de datos en un formato físico, como la pantalla de un ordenador o una copia impresa en papel. En las explicaciones de las realizaciones que se recogen en otro apartado de este documento, se entenderá que los métodos descritos se pueden combinar con la producción de datos ejecutables en un formato sobre el que el médico puede actuar. Por otra parte, los métodos descritos se pueden combinar con la ejecución real de una decisión clínica que resulta en un tratamiento clínico o la ejecución de una decisión clínica de no emprender medidas. Algunas de las realizaciones descritas en el documento para determinar los datos genéticos pertenecientes a un individuo diana se pueden combinar con la decisión de seleccionar uno o más embriones para la transferencia en el contexto de la IVF, opcionalmente combinada con el proceso de transferencia del embrión al útero de la futura madre. Algunas de las realizaciones descritas en el

documento para determinar los datos genéticos pertenecientes a un individuo diana se pueden combinar con la notificación de una potencial anomalía cromosómica, o la ausencia de esta, a un profesional médico, combinada opcionalmente con la decisión de abortar o no un feto en el contexto del diagnóstico prenatal. Algunas de las realizaciones descritas en el presente documento se pueden combinar con la producción de datos ejecutables y la ejecución de una decisión clínica que resulta en un tratamiento clínico o la ejecución de una decisión clínica de no emprender medidas.

*Enriquecimiento focalizado y secuenciación*

El uso de una técnica para enriquecer una muestra de ADN en un conjunto de loci diana seguida de la secuenciación como parte de un método no invasivo para la determinación del alelo prenatal o la determinación del estado de ploidía de conformidad con la presente divulgación puede ofrecer una serie de ventajas no esperadas. En algunas realizaciones de la presente divulgación, el método implica la medición de datos genéticos para su uso con un método basado en la informática, como PARENTAL SUPPORT™ (PS). El resultado último de algunas de las realizaciones son los datos genéticos ejecutables de un embrión o feto. Existen múltiples métodos que se pueden utilizar para medir los datos genéticos del individuo y/o los individuos relacionados como parte de los métodos realizados. En una realización, se divulga en el presente documento un método para enriquecer la concentración de un conjunto de alelos focalizados, donde el método comprende uno o más de los pasos siguientes: amplificación focalizada del material genético, adición de sondas de oligonucleótidos de loci específico, unión de cadenas de ADN especificadas, aislamiento de conjuntos de ADN deseados, eliminación de componentes no deseados de una reacción, detección de determinadas secuencias de ADN mediante hibridación, y detección de la secuencia de una o una pluralidad de cadenas de ADN a través de métodos de secuenciación de ADN. En algunos casos las cadenas de ADN se pueden referir a material genético diana, en algunos casos se pueden referir a cebadores, en algunos casos se pueden referir a secuencias sintetizadas o combinaciones de estos. Estos pasos se pueden llevar a cabo en diferentes órdenes. Dada la naturaleza altamente variable de la biología molecular, generalmente no resulta obvio qué métodos y combinaciones de pasos darán mal resultado o buen resultado en las diversas situaciones.

Por ejemplo, un paso de amplificación universal del ADN antes de la amplificación focalizada puede conferir múltiples ventajas, como la eliminación del riesgo de cuellos de botella y la reducción del sesgo alélico. El ADN se puede mezclar con una sonda de oligonucleótidos que se pueden hibridar con dos regiones adyacentes de la secuencia diana, una a cada lado. Tras la hibridación, los extremos de la sonda se pueden conectar añadiendo una polimerasa, un medio para la unión, y cualquier reactivo necesario para permitir la circularización de la sonda. Tras la circularización se puede añadir una exonucleasa para digerir el material genético no circularizado, seguido de la detección de la sonda circularizada. El ADN se puede mezclar con cebadores para la PCR que se pueden hibridar con dos regiones adyacentes de la secuencia diana, una a cada lado. Tras la hibridación, los extremos de la sonda se pueden conectar añadiendo una polimerasa, un medio para la unión, y cualquier reactivo necesario para completar la amplificación por PCR. El ADN amplificado o no amplificado puede ser focalizado por sondas de captura híbridas dirigidas a un conjunto de loci; tras la hibridación, la sonda puede ser localizada y separada de la mezcla para proporcionar una mezcla de ADN enriquecida en secuencias diana.

En algunas realizaciones la detección del material genético diana se puede realizar de forma multiplexada. El número de secuencias diana genéticas que se pueden ejecutar en paralelo pueden oscilar entre una y diez, diez y cien, cien y mil, mil y diez mil, diez mil y cien mil, cien mil y un millón o un millón y diez millones. Cabe señalar que la técnica existente incluye divulgaciones de reacciones por PCR multiplexadas que han prosperado y que implican conjuntos de hasta unos 50 o 100 cebadores y no más. Los intentos previos de multiplexar más de 100 cebadores por grupo han generado problemas significativos con reacciones secundarias no deseadas, como la formación de un cebador-dímero.

En algunas realizaciones, este método se puede utilizar para determinar el genotipo de una sola célula, un pequeño número de células, entre dos y cinco células, entre seis y diez células, entre diez y veinte células, entre veinte y cincuenta células, entre cincuenta y cien células, entre cien y mil células, o una pequeña cantidad de ADN extracelular, por ejemplo, entre uno y diez picogramos, entre diez y cien picogramos, entre cien picogramos y un nanogramo, entre uno y diez nanogramos, entre diez y cien nanogramos o entre cien nanogramos y un microgramo.

El uso de un método para focalizar determinados loci seguido de la secuenciación como parte de un método para la determinación de los alelos o la determinación del estado de ploidía puede conferir una serie de ventajas no esperadas. Algunos métodos por los que se puede focalizar el ADN, o enriquecer de forma preferente, incluyen el uso de sondas de circularización, sondas invertidas enlazadas (LIP, MIP), captura por métodos de hibridación como SURESELECT, y estrategias de amplificación como PCR focalizada o PCR mediada por unión.

En algunas realizaciones, un método de la presente divulgación implica la medición de datos genéticos para su uso con un método basado en la informática, como PARENTAL SUPPORT™ (PS). PARENTAL SUPPORT™ es un método basado en la informática para manipular datos genéticos, cuyos aspectos se describen en el presente documento. El resultado último de algunas de las realizaciones son los datos genéticos ejecutables de un embrión o feto, seguidos de una decisión clínica basada en los datos ejecutables. Los algoritmos utilizados por el método PS toman los datos genéticos medidos del individuo diana, normalmente un embrión o feto, y los datos genéticos medidos de individuos relacionados, y son capaces de incrementar la eficacia con la que se conoce el estado genético del individuo diana. En una realización, los datos genéticos medidos se utilizan en el contexto de la realización de determinaciones del estado de ploidía durante el diagnóstico genético prenatal. En una realización, los

datos genéticos medidos se utilizan en el contexto de la realización de determinaciones del estado de ploidía o determinaciones de los alelos en los embriones durante la fertilización in vitro. Existen múltiples métodos que se pueden utilizar para medir los datos genéticos del individuo y/o los individuos relacionados en los mencionados contextos. Los diferentes métodos comprenden una serie de pasos, pasos que a menudo implican la amplificación de material genético, adición de sondas de oligonucleótidos, unión de cadenas de ADN especificadas, aislamiento de series de ADN deseadas, eliminación de componentes no deseados de una reacción, detección de determinadas secuencias de ADN mediante hibridación, y detección de la secuencia de una o una pluralidad de cadenas de ADN a través de métodos de secuenciación de ADN. En algunos casos las cadenas de ADN se pueden referir a material genético diana, en algunos casos se pueden referir a cebadores, en algunos casos se pueden referir a secuencias sintetizadas o combinaciones de estos. Estos pasos se pueden llevar a cabo en diferentes órdenes. Dada la naturaleza altamente variable de la biología molecular, generalmente no resulta obvio qué métodos y combinaciones de pasos darán mal resultado o buen resultado en las diversas situaciones.

Cabe señalar que en teoría resulta posible focalizar cualquier número de loci en el genoma, desde uno hasta más de un millón de loci. En una muestra el ADN es sometido a focalización y a continuación secuenciado; el porcentaje de alelos leídos por el secuenciador se enriquecerá con respecto a su abundancia natural en la muestra. El grado de enriquecimiento puede ser desde un uno por ciento (o incluso menos), hasta multiplicado por diez, multiplicado por cien, multiplicado por mil o incluso por varios millones. En el genoma humano hay aproximadamente 3000 millones de pares de bases y los nucleótidos comprenden unos 75 millones de loci polimórficos. Cuando mayor sea el número de loci focalizados, menor será el grado de enriquecimiento posible. Cuanto menor sea el número de loci focalizados, mayor será el grado de enriquecimiento posible, y mayor la profundidad de lectura que se podrá conseguir en estos loci para un determinado número de lecturas de secuencia.

En una realización de la presente divulgación, la focalización o preferencia se puede centrar completamente en los SNP. En una realización, la focalización o preferencia se puede centrar en cualquier punto polimórfico. Existen una serie de productos de focalización comercializados para enriquecer exones. Sorprende que la focalización exclusiva de SNP o exclusiva de loci polimórficos resulte particularmente ventajosa cuando se utiliza un método para el NPD que se basa en las distribuciones de alelos. También hay métodos publicados para el NPD utilizando la secuenciación, por ejemplo, en la Patente USA 7 888 017, que implica un análisis del recuento leído donde el recuento leído se concentra en el recuento del número de lecturas que corresponden a un cromosoma dado, y donde las lecturas de la secuencia analizada no se han centrado en regiones del genoma que son polimórficas. Estos tipos de metodología que no se enfocan en alelos polimórficos no se beneficiarían tanto de la focalización o el enriquecimiento preferente de un conjunto de alelos.

En una realización de la presente divulgación, resulta posible utilizar un método de focalización que se centra en los SNP para enriquecer una muestra genética en regiones polimórficas del genoma. En una realización, resulta posible centrarse en un pequeño número de SNP, por ejemplo, entre 1 y 100 SNP, o un número mayor, por ejemplo, entre 100 y 1000, entre 1000, entre 10 000 y 100 000 o más de 100 000 SNP. En una realización, resulta posible centrarse en uno o un pequeño número de cromosomas que están correlacionados con nacimientos trisómicos vivos, por ejemplo, los cromosomas 13, 18, 21, X e Y, alguna combinación de estos.

En una realización, resulta posible enriquecer los SNP focalizados en un pequeño factor, por ejemplo, entre 1,01 y 100 veces, o en un factor mayor, por ejemplo, entre 100 y 1 000 000 de veces, o incluso más. En una realización de la presente divulgación, resulta posible utilizar un método de focalización para crear una muestra de ADN que es preferentemente enriquecida en las regiones polimórficas del genoma. En una realización, se puede utilizar este método para crear una mezcla de ADN con cualquiera de estas características donde la mezcla de ADN contiene ADN materno y también ADN fetal flotante libre. En una realización, resulta posible utilizar este método para crear una mezcla de ADN que presenta cualquier combinación de estos factores. Por ejemplo, el método descrito en el presente documento se puede utilizar para producir una mezcla de ADN que comprende ADN materno y ADN fetal, y que está preferentemente enriquecida en ADN que corresponde a 200 SNP, todos ellos ubicados en el cromosoma 18 o 21, y que están enriquecidos una media de 1000 veces. En otro ejemplo, se puede utilizar el método para crear una mezcla de ADN que está preferentemente enriquecida en 10 000 SNP, estando todos o prácticamente todos ubicados en los cromosomas 13, 18, 21, X e Y, y el enriquecimiento medio por loci es superior a 500 veces. Cualquiera de los métodos de focalización descritos en el presente documento puede ser utilizado para crear mezclas de ADN que están preferentemente enriquecidas en determinados loci.

En algunas realizaciones, un método de la presente divulgación incluye también la medición del ADN en la fracción mezclada utilizando un secuenciador de ADN de alto rendimiento, donde el ADN de la fracción mezclada contiene un número desproporcionado de secuencias de uno o más cromosomas, y donde el cromosoma o los cromosomas son tomados del grupo compuestos por el cromosoma 13, el cromosoma 18, el cromosoma 21, el cromosoma X, el cromosoma Y, y combinaciones de estos.

En el presente documento se divulgan tres métodos: PCR multiplexada, captura focalizada mediante hibridación y sondas invertidas enlazadas (LIP), que se pueden utilizar para obtener y analizar mediciones de un número suficiente de loci polimórficos de una muestra de plasma materno con el fin de detectar aneuploidía fetal; esto no significa que se excluyan otros métodos de enriquecimiento selectivo de loci focalizados. También se pueden utilizar otros métodos sin variar la esencia del método. En cada caso, el polimorfismo sometido a ensayo puede incluir polimorfismos de un solo nucleótido (SNP), pequeñas inserciones/deleciones, o STR. Un método preferible implica el

uso de SNP. Cada método produce datos de frecuencia alélica; los datos de frecuencia alélica para cada locus focalizado y/o las distribuciones de la frecuencia alélica conjuntas de estos loci se pueden analizar para determinar la ploidía del feto. Cada método tiene sus propias consideraciones debido al material de origen limitado y al hecho de que el plasma materno se compone de una mezcla de ADN materno y fetal. Este método se puede combinar con otros métodos para proporcionar una determinación más precisa. En una realización, este método se puede combinar con un método de recuento de secuencias como el que se describe en la Patente USA 7 888 017. Los métodos descritos también se podrían utilizar para detectar la paternidad del feto de forma no invasiva, utilizando muestras de plasma materno. Además, cada método se puede aplicar a otras mezclas de ADN o muestras de ADN puro para detectar la presencia o ausencia de cromosomas aneuploides, para determinar el genotipo de un gran número de SNP de muestras de ADN degradado, para detectar variaciones en el número de copias de segmentos (CNV), para detectar otros estados genotípicos de interés, o alguna combinación de los mismos.

*Medición precisa de las distribuciones alélicas en una muestra*

Los métodos de secuenciación actuales se pueden utilizar para estimar la distribución de los alelos en una muestra. Un método implica realizar muestreos aleatorios de secuencias de un conjunto de ADN, lo que se conoce como secuenciación por fuerza bruta. La proporción de un alelo concreto en los datos de secuencia normalmente es muy baja y se puede determinar por simple estadística. El genoma humano contiene aproximadamente 3000 millones de pares de bases. Por tanto, si el método de secuenciación utilizado realiza lecturas de 100 pb, un alelo concreto se medirá aproximadamente una vez cada 30 millones de lecturas de secuencias.

En una realización, se utiliza un método de la presente divulgación para determinar la presencia o ausencia de dos o más haplotipos diferentes que contienen el mismo conjunto de loci en una muestra de ADN de las distribuciones de alelos medidas de los loci de ese cromosoma. Los diferentes haplotipos podrían representar dos cromosomas homólogos diferentes de un individuo, tres cromosomas homólogos diferentes de un individuo trisómico, tres haplotipos homólogos diferentes de una madre y un feto, donde uno de los haplotipos es compartido entre la madre y el feto, tres o cuatro haplotipos de una madre y un feto, donde uno o dos de los haplotipos son compartidos entre la madre y el feto, u otras combinaciones. Los alelos que son polimórficos entre los haplotipos tienden a ser más informativos; sin embargo, cualesquiera alelos en los que la madre y el padre no sean ambos homocigotos para el mismo alelo producirán información útil a través de las distribuciones de alelos medidas, además de la información disponible de un simple análisis del recuento de las lecturas.

Sin embargo, la secuenciación por fuerza bruta de esta muestra es extremadamente ineficiente, dado que resulta en múltiples secuencias para regiones que no son polimórficas entre los diferentes haplotipos de la muestra, o corresponden a cromosomas que no son de interés, y por tanto no revelan ninguna información acerca de la proporción de los haplotipos diana. En el presente documento se describen métodos que se dirigen de forma específica y/o enriquecen preferentemente segmentos de ADN de la muestra que es más probable que sean polimórficos en el genoma, para aumentar la producción de información alélica obtenida a través de la secuenciación. Cabe señalar que para que las distribuciones alélicas medidas en una muestra enriquecida sean realmente representativas de las cantidades reales presentes en el individuo diana, resulta fundamental que el enriquecimiento preferente de un alelo en comparación con otro alelo de un loci determinado de los segmentos focalizados sea escaso o nulo. Los métodos actuales conocidos en la técnica para focalizar alelos polimórficos están diseñados para garantizar que se detecten al menos algunos de los alelos presentes. Sin embargo, estos métodos no han sido diseñados con el fin de medir distribuciones alélicas no sesgadas de los alelos polimórficos presentes en la mezcla original. No está claro que ningún método concreto de enriquecimiento de la diana sea capaz de producir una muestra enriquecida en la que las distribuciones alélicas medidas representen de forma precisa las distribuciones alélicas presentes en la muestra original no amplificada mejor que cualquier otro método. Aunque cabe esperar múltiples métodos de enriquecimiento, en teoría, para lograr este objetivo, un experto en la técnica sabrá que existe un importante sesgo estocástico y determinístico en la amplificación, la focalización y otros métodos de enriquecimiento preferente actuales. Una realización de un método descrito en el presente documento permite que una pluralidad de alelos que se encuentran en una muestra de ADN y que corresponden a un determinado locus del genoma sean amplificados o enriquecidos de forma preferente de forma que el grado de enriquecimiento de cada uno de los alelos sea prácticamente igual. Otra forma de decir esto es que el método permite aumentar la cantidad relativa de alelos presentes en la mezcla en su conjunto, al tiempo que el ratio entre los alelos correspondientes a cada locus sigue siendo básicamente el mismo que el que presentaba la mezcla original de ADN. Los métodos de enriquecimiento preferente de loci de la técnica existente pueden presentar sesgos alélicos de más del 1%, más del 2%, más del 5% e incluso más del 10%. El enriquecimiento preferente se puede deber a un sesgo de captura cuando se utiliza un método de captura por hibridación, o a un sesgo de amplificación que puede ser pequeño para cada ciclo, pero puede ser grande cuando se acumula durante 20, 30 o 40 ciclos. A efectos de la presente divulgación, para que el ratio se mantenga básicamente sin cambios es necesario que el ratio de los alelos de la mezcla original dividido por el ratio de los alelos de la mezcla resultante se sitúe entre 0,95 y 1,05, entre 0,98 y 1,02, entre 0,99 y 1,01, entre 0,995 y 1,005, entre 0,998 y 1,002, entre 0,999 y 1,001, o entre 0,9999 y 1,0001. Cabe señalar que el cálculo de los ratios de alelos presentado en el presente documento no se puede utilizar en la determinación del estado de ploidía del individuo diana y que es solo una métrica utilizada para medir el sesgo alélico.

En una realización de la divulgación, una vez que una mezcla se ha enriquecido preferentemente en el conjunto de loci diana, se puede secuenciar utilizando cualquiera de los instrumentos de secuenciación antiguos, actuales o de

nueva generación que secuencian una muestra clónica (una muestra generada a partir de una única molécula; algunos ejemplos incluyen ILLUMINA GAIIx, ILLUMINA HISEQ, LIFE TECHNOLOGIES SOLiD, 5500XL). Los ratios se pueden evaluar mediante secuenciación a través de los alelos específicos de la región diana. Las lecturas de esta secuenciación se pueden analizar y recontar en función del tipo de alelo y de los ratios de los diferentes alelos determinados en consecuencia. Para las variaciones que tienen entre una y unas cuantas bases de longitud, la detección de los alelos se realizará mediante secuenciación y es fundamental que la lectura de la secuenciación abarque el alelo en cuestión para evaluar la composición alélica de esa molécula capturada. El número total de moléculas capturadas sometidas a ensayo para el genotipo se puede aumentar incrementando la longitud de la lectura de secuenciación. La secuenciación completa de todas las moléculas garantizaría la recopilación de la cantidad máxima de datos disponibles en el conjunto enriquecido. Sin embargo, actualmente la secuenciación resulta cara y un método que puede medir distribuciones alélicas utilizando un número menor de lecturas de secuencia tendrá un gran valor. Por otra parte, existen limitaciones técnicas para la longitud máxima posible de lectura, así como limitaciones de precisión a medida que aumentan las longitudes de la lectura. Los alelos de máxima utilidad tendrán entre una y unas pocas bases de longitud, aunque teóricamente se puede utilizar cualquier alelo más corto que la longitud de la lectura de secuenciación. A pesar de que las variaciones alélicas se encuentran en todos los tipos, los ejemplos proporcionados en el presente documento se centran en los SNP o en las variantes que se encuentran a tan solo algunos pares de bases de distancia. Las variantes más largas como las variantes del número de copias segmentarias se pueden detectar mediante la adición de estas variaciones más pequeñas en muchos casos, dado que las recopilaciones completas del SNP interno del segmento están duplicadas. Las variantes más largas que unas pocas bases, como los STR, requieren una consideración especial y algunos métodos de focalización funcionan y otros no.

Existen múltiples métodos de focalización que se pueden utilizar para aislar de forma específica y enriquecer uno o una pluralidad de posiciones variables en el genoma. Típicamente, estas se basan en aprovechar la secuencia invariable que flanquea a la secuencia variable. Hay una técnica anterior relacionada con la focalización en el contexto de la secuenciación donde el sustrato es plasma materno (ver, por ejemplo, Liao et al., Clin. Chem. 2011; 57(1): pp. 92-101). Sin embargo, los métodos de la técnica anterior utilizan sondas de focalización dirigidas a exones y no se centran en focalizar regiones polimórficas del genoma. En una realización, un método de la presente divulgación implica el uso de sondas de focalización que se centran de forma exclusiva o de forma casi exclusiva en las regiones polimórficas. En una realización, un método de la presente divulgación implica el uso de sondas de focalización que se centran de forma exclusiva o de forma casi exclusiva en los SNP. En algunas realizaciones de la presente divulgación, los puntos polimórficos focalizados se componen de al menos un 10% de SNP, al menos un 20% de SNP, al menos un 30% de SNP, al menos un 40% de SNP, al menos un 50% de SNP, al menos un 60% de SNP, al menos un 70% de SNP, al menos un 80% de SNP, al menos un 90% de SNP, al menos un 95% de SNP, al menos un 98% de SNP, al menos un 99% de SNP, al menos un 99.9% de SNP, o exclusivamente SNP.

En una realización, se puede utilizar un método de la presente divulgación para determinar genotipos (composición de bases del ADN en un loci específico) y las proporciones relativas de estos genotipos a partir de una mezcla de moléculas de ADN, donde dichas moléculas de ADN pueden proceder de uno o de varios individuos genéticamente distintos. En una realización, se puede utilizar un método de la presente divulgación para determinar los genotipos de un conjunto de loci polimórficos y los ratios relativos de la cantidad de alelos diferentes presentes en esos loci.

En una realización, los loci polimórficos pueden estar compuestos completamente por SNP. En una realización, los loci polimórficos pueden comprender SNP, repeticiones en tándem simples y otros polimorfismos. En una realización, se puede utilizar un método de la presente divulgación para determinar las distribuciones relativas de alelos en un conjunto de loci polimórficos en una mezcla de ADN, donde la mezcla de ADN comprende ADN procedente de la madre y ADN procedente del feto. En una realización, las distribuciones alélicas conjuntas se pueden determinar en una muestra de ADN aislado de sangre de una mujer embarazada. En una realización, las distribuciones alélicas en un conjunto de loci se pueden utilizar para determinar el estado de ploidía de uno o más cromosomas de un feto en gestación.

En una realización de la divulgación, la mezcla de moléculas de ADN se podría obtener de ADN extraído de múltiples células de un individuo. En una realización, la recopilación original de células de las que se obtiene el ADN puede comprender una mezcla de células diploides o haploides del mismo o diferentes genotipos, si ese individuo presenta mosaicismo (germinal o somático). En una realización, la mezcla de moléculas de ADN también se podría obtener de ADN extraído de células únicas. En una realización, la mezcla de moléculas de ADN también se podría obtener de ADN extraído de una mezcla de dos o más células del mismo individuo o de diferentes individuos. En una realización, la mezcla de moléculas de ADN se podría obtener de ADN aislado de material biológico que ya ha sido liberado por células como el plasma sanguíneo, que se sabe que contienen cfADN. En una realización, este material biológico puede ser una mezcla de ADN de uno o más individuos, como sucede durante el embarazo, donde se ha demostrado que hay ADN fetal presente en la mezcla. En una realización, el material biológico podría proceder de una mezcla de células que se encuentran en la sangre materna, donde algunas de las células son de origen fetal. En una realización, el material biológico podrían ser células sanguíneas de una embarazada que han sido enriquecidas en células fetales.

*Sondas de circularización*

Algunas realizaciones de la presente divulgación implican el uso de Sondas Invertidas Enlazadas (LIP), que han sido anteriormente descritas en la bibliografía. LIP es un término genérico que pretende abarcar tecnologías que implican la creación de una molécula circular de ADN, donde las sondas están diseñadas para hibridarse con una región focalizada de ADN a uno de los lados de un alelo focalizado, de forma que la adición de las polimerasas y/o ligasas adecuadas, y las condiciones, tampones y otros reactivos adecuados, completará la región invertida complementaria del ADN del alelo focalizado para crear un bucle circular de ADN que captura la información que se encuentra en el alelo focalizado. Las LIP también se pueden denominar sondas precircularizadas, sondas de precircularización o sondas de circularización. Las LIP pueden ser una molécula de ADN lineal de entre 50 y 500 nucleótidos de longitud, y en una realización de entre 70 y 100 nucleótidos de longitud; en algunas realizaciones, puede ser más larga o más corta de lo que se describe en el presente documento. Otras realizaciones de la presente divulgación implican diferentes encarnaciones de la tecnología de las LIP, tales como sondas candado (padlock) y sondas de inversión molecular (MIP).

Un método para focalizar ubicaciones específicas para la secuenciación consiste en sintetizar sondas en las que los extremos 3' y 5' de las sondas se hibridan con ADN diana en ubicaciones adyacentes a cualquiera de los lados de la región focalizada, de manera invertida, de forma que la adición de ADN polimerasa y ADN ligasa resulta en la extensión desde el extremo 3', añadiendo bases a la sonda de cadena única que son complementarias de la molécula diana (relleno de huecos), seguida de la unión del nuevo extremo 3' al extremo 5' de la sonda original, lo que resulta en una molécula de ADN circular que posteriormente se puede aislar del ADN de fondo. Los extremos de la sonda están diseñados para flanquear la región de interés focalizada. Un aspecto de este método se denomina habitualmente MIPS y se ha utilizado conjuntamente con tecnologías de array para determinar la naturaleza de la secuencia rellenada. Una desventaja del uso de MIP en el contexto de la medición de los ratios de alelos es que los pasos de hibridación, circularización y amplificación no se producen a los mismos ratios para los diferentes alelos de los mismos loci. Esto resulta en unos ratios de alelos medidos que no son representativos de los ratios de alelos reales presentes en la mezcla original.

En una realización de la divulgación, las sondas de circularización están construidas de forma que la región de la sonda que está diseñada para hibridarse en sentido ascendente del locus polimórfico focalizado y la región de la sonda que está diseñada para hibridarse en sentido descendente del locus polimórfico focalizado están conectados covalentemente a través de una estructura de ácido no nucleico. Esta estructura puede ser cualquier molécula biocompatible o combinación de moléculas biocompatibles. Algunos ejemplos de posibles moléculas biocompatibles son poli(etilen glicol), policarbonatos, poliuretanos, polietilenos, polipropilenos, polímeros sulfonados, silicona, celulosa, fluoropolímeros, compuestos acrílicos, copolímeros de bloque estirénicos y otros copolímeros de bloque.

En una realización de la presente divulgación, este método ha sido modificado para que se pueda utilizar fácilmente para la secuenciación como medio de analizar la secuencia rellenada. A fin de conservar las proporciones alélicas originales de la muestra original, se debe tener en cuenta al menos una consideración clave. Las posiciones variables entre diferentes alelos en la región de relleno de huecos no deben estar demasiado cerca de los puntos de unión de la sonda, dado que puede darse un sesgo iniciador de ADN polimerasa que provoque un diferencial de las variantes. Otra consideración es que puede haber otras variantes presentes en los puntos de unión de la sonda que están correlacionadas con las variantes de la región de relleno de huecos, lo que puede producir una amplificación heterogénea de los diferentes alelos. En una realización de la presente divulgación, los extremos 3' y 5' de la sonda precircularizada están diseñados para hibridarse con bases que se encuentran a una o unas cuantas posiciones de distancia de las posiciones variables (puntos polimórficos) del alelo focalizado. El número de bases entre el punto polimórfico (SNP u otro) y la base a cuyo extremo 3' o 5' se hibridará la sonda precircularizada puede ser de una base, puede ser de dos bases, puede ser de tres bases, puede ser de cuatro bases, puede ser de cinco bases, puede ser de seis bases, puede ser de siete a diez bases, puede ser de once a quince bases, o puede ser de dieciséis a veinte bases, de veinte a treinta bases, o de treinta a sesenta bases. Los cebadores directos e inversos pueden estar diseñados para hibridarse a un número de bases de distancia del punto polimórfico. Las sondas de circularización se pueden generar en grandes cantidades con la tecnología de síntesis de ADN actual, lo que permite producir cantidades muy importantes de sondas y potencialmente agruparlas, permitiendo el análisis de múltiples loci de forma simultánea. Se ha documentado que funciona con más de 300 000 sondas. Dos documentos que exponen un método que implica sondas de circularización que se pueden utilizar para medir los datos genómicos del individuo diana incluyen: Porreca et al., Nature Methods, 2007 4(11), pp. 931-936.; y también Turner et al., Nature Methods, 2009, 6(5), pp. 315-316. Los métodos descritos en estos documentos se pueden utilizar en combinación con otros métodos descritos en el presente documento. Determinados pasos del método de estos dos documentos se pueden utilizar en combinación con otros pasos de otros métodos que se describen en el presente documento.

En algunas realizaciones de los métodos divulgados en el presente documento, el material genético del individuo diana es opcionalmente amplificado, seguido de una hibridación de las sondas precircularizadas, realizando un relleno de huecos para rellenar las bases entre los dos extremos de las sondas hibridadas, uniendo los dos extremos para formar una sonda circularizada, y amplificando la sonda circularizada, por ejemplo, mediante amplificación por círculo rodante. Una vez que se ha capturado la información genética de los alelos diana deseados mediante la circularización de las sondas de oligonucleótidos convenientemente diseñadas, como en el sistema LIP, la secuencia genética de las sondas circularizadas puede ser medida para obtener los datos de la secuencia deseada. En una realización, las sondas de oligonucleótidos convenientemente diseñadas pueden ser circularizadas directamente en el material genético no amplificado del individuo diana y amplificarse posteriormente. Cabe señalar que se pueden

utilizar diversos procedimientos de amplificación para amplificar el material genético original, o las LIP circularizadas, incluyendo la amplificación por círculo rodante, la MDA u otros protocolos de amplificación. Se pueden utilizar diferentes métodos para medir la información genética del genoma diana, por ejemplo, utilizando una secuenciación de alto rendimiento, secuenciación de Sanger, otros métodos de secuenciación, captura por hibridación, captura por circularización, PCR multiplexada, otros métodos de hibridación y combinaciones de estos.

Una vez que el material genético del individuo se ha medido utilizando uno o una combinación de los métodos anteriores, se puede utilizar un método informático, como el método PARENTAL SUPPORT™, junto con las mediciones genéticas correspondientes, para la determinación del estado de ploidía de uno o más cromosomas del individuo, y/o el estado genético de uno o un conjunto de alelos, concretamente de aquellos alelos que están correlacionados con una enfermedad o estado genético de interés. Cabe señalar que el uso de LIP ha sido documentado para la captura multiplexada de secuencias genéticas, seguida de la determinación del genotipo mediante secuenciación. Sin embargo, los datos de secuencia resultantes de una estrategia basada en LIP para la amplificación del material genético que se encuentra en una única célula, una pequeña cantidad de células o ADN extracelular, no han sido utilizados al objeto de determinar el estado de ploidía de un individuo diana.

La aplicación de un método basado en la informática para determinar el estado de ploidía de un individuo a partir de los datos genéticos medidos mediante arrays de hibridación, como el array ILLUMINA INFINIUM o el chip genético AFFYMETRIX, ha sido descrita en documentos a los que se hace referencia en otros apartados del presente documento. Sin embargo, el método descrito en el presente documento muestra mejoras respecto de los métodos anteriormente descritos en la bibliografía. Por ejemplo, el método basado en LIP seguido de una secuenciación de alto rendimiento sorprendentemente proporciona unos mejores datos genotípicos debido a que el método tiene una mayor capacidad de multiplexado, una mejor especificidad de captura, una mejor uniformidad y un menor sesgo alélico. La mayor capacidad de multiplexado permite focalizar un mayor número de alelos, dando unos resultados más precisos. La mejora de la uniformidad permite la medición de un mayor número de los alelos focalizados, dando unos resultados más precisos. Los bajos índices de sesgo alélico permiten unas tasas más reducidas de errores en la determinación, dando unos resultados más precisos. Unos resultados más precisos suponen una mejora de los resultados clínicos y una mejor atención médica.

Es importante señalar que las LIP se pueden utilizar como método para focalizar loci específicos en una muestra de ADN para la determinación del genotipo a través de métodos distintos de la secuenciación. Por ejemplo, se pueden utilizar LIP con el ADN diana para determinar el genotipo utilizando arrays de SNP u otros microarrays basados en ADN o ARN.

*PCR mediada por unión*

La PCR mediada por unión es un método de PCR utilizado para enriquecer preferentemente una muestra de ADN amplificando uno de una pluralidad de loci en una mezcla de ADN, donde el método consiste en lo siguiente: obtener un conjunto de pares de cebadores, donde cada cebador del par contiene una secuencia específica diana y una secuencia no diana, donde la secuencia específica diana está diseñada para hibridarse con una región diana, en sentido ascendente y en sentido descendente del punto polimórfico. y que puede estar separado del punto polimórfico por 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-20, 21-30, 31-40, 41-50, 51-100, o más de 100; la polimerización del ADN desde el extremo 3' del cebador en sentido ascendente para rellenar la región de cadena simple entre este y el extremo 5' del cebador en sentido descendente con nucleótidos complementarios de la molécula diana; la unión de la última base polimerizada del cebador en sentido ascendente a la base del extremo 5' adyacente del cebador en sentido descendente; y amplificación solo de las moléculas polimerizadas y ligadas utilizando las secuencias no diana contenidas en el extremo 5' del cebador en sentido ascendente y el extremo 3' del cebador en sentido descendente. Los pares de cebadores para las distintas dianas pueden estar mezclados en la misma reacción. Las secuencias no diana sirven como secuencias universales de forma que todos los pares de cebadores que han sido polimerizados y ligados con éxito se pueden amplificar con un único par de cebadores de amplificación.

*Captura por hibridación*

El enriquecimiento preferente de un conjunto específico de secuencias en un genoma diana se puede realizar de múltiples maneras. En otro apartado de este documento se recoge una descripción de cómo se pueden utilizar las LIP para focalizar un conjunto específico de secuencias, pero en todas esas aplicaciones se pueden utilizar otros métodos de focalización y/o enriquecimiento preferente igualmente bien para los mismos extremos. Un ejemplo de otro método de focalización es la captura mediante un método de hibridación. Algunos ejemplos de tecnologías de captura por hibridación disponibles en el mercado incluyen AGILENT's SURE SELECT y TruSeq de ILLUMINA. En la captura por hibridación, se permite que se hibride un conjunto de oligonucleótidos que es complementario o en gran medida complementario de las secuencias focalizadas deseadas con una mezcla de ADN y, a continuación, se separan físicamente de la muestra. Una vez que las secuencias deseadas se han hibridado con los oligonucleótidos de focalización, el efecto de separar físicamente los oligonucleótidos de focalización incluye también la eliminación de las secuencias focalizadas. Una vez que se han retirado los oligonucleótidos hibridados, pueden ser calentados por encima de su temperatura de fusión y amplificados. Algunas maneras de eliminar físicamente los oligonucleótidos de focalización consisten en enlazar covalentemente los oligonucleótidos de focalización con un soporte sólido, por ejemplo, un gránulo magnético o un chip. Otra forma de eliminar físicamente los oligonucleótidos de focalización consiste en enlazarlos covalentemente con una fracción molecular con una fuerte afinidad por otra fracción molecular. Un ejemplo de este par de moléculas es biotina y estreptavidina, como el que se utiliza en SURE

SELECT. De esta forma, las secuencias focalizadas se podrían unir covalentemente a una molécula de biotina y, después de la hibridación, utilizar un soporte sólido con estreptavidina para hacer bajar los oligonucleótidos biotinilados, a los que se hibridan las secuencias focalizadas.

La captura híbrida implica la hibridación de sondas que son complementarias de las dianas de interés de las moléculas diana. Las sondas de captura híbrida fueron originalmente desarrolladas para focalizar y enriquecer fracciones grandes del genoma con una uniformidad relativa entre dianas. En esa aplicación, era importante que todas las dianas se amplificasen con una uniformidad suficiente para que todas las regiones se pudiesen detectar mediante secuenciación, aunque no se prestó atención a conservar la proporción de alelos de la muestra original. Tras la captura, los alelos presentes en la muestra se pueden determinar mediante secuenciación directa de las moléculas capturadas. Estas lecturas de secuenciación se pueden analizar y recontar en función del tipo de alelos. Sin embargo, utilizando la tecnología actual, las distribuciones alélicas medidas de las secuencias capturadas típicamente no son representativas de las distribuciones alélicas originales.

En una realización de la divulgación, la detección de los alelos se realiza mediante secuenciación. A fin de capturar la identidad alélica del punto polimórfico, es esencial que la lectura de secuenciación abarque el alelo en cuestión a fin de evaluar la composición alélica de esa molécula capturada. Dado que las moléculas capturadas suelen tener una longitud variable tras la secuenciación no se puede garantizar que solapen las posiciones variables a menos que se secuencie la molécula completa. Sin embargo, las consideraciones de costes y las limitaciones técnicas asociadas a la máxima longitud posible y a la precisión de las lecturas de secuenciación hacen que la secuenciación de la molécula completa resulte inviable. En una realización, la longitud de la lectura se puede aumentar desde unas 30 hasta unas 50 o unas 70 bases y puede aumentar en gran medida el número de lecturas que solapan las posiciones variables dentro de las secuencias focalizadas.

Otra forma de aumentar el número de lecturas que analizan la posición de interés consiste en reducir la longitud de la sonda, de forma que no resulte en un sesgo en los alelos enriquecidos subyacentes. La longitud de la sonda sintetizada debería ser suficiente para que dos sondas diseñadas para hibridarse con dos alelos diferentes que se encuentran en un locus se hibriden con una afinidad prácticamente igual a los diversos alelos de la muestra original. En la actualidad, los métodos conocidos en la técnica describen sondas que tienen típicamente más de 120 bases de largo. En una realización actual, si el alelo tiene entre una y unas pocas bases entonces las sondas de captura pueden tener menos de unas 110 bases, menos de unas 100 bases, menos de unas 90 bases, menos de unas 80 bases, menos de unas 70 bases, menos de unas 60 bases, menos de unas 50 bases, menos de unas 40 bases, menos de unas 30 bases y menos de unas 25 bases, y esto es suficiente para garantizar un enriquecimiento homogéneo de todos los alelos. Cuando la mezcla de ADN que se va a enriquecer utilizando la tecnología de captura híbrida es una mezcla que comprende ADN flotante libre aislado de una muestra de sangre, por ejemplo, de sangre materna, la longitud media del ADN es bastante corta, típicamente menos de 200 bases. El uso de sondas más cortas resulta en una mayor probabilidad de que las sondas de captura híbrida capturen los fragmentos de ADN deseados. Las variaciones más largas pueden requerir sondas más largas. En una realización, las variaciones de interés tienen entre una (un SNP) y unas pocas bases de longitud. En una realización, las regiones focalizadas del genoma se pueden enriquecer preferentemente utilizando sondas de captura híbrida donde las sondas de captura híbrida tienen una longitud inferior a 90 bases y pueden tener menos de 80 bases, menos de 70 bases, menos de 60 bases, menos de 50 bases, menos de 40 bases, menos de 30 bases o menos de 25 bases. En una realización, para aumentar la probabilidad de que el alelo deseado sea secuenciado, la longitud de la sonda diseñada para hibridarse con las regiones que flanquean la ubicación del alelo polimórfico se puede reducir desde más de 90 bases hasta unas 80 bases, o hasta unas 70 bases, o hasta unas 60 bases, o hasta unas 50 bases, o hasta unas 40 bases, o hasta unas 30 bases o hasta unas 25 bases.

Existe un solapamiento mínimo entre la sonda sintetizada y la molécula diana para permitir la captura. Esta sonda sintetizada se puede hacer tan corta como resulte posible, mientras siga siendo más larga que este solapamiento mínimo requerido. El efecto de utilizar una longitud de sonda más corta para focalizar una región polimórfica es que habrá más moléculas que solapen la región alélica diana. El estado de fragmentación de las moléculas de ADN originales también afecta al número de lecturas que solaparán los alelos focalizados. Algunas muestras de ADN, como muestras de plasma, ya están fragmentadas debido a procesos biológicos que se producen in vivo. Sin embargo, las muestras con fragmentos más largos pueden beneficiarse de una fragmentación previa a la secuenciación, la preparación de la biblioteca y el enriquecimiento. Cuando tanto las sondas como los fragmentos son cortos (unos 60-80 pares de bases), se puede conseguir una especificidad máxima, dado que un número relativamente pequeño de lecturas de secuencias no solapan la región de interés crítica.

En una realización de la divulgación, las condiciones de hibridación se pueden ajustar para maximizar la uniformidad en la captura de diferentes alelos presentes en la muestra original. En una realización, las temperaturas de hibridación se reducen para minimizar las diferencias en el sesgo de hibridación entre alelos. Los métodos conocidos en la técnica evitan utilizar temperaturas inferiores para la hibridación, porque la bajada de la temperatura tiene el efecto de aumentar la hibridación de las sondas con dianas no deseadas. Sin embargo, cuando el objetivo consiste en preservar los ratios de alelos con la máxima fidelidad, el método de utilizar bajas temperaturas de hibridación proporciona unos ratios de alelos óptimamente precisos, a pesar del hecho de que la técnica actual no recomienda este método. La temperatura de hibridación también se puede aumentar para exigir un mayor solapamiento entre la diana y la sonda sintetizada, de forma que solo se capturan las dianas con un solapamiento sustancial de la región

focalizada. En algunas realizaciones de la presente divulgación, la temperatura de hibridación se reduce desde la temperatura de hibridación normal hasta unos 40ºC, 45ºC, 50ºC, 55ºC, 60ºC, 65ºC o 70ºC.

En una realización de la divulgación, las sondas de captura híbrida pueden estar diseñadas de forma que la región de la sonda de captura con el ADN que sea complementario al ADN que se encuentra en las regiones que flanquean el alelo polimórfico no se encuentra inmediatamente adyacente al punto polimórfico.

En vez de esto, la sonda de captura puede estar diseñada de forma que la región de la sonda de captura que está diseñada para hibridarse con el ADN que flanquea el punto polimórfico de la diana está separado de la porción de la sonda de captura que se podrá en contacto conforme a las fuerzas de van der Waals con el punto polimórfico por una pequeña distancia que tiene una longitud equivalente a una o un pequeño número de bases. En una realización, la sonda de captura híbrida está diseñada para hibridarse con una región que flanquea el alelo polimórfico pero no lo cruza; esto se puede denominar sonda de captura de flanqueo. La longitud de la sonda de captura de flanqueo puede ser inferior a unas 120 bases, menos de unas 110 bases, menos de unas 100 bases, menos de unas 90 bases, menos de unas 80 bases, menos de unas 70 bases, menos de unas 60 bases, menos de unas 50 bases, menos de unas 40 bases, menos de unas 30 bases o menos de unas 25 bases. La región del genoma focalizada por la sonda de captura de flanqueo puede estar separada del locus polimórfico por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-20, o más de 20 pares de bases.

Descripción de una prueba de detección de una enfermedad basada en la captura focalizada utilizando una captura de secuencia focalizada. Captura de secuencia focalizada adaptada, como las que ofrecen actualmente AGILENT (SURE SELECT), ROCHE-NIMBLEGEN, o ILLUMINA. Las sondas de captura podrían diseñarse de forma adaptada para garantizar la captura de diversos tipos de mutaciones. Para las mutaciones puntuales, una o más sondas que solapan la mutación puntual deberían ser suficientes para capturar y secuenciar la mutación.

Para pequeñas inserciones o deleciones, una o más sondas que solapan la mutación pueden ser suficientes para capturar y secuenciar los fragmentos que comprenden la mutación. La hibridación puede ser menos eficiente por lo que respecta a la eficacia de captura con limitación de sonda, típicamente diseñada para la secuencia del genoma de referencia. Para garantizar la captura de fragmentos que comprenden la mutación, se podrían diseñar dos sondas, una correspondiente al alelo normal y otra correspondiente al alelo mutante. Una sonda más larga puede mejorar la hibridación. Múltiples sondas con solapamiento pueden mejorar la captura. Por último, la colocación de una sonda inmediatamente adyacente a la mutación, pero sin solapamiento, puede permitir una eficacia de captura relativamente similar de los alelos normales y mutantes.

En el caso de las repeticiones en tándem simples (STR), es poco probable que una sonda que solapa estos puntos altamente variables capture bien el fragmento. Para mejorar la captura se podría colocar una sonda adyacente al punto variable, aunque sin solapamiento. A continuación, se podría secuenciar el fragmento con normalidad para revelar la longitud y la composición de la STR.

En el caso de las grandes deleciones, puede funcionar el uso de una serie de sondas de solapamiento, un método habitual que se utiliza actualmente en los sistemas de captura de exomas. Sin embargo, con este método puede resultar difícil determinar si un individuo es heterocigoto o no. La focalización y evaluación de los SNP de la región capturada podrían potencialmente revelar la pérdida de heterocigosidad en la región, indicando que el individuo es portador. En una realización, es posible colocar sondas sin solapamiento o sondas únicas en la región potencialmente sometida a deleción y utilizar el número de fragmentos capturados como medida de la heterocigosidad. En caso de que un individuo porte una gran deleción, se espera que la mitad del número de fragmentos esté disponible para la captura con respecto a un locus de referencia sin deleción (diploide). Por consiguiente, el número de lecturas obtenidas de las regiones sometidas a deleción debería ser aproximadamente la mitad que el obtenido de un locus diploide normal. La agrupación y la determinación de la media de la profundidad de las lecturas de secuenciación de múltiples sondas únicas en una región potencialmente sometida a deleción pueden mejorar la señal y la certeza del diagnóstico. Los dos métodos, la focalización de SNP para identificar la pérdida de heterocigosidad y el uso de múltiples sondas únicas para obtener una medición cuantitativa de la cantidad de fragmentos subyacentes de ese locus, también se pueden combinar. Cualquiera de estas estrategias o ambas se pueden combinar con otras estrategias para obtener mejores resultados.

Si durante las pruebas del cfADN, la detección de un feto varón, indicado por la presencia de los fragmentos del cromosoma Y, capturados y secuenciados en la misma prueba, y una mutación dominante vinculada a X donde la madre y el padre no se ven afectados, o una mutación dominante donde la madre no se ve afectada, indicaría un mayor riesgo para el feto. La detección de dos alelos recesivos mutantes en el mismo gen en una madre no afectada implicaría que el feto ha heredado un alelo mutante del padre y potencialmente un segundo alelo mutante de la madre. En todos los casos, las pruebas de seguimiento mediante amniocentesis o muestreo del vello coriónico pueden estar indicadas.

Una prueba de detección de una enfermedad basada en la captura focalizada se podría combinar con una prueba de diagnóstico prenatal no invasiva basada en la captura focalizada para determinar la aneuploidía.

Existen diversas formas de reducir la variabilidad de la profundidad de lectura (DOR): por ejemplo, se podrían aumentar las concentraciones del cebador, se podrían utilizar sondas de amplificación focalizadas más largas, se podrían ejecutar más ciclos STA (como más de 25, más de 30, más de 35 o incluso más de 40).

*PCR focalizada*

En algunas realizaciones de la invención, la PCR se puede utilizar para focalizar ubicaciones específicas del genoma. En muestras de plasma, el ADN original está altamente fragmentado (típicamente menos de 500 pb, con una longitud media inferior a 200 pb). En la PCR, tanto los cebadores directos como inversos deben hibridarse con el mismo fragmento para permitir la amplificación. Por tanto, si los fragmentos son cortos, los ensayos de PCR deben amplificar también regiones relativamente cortas. Al igual que en las MIP, si las posiciones polimórficas están demasiado cerca del punto de unión de la polimerasa, podrían producirse sesgos en la amplificación de los diferentes alelos. En la actualidad, los cebadores de la PCR que focalizan regiones polimórficas, como aquellos que contienen SNP, están típicamente diseñados de forma que el extremo 3' del cebador se hibridará con la base inmediatamente adyacente a la base o las bases polimórficas. En una realización de la presente divulgación, los extremos 3' de los cebadores directos e inversos de la PCR están diseñados para hibridarse con bases que se encuentran a una o unas cuantas posiciones de distancia de las posiciones variables (puntos polimórficos) del alelo focalizado. El número de bases entre el punto polimórfico (SNP u otro) y la base a cuyo extremo 3' se hibridará el cebador puede ser de una base, puede ser de dos bases, puede ser de tres bases, puede ser de cuatro bases, puede ser de cinco bases, puede ser de seis bases, puede ser de siete a diez bases, puede ser de once a quince bases, o puede ser de dieciséis a veinte bases. Los cebadores directos e inversos pueden estar diseñados para hibridarse a un número de bases de distancia del punto polimórfico.

Los ensayos de PCR se pueden generar en grandes cantidades; sin embargo, las interacciones entre diferentes ensayos de PCR dificultan su multiplexado cuando se superan aproximadamente los cien ensayos. Se pueden utilizar diversos métodos moleculares complejos para aumentar el nivel de multiplexado, pero aun así puede estar limitado a menos de 100, quizás menos de 200 o posiblemente menos de 500 ensayos por reacción. Las muestras con grandes cantidades de ADN se pueden dividir en múltiples subreacciones y posteriormente recombinarse antes de la secuenciación. Para las muestras en las que la muestra completa o alguna subpoblación de moléculas de ADN es limitada, la división de la muestra introduciría ruido estadístico. En una realización, una cantidad pequeña o limitada de ADN se puede referir a una cantidad inferior a 10 pg, entre 10 y 100 pg, entre 100 pg y 1 ng, entre 1 y 10 ng, o entre 10 y 100 ng. Cabe señalar que a pesar de que este método resulta particularmente útil con pequeñas cantidades de ADN donde otros métodos que implican la división en múltiples grupos pueden provocar problemas significativos asociados con el ruido estocástico introducido, este método ofrece la ventaja de minimizar el sesgo cuando se ejecuta con muestras de cualquier cantidad de ADN. En estas situaciones, se puede utilizar un paso de preamplificación universal para aumentar la cantidad total de la muestra. Idealmente, este paso de preamplificación no debería alterar de forma apreciable las distribuciones alélicas.

En una realización, un método de la presente invención puede generar productos de PCR que son específicos para un gran número de loci focalizados, concretamente de 1000 a 5000 loci, de 5000 a 10 000 loci o más de 10 000 loci, para la determinación del genotipo mediante secuenciación o algún otro método de determinación del genotipo, de muestras limitadas tales como las que contienen una sola célula o ADN de fluidos corporales. En la actualidad, la realización de reacciones por PCR multiplexada de más de 5 o 10 dianas presenta un desafío importante y a menudo se ve dificultada por los productos derivados de los cebadores, como los dímeros de cebadores y otros artefactos. A la hora de detectar secuencias diana utilizando microarrays con sondas de hibridación, los dímeros de cebadores y otros artefactos se pueden ignorar, dado que estos no se detectan. Sin embargo, cuando se utiliza la secuenciación como método de detección, la gran mayoría de las lecturas de secuenciación secuenciarían estos artefactos y no las secuencias diana deseadas de una muestra. Los métodos descritos en la técnica anterior utilizada para multiplexar más de 50 o 100 reacciones en una reacción, seguida de la secuenciación, resultarán típicamente en más de un 20%, y más a menudo más del 50%, en muchos casos más del 80% y en algunos casos más del 90% de lecturas de secuencia de elementos no diana.

En general, para realizar la secuenciación focalizada de múltiples (n) dianas de una muestra (mayor de 50, mayor de 100, mayor de 500 o mayor de 1000), se puede dividir la muestra en una serie de reacciones paralelas que amplifican una diana individual. Esto se ha realizado en placas multipocillo para PCR o se puede realizar en plataformas comercializadas como FLUIDIGM ACCESS ARRAY (48 reacciones por muestra en chips microfluídicos) o DROPLET PCR mediante RAIN DANCE TECHNOLOGY (entre varios cientos y algunos miles de dianas). Lamentablemente, estos métodos de división y reagrupación son problemáticos para las muestras con una cantidad limitada de ADN, dado que a menudo no existen suficientes copias del genoma para garantizar que exista una copia de cada región del genoma en cada pocillo. Esto representa un problema particularmente grave cuando los loci polimórficos son focalizados y las proporciones relativas de los alelos de los loci polimórficos son importantes, dado que el ruido estocástico introducido por la división y reagrupación causará mediciones muy poco precisas de las proporciones de los alelos presentes en la muestra original de ADN. En el presente documento se describe un método para amplificar de forma efectiva y eficaz muchas reacciones por PCR que resultan aplicables a los casos en los que solo se dispone de una cantidad limitada de ADN. En una realización, el método se puede aplicar para el análisis de células únicas, fluidos corporales, mezclas de ADN como el ADN flotante libre que se encuentra en el plasma materno, biopsias, mezclas ambientales y/o forenses.

En una realización de la invención, la secuenciación focalizada puede implicar uno, una pluralidad o todos los pasos siguientes: a) generar y amplificar una biblioteca con secuencias de adaptadores a ambos lados de los fragmentos de ADN, b) dividir en múltiples reacciones tras la amplificación de la biblioteca; c) generar y opcionalmente amplificar una biblioteca con secuencias de adaptadores a ambos extremos de los fragmentos de ADN; d) realizar una

amplificación de 1000 a 10 000-plex de las dianas seleccionadas utilizando un cebador directo (forward) específico diana para cada diana y un cebador específico de etiqueta; e) realizar una segunda amplificación de este producto utilizando cebadores inversos (reverse) específicos diana y uno (o más) cebadores específicos de una etiqueta universales que se hayan introducido como parte de los cebadores directos específicos diana en la primera ronda; f) realizar una preamplificación de 1000-plex de la diana seleccionada para un número limitado de ciclos; g) dividir el producto en múltiples partes alícuotas y amplificar subconjuntos de dianas en reacciones individuales (por ejemplo,50 a 500-plex, aunque puede ser hasta 1-plex); h) agrupar los productos de las reacciones paralelas de los subconjuntos; i) durante estas amplificaciones, los cebadores pueden portar etiquetas compatibles con la secuenciación (longitud parcial o total) de forma que los productos se puedan secuenciar.

*PCR altamente multiplexada*

En el presente documento se divulgan métodos de conformidad con la invención que permiten la amplificación focalizada de entre más de cien y varias decenas de miles de secuencias diana (por ejemplo, loci SNP) del ADN genómico obtenido del plasma. La muestra amplificada puede estar relativamente libre de productos del dímero del cebador y tener un escaso sesgo alélico en los loci diana. Si durante o después de la amplificación, se unen a los productos adaptadores compatibles con la secuenciación, el análisis de estos productos se puede realizar mediante secuenciación.

La realización de una amplificación por PCR altamente multiplexada utilizando métodos conocidos en la técnica resulta en la generación de productos del dímero del cebador que superan a los productos de amplificación deseados y no resultan adecuados para la secuenciación. Estos se pueden reducir empíricamente eliminando cebadores que forman estos productos o mediante una selección in silico de cebadores. Sin embargo, cuanto mayor es el número de ensayos, más difícil resulta este problema.

Una solución consiste en dividir la reacción de 5000-plex en varias amplificaciones con un multiplexado inferior, por ejemplo, cien reacciones de 50-plex o cincuenta reacciones de 100-plex, o utilizar microfluídicos o incluso dividir la muestra en reacciones por PCR individuales. Sin embargo, si la muestra de ADN es limitada, como en el diagnóstico prenatal no invasivo a partir del plasma de la madre embarazada, se debe evitar dividir la muestra entre múltiples reacciones, dado que esto generará cuellos de botella.

En el presente documento se describen métodos de conformidad con la invención para amplificar primero globalmente el ADN del plasma de una muestra y, a continuación, dividir la muestra en múltiples reacciones con enriquecimiento de la diana multiplexada con unas cifras más moderadas de secuencias diana por reacción. En una realización, se puede utilizar un método de la presente divulgación para enriquecer preferentemente una mezcla de ADN en diversos loci, donde el método se compone de uno o más de los pasos siguientes: generar y amplificar una biblioteca de una mezcla de ADN donde las moléculas de la biblioteca tienen secuencias de adaptadores ligadas a ambos extremos de los fragmentos de ADN, dividir la biblioteca amplificada en múltiples reacciones, realizar una primera ronda de amplificación multiplexada de las dianas seleccionadas utilizando un cebador directo (forward) específico diana por diana y uno o una pluralidad de cebadores inversos (reverse) universales específicos de adaptador. En una realización, un método de la presente divulgación incluye además realizar una segunda amplificación utilizando cebadores específicos diana inversos (reverse) y uno o una pluralidad de cebadores específicos de una etiqueta universal que se haya introducido como parte de los cebadores directos específicos diana en la primera ronda. En una realización, el método puede implicar un método de PCR anidado, hemi-anidado, semi-anidado, anidado unilateral, hemi-anidado unilateral o semi anidado unilateral. En una realización, un método de la presente divulgación se utiliza para enriquecer preferentemente una mezcla de ADN en diversos loci, donde el método consiste en realizar una preamplificación multiplexada de dianas seleccionadas para un número limitado de ciclos, dividir el producto en múltiples partes alícuotas y amplificar subconjuntos de dianas en reacciones individuales, y agrupar los productos en reacciones de subgrupos paralelos. Cabe señalar que este método se podría utilizar para realizar una amplificación focalizada de forma que resulte en bajos niveles de sesgo alélico para 50-500 loci, para 500-5000 loci, para 5000-50 000 loci, o incluso para 50 000-500 000 loci. En una realización, los cebadores portan etiquetas compatibles con la secuenciación parciales o completas.

El flujo de trabajo puede implicar: 1) extraer el ADN plasmático; 2) preparar una biblioteca de fragmentos con adaptadores universales a ambos extremos de los fragmentos; 3) amplificar la biblioteca utilizando cebadores universales específicos para los adaptadores; 4) dividir la "biblioteca" de la muestra amplificada en múltiples partes alícuotas; 5) realizar amplificaciones multiplexadas (por ejemplo, aproximadamente 100-plex, 1000 o 10 000-plex con un cebador específico diana por diana y un cebador específico de etiqueta) con las partes alícuotas; 6) agrupar las partes alícuotas de una muestra; 7) añadir un código de barras a la muestra; 8) mezclar las muestras y ajustar la concentración; 9) secuenciar la muestra. El flujo de trabajo puede comprender múltiples subpasos que contienen uno de los pasos enumerados (por ejemplo, el paso 2) de la preparación de la biblioteca podría implicar tres pasos enzimáticos (generar extremos romos, añadir una cola dA y unión de un adaptador) y tres pasos de purificación). Los pasos del flujo de trabajo se pueden combinar, dividir o realizar en un orden diferente (por ejemplo, añadiendo códigos de barras y agrupando las muestras).

Es importante señalar que la amplificación una biblioteca se puede realizar de forma que mantenga un sesgo para amplificar los fragmentos cortos de forma más eficiente. De esta manera resulta posible amplificar preferentemente las secuencias más cortas, por ejemplo, los fragmentos de ADN mono-nucleosomales, como el ADN fetal libre de células (de origen placentario) que se encuentra en la circulación de las mujeres embarazadas. Cabe señalar que los

ensayos por PCR pueden tener etiquetas, por ejemplo, etiquetas de secuenciación (normalmente una forma truncada de 15-25 bases). Tras el multiplexado, los productos de la PCR de una muestra se agrupan y, a continuación, las etiquetas son completadas (incluyendo la aplicación de un código de barras) mediante una PCR específica de etiqueta (también se podría realizar mediante unión). Por otra parte, se pueden añadir etiquetas de secuenciación completas en la misma reacción que el multiplexado. En los primeros ciclos, las dianas se pueden amplificar con los cebadores específicos diana, posteriormente los cebadores específicos de etiqueta se encargan de completar la secuencia del SQ-adaptador. Los cebadores de la PCR no portan ninguna etiqueta. Las etiquetas de secuenciación se pueden unir a los productos de la amplificación mediante unión.

En una realización de la invención, la PCR altamente multiplexada seguida de la evaluación de material amplificado mediante secuenciación clonal se puede utilizar para detectar la aneuploidía fetal. A pesar de que las PCR multiplexadas tradicionales evalúan hasta 50 loci simultáneamente, el método descrito en el presente documento se puede utilizar para permitir la evaluación simultánea de más de 50 loci, más de 100 loci, más de 500 loci, más de 1000 loci, más de 5 000 loci, más de 10 000 loci, más de 50 000 loci y más de 100 000 loci. Los experimentos han demostrado que se pueden evaluar simultáneamente hasta incluso y más de 10 000 loci distintos en una única reacción, con una eficiencia y especificidad suficientes para realizar diagnósticos prenatales no invasivos de aneuploidía y/o determinaciones del número de copias con una precisión elevada. Los ensayos se pueden combinar en una única reacción con la totalidad de una muestra de cfADN aislada del plasma materno, una fracción de esta, o un derivado procesado de una muestra de cfADN. El cfADN o el derivado también se puede dividir en múltiples reacciones multiplexadas paralelas. La división y el multiplexado óptimos de la muestra se determinan compensando diversas especificaciones de rendimiento. Debido a la cantidad limitada de material, la división de la muestra en múltiples fracciones puede introducir ruido de muestreo, sumar tiempo de manipulación y aumentar la posibilidad de error. Por el contrario, un multiplexado más elevado puede resultar en grandes cantidades de amplificación falsa y mayores desigualdades en la amplificación, dos factores que puede limitar el rendimiento de la prueba.

Dos consideraciones relacionadas cruciales en la aplicación de los métodos descritos en el presente documento son la cantidad limitada de plasma original y el número de moléculas originales en el material del que se obtienen la frecuencia alélica y otras mediciones. Si el número de moléculas originales cae por debajo de un determinado nivel, el ruido de muestreo aleatorio resulta significativo, y puede afectar a la precisión de la prueba.

Típicamente, se pueden obtener datos de calidad suficiente para realizar diagnósticos prenatales no invasivos de aneuploidía si las mediciones se realizan con una muestra que comprende el equivalente a 500-1000 moléculas originales por locus diana. Existen varias formas de aumentar el número de mediciones distintas, por ejemplo, aumentando el volumen de la muestra. Cada manipulación aplicada a la muestra también resulta potencialmente en pérdidas de material. Es esencial caracterizar las pérdidas sufridas por las diversas manipulaciones y evitar pérdidas o cuando sea necesario mejorar el rendimiento de determinadas manipulaciones para evitar pérdidas que puedan deteriorar el rendimiento de la prueba.

En una realización de la invención, resulta posible mitigar las potenciales pérdidas en pasos posteriores, amplificando la totalidad o una fracción de la muestra original de cfADN. Existen diversos métodos disponibles para amplificar la totalidad del material genético de una muestra, aumentando la cantidad disponible para los procedimientos en sentido descendente. En una realización, los fragmentos de ADN de la PCR mediada por unión (LM-PCR) son amplificados mediante PCR tras la unión de un adaptador distinto, dos adaptadores distintos y múltiples adaptadores distintos. En una realización, se utiliza la amplificación por desplazamiento múltiple (MDA) con polimerasa phi-29 para amplificar todo el ADN isotérmicamente. En la DOP-PCR y sus variantes, se utilizan cebadores aleatorios para amplificar el ADN original del material. Cada método presenta determinadas características, como la uniformidad de la amplificación en todas las regiones representadas del genoma, la eficiencia de la captura y amplificación del ADN original, y el rendimiento de la amplificación como una función de la longitud del fragmento.

En una realización de la invención, se puede utilizar LM-PCR con un único adaptador heteroduplexado que tiene una tirosina en el cebador 3'. El adaptador heteroduplexado permite el uso de una única molécula del adaptador que se puede convertir en dos secuencias distintas en los extremos de cebador 5' y 3' del fragmento de ADN original durante la primera ronda de la PCR. En una realización, se puede fraccionar la biblioteca amplificada mediante separaciones por tamaños, o con productos como AMPURE, TASS u otros métodos similares. Antes de la unión, la muestra de ADN se puede someter a un proceso para generar extremos romos y, a continuación, se añade una única base de adenosina al extremo del cebador 3'. Antes de la unión, el ADN se puede clivar utilizando una enzima de restricción o algún otro método de clivaje. Durante la unión, la adenosina del cebador 3' de los fragmentos de la muestra y la tirosina del cebador 3' complementario cubiertos por el adaptador pueden mejorar la eficiencia de la unión. El paso de extensión de la amplificación por PCR puede ser limitado por lo que respecta al tiempo, con el fin de reducir la amplificación de los fragmentos más largos de unos 200 pb, unos 300 pb, unos 400 pb, unos 500 pb o unos 1000 pb. Dado que el ADN más largo que se encuentra en el plasma materno es casi exclusivamente de origen materno, esto puede resultar en un enriquecimiento del ADN fetal del 10-50% y una mejora del rendimiento de la prueba. Se llevaron a cabo varias reacciones utilizando las condiciones especificadas por los kits disponibles en el mercado; el resultado fue la unión satisfactoria en menos del 10% de las moléculas de ADN de la muestra. Una serie de optimizaciones de las condiciones de reacción mejoraron la unión hasta aproximadamente un 70%.

*Mini-PCR*

El diseño del ensayo de PCR tradicional genera pérdidas significativas de distintas moléculas fetales, pero las pérdidas se pueden reducir en gran medida diseñando ensayos de PCR muy cortos, denominados ensayos de mini-PCR.

El cfADN fetal del suero materno se encuentra sumamente fragmentado y los tamaños de los fragmentos están distribuidos aproximadamente de forma gaussiana con una media de 160 pb, una desviación estándar de 15 pb, un tamaño mínimo de unos 100 pb y un tamaño máximo de unos 220 pb.

La distribución de las posiciones de principio y fin del fragmento con respecto a los polimorfismos focalizados, aunque no necesariamente aleatorios, varían en gran medida entre las dianas individuales y entre todas las dianas colectivamente y el punto polimórfico de un locus diana concreto puede ocupar cualquier posición entre el principio y el fin entre los diversos fragmentos que proceden de ese locus.

Cabe señalar que el término mini-PCR se podrá referir igualmente a la PCR normal sin ninguna restricción o limitación adicional.

Durante la PCR, la amplificación solamente se producirá con los fragmentos de ADN de plantilla que comprenden puntos del cebador directo e inverso.

Dado que los fragmentos de cfADN fetal son cortos, la probabilidad de que los dos puntos del cebador estén presentes, la probabilidad de que un fragmento fetal de longitud L comprenda los puntos de los cebadores directo e inverso es un ratio de la longitud del amplicón respecto a la longitud del fragmento.

En condiciones ideales, los ensayos en los que el amplicón tiene 45, 50, 55, 60, 65 o 70 pb se amplificará con éxito de un 72%, 69%, 66%, 63%, 59% o 56%, respectivamente, de las moléculas de los fragmentos de plantilla disponibles.

La longitud del amplicón es la distancia entre los extremos del cebador 5' de los puntos de los cebadores directo e inverso.

La longitud del amplicón que es menor de la típicamente utilizada por los expertos en la técnica puede generar mediciones más eficientes de los loci polimórficos deseados dado que solo requiere lecturas de secuencia cortas.

En una realización, una fracción sustancial de los amplicones debería tener menos de 100 pb, menos de 90 pb, menos de 80 pb, menos de 70 pb, menos de 65 pb, menos de 60 pb, menos de 55 pb, menos de 50 pb, o menos de 45 pb.

Cabe señalar que en los métodos conocidos en la técnica existente, los ensayos cortos como los que se describen en el presente documento suelen evitarse porque no son necesarios e imponen considerables restricciones por lo que respecta al diseño del cebador, al limitar la longitud del cebador, las características de hibridación y la distancia entre el cebador directo e inverso.

Cabe señalar que también existe un potencial de amplificación sesgada si el extremo del cebador 3' de alguno de los cebadores se encuentra aproximadamente a 1-6 bases del punto polimórfico. Esta diferencia de una única base en el punto de la unión de la polimerasa inicial puede resultar en una amplificación preferente de un alelo, lo que puede alterar las frecuencias alélicas observadas y afectar al rendimiento. Todas estas limitaciones hacen que resulte muy difícil identificar cebadores que amplificarán un locus concreto con éxito y también diseñar grandes conjuntos de cebadores que sean compatibles en la misma reacción de multiplexado. En una realización, el extremo 3' de los cebadores internos directo e inverso está diseñado para hibridarse con una región de ADN en sentido ascendente desde el punto polimórfico y separada del punto polimórfico por un pequeño número de bases. Idealmente, el número de bases puede ser de entre 6 y 10 bases, pero también puede ser de entre 4 y 15 bases, entre 3 y 20 bases, entre dos y 30 bases o entre 1 y 60 bases, y conseguir sustancialmente el mismo resultado.

La PCR multiplexada puede implicar una única ronda de PCR en la que todas las dianas sean amplificadas o puede implicar una ronda de PCR seguida por una o más rondas de PCTR anidada o alguna variante de la PCR anidada. La PCR anidada se compone de una ronda o rondas posteriores de amplificación por PCR utilizando uno o más cebadores nuevos que se unen internamente, por al menos un par de bases, a los cebadores utilizados en una ronda anterior. La PCR anidada reduce el número de dianas de amplificación falsas, amplificando, en reacciones posteriores, solo aquellos productos de la amplificación de la anterior que tienen la secuencia interna correcta. Reducir las dianas de amplificación falsas mejora el número de mediciones útiles que se pueden obtener, especialmente en la secuenciación. La PCR anidada típicamente implica el diseño de cebadores completamente internos de los puntos de unión del cebador anterior, aumentando necesariamente el tamaño de segmento de ADN mínimo requerido para la amplificación. Para las muestras como el cfADN de plasma materno, en las que el ADN se encuentra altamente fragmentado, un tamaño de ensayo mayor reduce el número de moléculas distintas de cfADN de las que se puede obtener una medición. En una realización, para compensar este efecto, se puede utilizar un método de anidado parcial donde uno o los dos cebadores de la segunda ronda solapan los primeros puntos de unión extendiéndose internamente algún número de bases para conseguir una especificidad adicional al tiempo que se aumenta mínimamente el tamaño del ensayo total.

En una realización de la invención, un conjunto multiplexado de ensayos de PCR son diseñados para amplificar SNP potencialmente heterocigotos u otros loci polimórficos o no polimórficos en uno o más cromosomas y estos ensayos

se utilizan en una única reacción para amplificar el ADN. El número de ensayos por PCR puede ser entre 50 y 200 ensayos por PCR, entre 200 y 1000 ensayos por PCR, entre 1000 y 5000 ensayos por PCR, o entre 5 000 y 20 000 ensayos por PCR (50 a 200-plex, 200 a 1000-plex, 1000 a 5000-plex, 5000 a 20 000-plex, más de 20 000-plex, respectivamente). En una realización, un conjunto multiplexado de unos 10 000 ensayos por PCR (10 000-plex) está diseñado para amplificar potencialmente loci de SNP heterocigotos en los cromosomas X, Y, 13, 18 y 21 y 1 o 2 y estos ensayos se utilizan en una única reacción para amplificar el cfADN obtenido de una muestra de plasma del material, muestras de vello coriónico, muestras de amniocentesis, una única o un grupo limitado de células, otros fluidos o tejidos corporales, cánceres u otro material genético. Las frecuencias de SNP de cada locus se pueden determinar mediante un método clonal u otro método de secuenciación de los amplicones. El análisis estadístico de los ratios o las distribuciones de la frecuencia alélica de todos los ensayos se puede utilizar para determinar si la muestra contiene una trisomía de uno o más cromosomas incluidos en el ensayo. En otra realización, las muestras originales de cfADN se dividen en dos muestras y se realizan en ensayos de 5000-plex paralelos. En otra realización, las muestras originales de cfADN se divide en muestras y se realizan ensayos de (aprox. 10 000/n-plex) paralelos donde n es entre 2 y 12 o entre 12 y 24 o entre 24 y 48 o entre 48 y 96. Los datos se recopilan y analizan de manera similar a la que ya se ha descrito. Cabe señalar que este método también se puede aplicar a la detección de translocaciones, deleciones, duplicaciones y otras anomalías cromosómicas.

En una realización de la invención, también se pueden añadir colas sin ninguna homología con el genoma diana al extremo 3' o al extremo 5' de cualquiera de los cebadores. Estas colas facilitan las posteriores manipulaciones, procedimientos o mediciones. En una realización, la secuencia de la cola puede ser la misma para los cebadores específicos diana directos e inversos. En una realización, se pueden utilizar diferentes colas para los cebadores específicos diana directos e inversos. En una realización, se puede utilizar una pluralidad de colas diferentes para diferentes loci o conjuntos de loci. Determinadas colas se pueden compartir entre todos los loci o entre subconjuntos de loci. Por ejemplo, el uso de colas directas a inversas correspondientes a las secuencias directas e inversas requeridas por cualquiera de las plataformas de secuenciación existentes puede permitir la secuenciación directa tras la amplificación. En una realización, las colas se pueden utilizar como puntos de cebado comunes entre todas las dianas amplificadas que se pueden usar para añadir otras secuencias útiles. En algunas realizaciones, los cebadores internos pueden contener una región que está diseñada para hibridarse en sentido ascendente o descendente del locus polimórfico diana. En algunas realizaciones, los cebadores pueden contener un código de barras molecular. En algunas realizaciones, el cebador puede contener una secuencia de cebado universal diseñada para permitir la amplificación por PCR.

En una realización de la invención, se crea un conjunto de ensayos por PCR de 10 000-plex de forma que los cebadores directos e inversos tienen colas correspondientes a las secuencias directas e inversas requeridas por un instrumento de secuenciación de alto rendimiento como HISEQ, GAIIX o MYSEQ comercializados por ILLUMINA. Por otra parte, el extremo 5' incluido en las colas de secuenciación es una secuencia adicional que se puede utilizar como punto de cebado en una PCR posterior para añadir secuencias de códigos de barras de nucleótidos a los amplicones, permitiendo la secuenciación multiplexada de múltiples muestras en una única línea del instrumento de secuenciación de alto rendimiento.

En una realización de la invención, se crea un conjunto de ensayos por PCR de 10 000-plex de forma que los cebadores inversos tienen colas correspondientes a las secuencias inversas requeridas por un instrumento de secuenciación de alto rendimiento. Tras la amplificación con el primer ensayo de 10 000-plex, se puede realizar una amplificación posterior por PCR utilizando otro conjunto de 10 000-plex que tiene cebadores directos parcialmente anidados (por ejemplo, anidados de 6 bases) para todas las dianas y un cebador inverso correspondiente a la cola de secuenciación inversa incluida en la primera ronda. Esta ronda posterior de amplificación parcialmente anidada con un solo cebador específico diana y un cebador universal limita el tamaño requerido del ensayo, reduciendo el ruido de muestreo, pero reduce en gran medida el número de amplicones falsos. Las etiquetas de secuenciación se pueden añadir a los adaptadores de unión añadidos y/o como parte de las sondas para PCR, de forma que la etiqueta sea parte del amplicón final.

La fracción fetal afecta al rendimiento de la prueba. Existen varias formas de enriquecer la fracción fetal del ADN que se encuentra en el plasma materno. La fracción fetal se puede incrementar a través del método de LM-PCR anteriormente descrito y ya expuesto, así como mediante la eliminación focalizada de los fragmentos maternos largos. En una realización, antes de la amplificación por PCR multiplexada de los loci diana, se puede realizar una reacción por PCR multiplexada adicional para eliminar selectivamente fragmentos largos y en gran medida maternos correspondientes a los loci focalizados en la posterior PCR multiplexada. Otros cebadores están diseñados para hibridarse con un punto a una distancia mayor del polimorfismo de la que se espera que esté presente entre los fragmentos del ADN fetal libre de células. Estos cebadores se pueden utilizar en una reacción por PCR multiplexada de un ciclo antes de la PCR multiplexada de los loci polimórficos diana. Estos cebadores distales son etiquetados con una molécula o fracción que puede permitir el reconocimiento selectivo de las piezas de ADN etiquetadas. En una realización, estas moléculas de ADN se pueden modificar covalentemente con una molécula de biotina que permite la eliminación del ADN de cadena doble recién formado que comprende estos cebadores tras un ciclo de PCR. El ADN de doble cadena formado durante esa primera ronda es probablemente de origen materno. La eliminación del material híbrido se puede realizar mediante el uso de perlas magnéticas de estreptavidina.

Existen otros métodos de etiquetado que pueden funcionar igual de bien. En una realización, se pueden utilizar métodos de selección del tamaño para enriquecer la muestra en cadenas más cortas de ADN, por ejemplo, las que

tienen menos de unos 800 pb, menos de unos 500 pb o menos de unos 300 pb. La amplificación de los fragmentos cortos se puede realizar después de la manera habitual.

El método de la mini-PCR descrito en la presente divulgación permite la amplificación altamente multiplexada y el análisis de entre cientos y miles o incluso millones de loci en una única reacción con una única muestra. Al mismo tiempo, la detección del ADN amplificado puede ser multiplexada; se pueden multiplexar entre decenas y cientos de muestras en una línea de secuenciación utilizando una PCR con códigos de barras. Esta detección multiplexada ha sido testada con éxito hasta 49-plex y es posible un grado de multiplexado mucho mayor. En efecto, esto permite establecer el genotipo de cientos de muestras en miles de SNP en una única secuenciación. Para estas muestras, el método permite la determinación del genotipo y la tasa de heterocigosidad y la determinación simultánea del número de copias, que se pueden utilizar con objeto de detectar una aneuploidía. Este método resulta particularmente útil en la detección de la aneuploidía de un feto en gestación a partir del ADN flotante libre que se encuentra en el plasma materno. Este método se puede utilizar como parte de un método para determinar el sexo de un feto y/o predecir la paternidad del feto. Esto se puede utilizar como parte de un método para determinar la dosis de mutación. Este método se puede utilizar para cualquier cantidad de ADN o ARN y las regiones focalizadas pueden ser SNP, otras regiones polimórficas, regiones no polimórficas y combinaciones de estas.

En algunas realizaciones de la invención, se puede utilizar la amplificación mediante PCR universal mediada por unión del ADN fragmentado. La amplificación por PCR universal mediada por unión se puede utilizar para amplificar el ADN del plasma, que después se puede dividir en múltiples reacciones paralelas. También se puede utilizar para amplificar preferentemente fragmentos cortos, enriqueciendo así la fracción fetal. En algunas realizaciones, la adición de etiquetas a los fragmentos mediante unión puede permitir la detección de fragmentos más cortos, el uso de porciones de los cebadores específicos de secuencia diana más cortos y/o la hibridación a temperaturas más elevadas que reduce las reacciones no específicas.

Los métodos descritos en el presente documento se pueden utilizar para una serie de fines donde haya un conjunto diana de ADN que está mezclado con una cantidad de ADN contaminante. En algunas realizaciones, el ADN diana y el ADN contaminante puede proceder de individuos genéticamente relacionados. Por ejemplo, las anomalías genéticas de un feto (diana) se pueden detectar a partir del plasma materno que contiene ADN fetal (diana) y también ADN materno (contaminante); las anomalías incluyen anomalías del cromosoma completo (por ejemplo, aneuploidía), anomalías cromosómicas parciales (por ejemplo, deleciones, duplicaciones, inversiones, translocaciones), polimorfismos de polinucleótidos (por ejemplo, STR), polimorfismos de un único nucleótido y/u otras diferencias o anomalías genéticas. En algunas realizaciones, el ADN diana y contaminante puede proceder del mismo individuo, pero donde el ADN diana y contaminante se diferencian en una o más mutaciones, por ejemplo, en el caso del cáncer (ver, por ejemplo, Mamon et al. Preferential Amplification of Apoptotic DNA from Plasma: Potential for Enhancing Detection of Minor DNA Alterations in Circulating DNA. Clinical Chemistry 54:9 (2008). En algunas realizaciones, el ADN se puede encontrar en el supernatante de un cultivo celular (apóptico). En algunas realizaciones, se puede inducir la apóptosis en muestras biológicas (por ejemplo, sangre) para la posterior preparación de la biblioteca, amplificación y/o secuenciación. Una serie de flujos de trabajo y protocolos que permiten conseguir este fin se presentan en otro apartado de la presente divulgación.

En algunas realizaciones, el ADN diana puede proceder de células únicas, de muestras de ADN que contienen menos de una copia del genoma diana, de pequeñas cantidades de ADN, de ADN de origen mixto (por ejemplo, plasma de una mujer embarazada: ADN placentario y materno; plasma de pacientes con cáncer y tumores: mezcla entre ADN sano y cancerígeno, trasplante, etc.), de otros fluidos corporales, de cultivos celulares, de supernatantes de cultivos, de muestras forenses de ADN, de muestras antiguas de ADN (por ejemplo, insectos fijados en ámbar), de otras muestras de ADN y combinaciones de estos.

En algunas realizaciones, se puede utilizar un tamaño de amplicón corto. Los tamaños de amplicones cortos resultan especialmente adecuados para el ADN fragmentado (ver, por ejemplo, A. Sikora, et al. *Detection of increased amounts of cell-free fetal DNA with short PCR amplicons*. Clin Chem. 2010 Jan;56(l):136-8.) El uso de tamaños de amplicones cortos puede proporcionar algunas ventajas significativas. Los tamaños de amplicones cortos pueden proporcionar una eficiencia de amplificación optimizada. Los tamaños de amplicones cortos típicamente producen productos más cortos y, por tanto, existen menos probabilidades de cebado no específico. Los productos más cortos se pueden agrupar de forma más densa en la célula del flujo de secuenciación, dado que los grupos serán más pequeños. Cabe señalar que los métodos descritos en el presente documento pueden funcionar igualmente bien para los amplicones de la PCR más largos. La longitud del amplicón se puede incrementar si es necesario, por ejemplo, cuando se secuencian tramos de secuencias más largas.

Se realizaron experimentos de amplificación focalizada de 146-plex con ensayos de 100 pb a 200 pb de largo como primer paso en un protocolo de PCR anidada con células únicas y con ADN genómico con resultados positivos.

En algunas realizaciones, los métodos descritos en el presente documento se pueden utilizar para amplificar y/o detectar SNP, el número de copias, la metilación de nucleótidos, niveles de ARN mensajero (mARN), otros tipos de niveles de expresión de ARN, y otras características genéticas y epigenéticas. Los métodos de mini-PCR descritos en el presente documento se pueden utilizar con la secuenciación de nueva generación; se pueden utilizar con otros métodos en sentido descendente como microarrays, recuento mediante PCR digital, PCR en tiempo real, análisis de espectrometría de masas, etc.

En algunas realizaciones, los métodos de amplificación por mini-PCR descritos en el presente documento se pueden utilizar como parte de un método para la cuantificación precisa de poblaciones minoritarias. Se puede utilizar para la cuantificación absoluta utilizando calibradores de picos. Se puede utilizar para la cuantificación de la mutación/alelos menores a través de una secuenciación muy profunda y se puede ejecutar de forma altamente multiplexada. Se puede utilizar para las pruebas estándar de paternidad e identidad de parientes y antepasados en seres humanos, animales, plantas u otras criaturas. Se puede utilizar para pruebas forenses. Se puede utilizar para la determinación rápida del genotipo y para el análisis del número de copias (CN), en cualquier tipo de material, por ejemplo, líquido amniótico y CVS, esperma, productos de la concepción (POC). Se puede utilizar para análisis de células únicas, como la determinación del genotipo en muestras obtenidas de biopsias de embriones. Se puede utilizar para un análisis rápido del embrión (a menos de un día, a uno o dos días de la biopsia) mediante secuenciación focalizada utilizando mini-PCR.

En algunas realizaciones, se puede utilizar para el análisis de tumores: las biopsias de tumores suelen ser una mezcla de células sanas y células tumorales. La PCR focalizada permite una secuenciación profunda de SNP y loci que prácticamente carecen de secuencias de fondo. Se puede utilizar para el análisis del número de copias y la pérdida de heterocigosidad en ADN tumoral. Dicho ADN tumoral puede estar presente en múltiples fluidos corporales diferentes o tejidos de pacientes con tumores. Se puede utilizar para la detección de la recurrencia tumoral y/o la detección de un tumor. Se puede utilizar para las pruebas del control de calidad de las semillas. Se puede utilizar para la cría o la pesca. Cabe señalar que algunos de estos métodos se podrían utilizar perfectamente para la focalización de loci no polimórficos para la determinación del estado de ploidía.

Alguna de la bibliografía que describe algunos de los métodos fundamentales subyacentes de los métodos divulgados en el presente documento incluyen: 1) Wang HY, Luo M, Tereshchenko IV, Frikker DM, Cui X, Li JY, Hu G, Chu Y, Azaro MA, Lin Y, Shen L, Yang Q, Kambouris ME, Gao R, Shih W, Li H. Genome Res. 2005 Feb;15(2):276-83. Department of Molecular Genetics, Microbiology and Immunology/The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, New Jersey 08903, EE. UU. 2) High-throughput genotyping of single nucleotide polymorphisms with high sensitivity. Li H, Wang HY, Cui X, Luo M, Hu G, Greenawalt DM, Tereshchenko IV, Li JY, Chu Y, Gao R. Methods Mol Biol. 2007;396 - PubMed PMID: 18025699. 3) Un método que comprende el multiplexado de una media de nueve ensayos para la secuenciación se describe en: Nested Patch PCR enables highly multiplexed mutation discovery in candidate 5 genes. Varley KE, Mitra RD. Genome Res. 2008 Nov;18(ll):1844-50. Epub 2008 Oct 10. Cabe señalar que los métodos divulgados en el presente documento permiten el multiplexado de órdenes de magnitud superiores a los indicados en las referencias anteriores.

*Diseño de cebadores*

A menudo la PCR altamente multiplexada puede resultar en la producción de una proporción muy elevada de ADN como consecuencia de reacciones secundarias no productivas como la formación de dímeros de cebadores. En una realización, los cebadores concretos que es más probable que causen reacciones secundarias no productivas se pueden eliminar de la biblioteca de cebadores para obtener una biblioteca de cebadores que proporcionará una proporción mayor de ADN amplificado correspondiente al genoma. El paso de eliminar cebadores problemáticos, es decir aquellos cebadores que es particularmente probable que formen dímeros ha proporcionado sorprendentemente unos niveles de multiplexado por PCR muy elevados para el posterior análisis mediante secuenciación. En sistemas como la secuenciación, donde el rendimiento se ve reducido de forma significativa por los dímeros de cebadores y/u otros productos no deseados, se ha conseguido un multiplexado más de 10, más de 50 y más de 100 veces mayor que en el caso de otros multiplexados descritos. Cabe señalar que esto contrasta con los métodos de detección basados en sondas, por ejemplo, microarrays, TaqMan, PCR, etc., donde un exceso de dímeros de cebadores no afectará al resultado de forma apreciable. Cabe señalar también que la idea general de la técnica es que la PCR por multiplexado para la secuenciación se limita a unos 100 ensayos en el mismo pocillo. Por ejemplo, Fluidigm y Rain Dance ofrecen plataformas para realizar 48 o 1000 ensayos por PCR en reacciones paralelas para una muestra.

Existen diversas formas de seleccionar los cebadores para una biblioteca en la que se minimiza la cantidad de dímeros de cebadores no correspondientes u otros productos no deseados de los cebadores. Los datos empíricos indican que un pequeño número de cebadores "malos" son responsables de una gran cantidad de reacciones secundarias de dímeros de cebadores no correspondientes. La eliminación de estos cebadores "malos" puede aumentar el porcentaje de lecturas de secuencia que se corresponden con los loci focalizados. Una forma de identificar los cebadores "malos" consiste en analizar los datos de secuencia del ADN que se ha amplificado mediante amplificación focalizada; los dímeros de cebadores que se observan con mayor frecuencia se pueden eliminar para obtener una biblioteca de cebadores que es menos probable que produzcan ADN de productos secundarios que no se corresponda con el genoma. También hay programas a disposición del público que pueden calcular la energía de unión de diversas combinaciones de cebadores y la eliminación de que los presentan la energía de unión más elevada también proporcionará una biblioteca de cebadores que es menos probable que produzcan ADN de productos secundarios que no se corresponda con el genoma.

El multiplexado de grandes cantidades de cebadores impone considerables limitaciones sobre los análisis que se pueden incluir. Los ensayos que interactúan de forma no deseada provocan productos de la amplificación falsos. Las limitaciones de tamaño de la mini-PCR pueden provocar otras limitaciones. En una realización, se puede comenzar con una cantidad muy importante de potenciales dianas de SNP (entre unas 500 y más de un millón) e intentar diseñar cebadores para amplificar cada SNP. Si los cebadores pueden ser diseñados, se puede intentar identificar

pares de cebadores que es probable que formen productos falsos evaluando la probabilidad de formación dúplex de cebadores falsos entre todos los pares de cebadores posibles utilizando parámetros termodinámicos publicados para la formación de dúplex de ADN. Las interacciones de cebadores se pueden clasificar mediante una función de puntuación relacionada con la interacción y los cebadores con las peores puntuaciones de interacción son eliminados hasta que se alcance el número de cebadores deseado. En los casos en los que los SNP que es probable que sean heterocigotos los más útiles, también se puede clasificar la lista de ensayos y seleccionar los ensayos más compatibles con los heterocigotos. Los experimentos han confirmado que los cebadores con las puntuaciones de interacción elevadas tienen más probabilidades de formar dímeros de cebadores. Con un multiplexado elevado no resulta posible eliminar todas las interacciones falsas, pero resulta fundamental eliminar los cebadores o pares de cebadores con las puntuaciones de interacción más elevadas in silico, dado que pueden dominar una reacción completa, limitando en gran medida la amplificación de las dianas previstas. Hemos realizado este procedimiento para crear conjuntos de cebadores de multiplexado de hasta 10 000 cebadores. La mejora obtenida con este procedimiento es sustancial, dado que permite la amplificación de más del 80%, más del 90%, más del 95%, más del 98% e incluso más del 99% en productos diana determinados mediante secuenciación de todos los productos de la PCR, en comparación con el 10% de una reacción en la que no se han eliminado los peores cebadores. Cuando se combina con un método semi-anidado parcial como el anteriormente descrito, más del 90% e incluso más del 95% de los amplicones se pueden corresponder con las secuencias focalizadas.

Cabe señalar que existen otros métodos para determinar las sondas para PCR que es probable que formen dímeros. En una realización, el análisis de un conjunto de ADN que se ha amplificado utilizando un grupo no optimizado de cebadores puede ser suficiente para determinar cebadores problemáticos. Por ejemplo, el análisis se puede realizar utilizando secuenciación y se determina que aquellos dímeros que se encuentran presentes en mayor número es más probable que formen dímeros y pueden ser eliminados.

Este método tiene una serie de aplicaciones potenciales, por ejemplo, la determinación del genotipo del SNP, la determinación de la tasa de heterocigosidad, la medición del número de copias y otras aplicaciones de secuenciación focalizada. En una realización, el método del diseño del cebador se puede utilizar en combinación con el método de la mini-PCR descrito en otro apartado de este documento. En algunas realizaciones, el método del diseño del cebador se puede utilizar como parte de un método de PCR multiplexada masiva.

El uso de etiquetas en los cebadores puede reducir la amplificación y secuenciación de dímeros del cebador. Los cebadores con etiquetas se pueden utilizar para acortar la secuencia específica de diana requerida hasta menos de 20, menos de 15, menos de 12 e incluso menos de 10 pares de bases. Esto se puede producir de forma fortuita con el diseño del cebador estándar cuando la secuencia diana está fragmentada dentro del punto de unión del cebador o se puede diseñar durante el diseño del cebador. Las ventajas de este método incluyen: aumenta el número de ensayos que se pueden diseñar para una determinada longitud máxima del amplicón, y reduce la secuenciación "no informativa" de la secuencia del cebador. Se puede utilizar en combinación con un etiquetado interno (ver en otro apartado de este documento).

En una realización, la cantidad relativa de productos no productivos en una amplificación por PCR focalizada multiplexada se puede reducir aumentando la temperatura de hibridación. En los casos en los que se amplifican bibliotecas con la misma etiqueta que los cebadores de diana específica, la temperatura de hibridación se puede aumentar en comparación con el ADN genómico, dado que las etiquetas contribuyen a la unión del cebador. En algunas realizaciones de la invención, utilizamos unas concentraciones de cebador considerablemente menores de las que se habían documentado con anterioridad y también utilizamos unos tiempos de hibridación superiores a los documentados. En algunas realizaciones, los tiempos de hibridación puede ser superiores en más de 10 minutos, más de 20 minutos, más de 30 minutos, más de 60 minutos, más de 120 minutos, más de 240 minutos, más de 480 minutos e incluso más de 960 minutos. En una realización, se utilizan tiempos de hibridación superiores que en informes anteriores, lo que permite reducir las concentraciones del cebador. En algunas realizaciones, las concentraciones del cebador son de tan solo 50 nM, 20 nM, 10 nM, 5 nM, 1 nM e inferiores a 1uM. Sorprendentemente esto da como resultado un excelente rendimiento de las reacciones altamente multiplexadas, por ejemplo, reacciones de 1000-plex, reacciones de 2000-plex, reacciones de 5000-plex, reacciones de 10 000-plex, reacciones de 20 000-plex, reacciones de 50 000-plex e incluso reacciones de 100 000-plex. En una realización, la amplificación utiliza uno, dos, tres, cuatro o cinco ciclos ejecutados con tiempos de hibridación prolongados, seguidos por ciclos de PCR con tiempos de hibridación más habituales con los cebadores etiquetados.

Para seleccionar ubicaciones diana, se puede comenzar con una serie de diseños de pares de cebadores candidatos y crear un modelo termodinámico de interacciones potencialmente adversas entre los pares de cebadores y, a continuación, utilizar el modelo para eliminar los diseños que son incompatibles con otros diseños de la serie.

*Variantes de PCR focalizada - Anidado*

Existen múltiples flujos de trabajo posibles cuando se realiza una PCR; a continuación, se describen algunos flujos de trabajo típicos para los métodos divulgados. Los pasos descritos en el presente documento no pretenden excluir otros posibles pasos ni implican que ninguno de los pasos descritos en el presente documento sea necesario para que el método funcione correctamente. En la bibliografía se documenta un gran número de variaciones de los parámetros y otras modificaciones que se pueden realizar sin afectar a la esencia de la invención. Un flujo de trabajo concreto generalizado se expone a continuación seguido de una serie de posibles variantes. Las variantes se

refieren típicamente a posibles reacciones de PCR secundarias, por ejemplo, diferentes tipos de animado que se pueden realizar (paso 3). Es importante señalar que se pueden realizar variantes en todo momento en un orden diferente del que se describe de forma explícita en el presente documento.

1.El ADN de la muestra puede tener adaptadores de unión, a menudo denominados etiquetas de bibliotecas o etiquetas de adaptadores de unión (LT), unidos, donde los adaptadores de unión contienen una secuencia de cebado universal seguida de una amplificación universal. En una realización, esto se puede realizar utilizando un protocolo estándar diseñado para crear bibliotecas de secuenciación tras la fragmentación. En una realización, la muestra de ADN se puede someter a un proceso para generar extremos romos y, a continuación, se puede añadir una A al extremo 3'. Se pueden añadir y unir un adaptador Y con una proyección T. En algunas reacciones, se pueden utilizar extremos adhesivos distintos de una proyección A o T. En algunas realizaciones, se pueden añadir otros adaptadores, por ejemplo, adaptadores de unión en bucle. En algunas realizaciones, los adaptadores pueden tener una etiqueta diseñada para la amplificación por PCR.

2.Amplificación de diana específica (STA): La preamplificación de cientos a miles, a de decenas de miles e incluso a cientos de miles de dianas puede ser multiplexada en una reacción. La STA típicamente se ejecuta en 10-30 ciclos, aunque se puede ejecutar en 5-40 ciclos, 2-50 ciclos e incluso 1-100 ciclos. Los cebadores pueden tener cola, por ejemplo, para facilitar el flujo de trabajo o evitar la secuenciación de una gran proporción de dímeros. Cabe señalar, que típicamente los dímeros de los dos cebadores que portan la misma etiqueta no se amplificarán ni secuenciarán de forma eficiente. En algunas realizaciones, se pueden realizar entre 1 y 10 ciclos de PCR; en algunas realizaciones se pueden realizar entre 10 y 20 ciclos de PCR; en algunas realizaciones se pueden realizar entre 20 y 30 ciclos de PCR; en algunas realizaciones se pueden realizar entre 30 y 40 ciclos de PCR; en algunas realizaciones se pueden realizar más de 40 ciclos de PCR. La amplificación puede ser una amplificación lineal. El número de ciclos de PCR puede ser optimizado para obtener un perfil de profundidad de lectura (DOR) óptimo. Diferentes perfiles de DOR pueden ser recomendables para diferentes fines. En algunas realizaciones, una distribución más uniforme de las lecturas entre todos los ensayos resulta recomendable; si la DOR es demasiado limitada para algunos ensayos, el ruido estocástico puede ser demasiado elevado para que los datos resulten útiles, mientras que si la profundidad de lectura es demasiado elevada, la utilidad marginal de cada lectura adicional es relativamente pequeña.

La adición de colas a los cebadores mejora la detección del ADN fragmentado de bibliotecas universalmente etiquetadas. Si la etiqueta de la biblioteca y las colas del cebador contienen una secuencia homóloga, la hibridación se puede mejorar (por ejemplo, reduciendo la temperatura de fusión (Tm) y los cebadores se pueden ampliar únicamente si una porción de la secuencia diana del cebador se encuentra en el fragmento de ADN de la muestra. En algunas realizaciones, se pueden utilizar 13 o más pares de bases específicos diana. En algunas realizaciones, se pueden utilizar 10-12 pares de bases específicos diana. En algunas realizaciones, se pueden utilizar 8-9 pares de bases específicos diana. En algunas realizaciones, se pueden utilizar 6-7 pares de bases específicos diana. En algunas realizaciones, se puede realizar la STA con el ADN preamplificado, por ejemplo, MDA, RCA, otras amplificaciones del genoma completo o PCR universal mediada por adaptador. En algunas realizaciones, se puede realizar la STA con muestras que están enriquecidas o en las que se han reducido determinadas secuencias y poblaciones, por ejemplo, selección por tamaño, captura de diana, degradación dirigida.

3.En algunas realizaciones se pueden realizar PCR multiplexadas secundarias o reacciones de extensión del cebador para aumentar la especificidad y reducir los productos no deseables. Por ejemplo, el anidado, semi-anidado, hemi-anidado y/o subdivisión en reacciones paralelas de series de ensayo de menor tamaño son técnicas que se pueden utilizar para aumentar la especificidad. Los experimentos han demostrado que dividir una muestra en tres reacciones de 400-plex resultó en un ADN con una mayor especificidad que una reacción de 1200-plex con exactamente los mismos cebadores. Del mismo modo, los experimentos han demostrado que dividir una muestra en cuatro reacciones de 2400-plex resultó en un ADN con una mayor especificidad que una reacción de 9600-plex con exactamente los mismos cebadores. En una realización, se pueden utilizar cebadores específicos diana y específicos de etiqueta de la misma direccionalidad y de direccionalidad opuesta.

4.En algunas realizaciones se puede amplificar una muestra de ADN (dilución, purificación u otro) producida por una reacción de STA utilizando cebadores específicos diana y "amplificación universal", es decir, amplificar muchas o todas de las dianas etiquetadas y preamplificadas. Los cebadores pueden contener secuencias funcionales adicionales, por ejemplo, códigos de barras o una secuencia de adaptador completa necesaria para la secuenciación en una plataforma de secuenciación de alto rendimiento.

Estos métodos se pueden utilizar para el análisis de cualquier muestra de ADN y resultan especialmente útiles cuando la muestra de ADN es particularmente pequeña o cuando se trata de una muestra de ADN en la que el ADN procede de más de un individuo, como en el caso del plasma materno. Estos métodos se pueden utilizar con muestras de ADN como las que contienen una única célula o un reducido número de células, ADN genómico, ADN plasmático, bibliotecas de plasma amplificadas, bibliotecas de supernatante apoptótico amplificado u otras muestras de ADN mixto. En una realización, estos métodos se pueden utilizar en caso de que pueda haber presentes células de diferente constitución genética en un único individuo, como individuos con cáncer o trasplantes.

*Variantes del protocolo (variantes y/o adiciones al flujo de trabajo anterior)*

*Mini-PCR multiplexada directa*: La amplificación de diana específica (STA) de una pluralidad de secuencias diana con cebadores etiquetados se muestra en la Figura 1. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con cebadores para PCR hibridados. 104 denota el producto final de la PCR. En algunas realizaciones, la STA se puede realizar en más de 100, más de 200, más de 500, más de 1000, más de 2000, más de 5000, más de 10 000, más de 20 000, más de 50 000, más de 100 000 o más de 200 000 dianas. En una reacción posterior, los cebadores con etiqueta específicos amplifican todas las secuencias diana y alargan las etiquetas para incluir todas las secuencias necesarias para la secuenciación, incluyendo los índices de las muestras. En una realización, los cebadores pueden no estar etiquetados o puede que solo determinados cebadores estén etiquetados. Se pueden añadir adaptadores de secuenciación mediante unión convencional del adaptador. En una realización, los cebadores iniciales pueden portar las etiquetas.

En una realización, los cebadores están diseñados para que la longitud del ADN amplificado sea inesperadamente corta. La técnica existente demuestra que los expertos con conocimientos ordinarios de la técnica típicamente diseñan amplicones de más de 100 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 80 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 70 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 60 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 50 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 45 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 40 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan menos de 35 pb. En una realización, los amplicones pueden estar diseñados de forma que tengan entre 40 y 65 pb.

Se realizó un experimento utilizando este protocolo con la amplificación de 1200-plex. Se utilizó tanto ADN genómico como plasma del embarazo; aproximadamente el 70% de las lecturas de secuencia se correspondían a las secuencias focalizadas. Los detalles se recogen en otro apartado del presente documento. La secuenciación de 1042-plex sin diseño y selección de ensayos provocó que más del 99% de las secuencias fuesen productos de dímeros del cebador.

*PCR secuencial*: Tras STA1 se pueden amplificar múltiples partes alícuotas del producto en paralelo con grupos de complejidad reducida con los mismos cebadores. La primera amplificación puede dar material suficiente para la división. El método resulta especialmente apropiado para pequeñas muestras, por ejemplo, las de aproximadamente 6-100 pg, aproximadamente 100 pg a 1 ng, aproximadamente 1 ng a 10 ng, o aproximadamente 10 ng a 100 ng. El protocolo se realizó con 1200-plex en tres 400-plexes. La correspondencia de las lecturas de secuencia aumentó de aproximadamente un 60-70% en 1200-plex hasta superar el 95%.

*Mini-PCR semi-anidada*: (ver la Figura 2) Tras la STA1, se realiza una segunda STA que consiste en un conjunto multiplexado de cebadores directos anidados internos (103B, 105b) y uno o unos cuantos cebadores inversos específicos de etiqueta (103A). 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con el cebador directo B y el cebador inverso A hibridados. 104 denota el producto de la PCR de 103. 105 denota el producto de 104 con el cebador directo anidado b hibridado, y la etiqueta inversa A formando ya parte de la molécula de la PCR que ha ocurrido entre 103 y 104. 106 denota el producto final de la PCR. Con este flujo de trabajo, normalmente más del 95% del mapa de secuencias se corresponden con dianas no previstas. El cebador anidado puede solaparse con la secuencia externa del cebador directo pero introduce tres bases adicionales en el extremo 3'. En algunas realizaciones, se pueden utilizar entre una y 20 bases adicionales en el extremo 3'. Los experimentos han demostrado que el uso de 9 o más bases adicionales en el extremo 3' con diseños de 1200-plex funciona bien.

*Mini-PCR completamente anidada*: (ver la Figura 3) Tras el paso 1 de la STA, se puede realizar una segunda PCR multiplexada (o PCR multiplexadas paralelas de complejidad reducida) con dos cebadores anidados que portan etiquetas (A, a, B, b). 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con el cebador directo B y el cebador inverso A hibridados. 104 denota el producto de la PCR de 103. 105 denota el producto de 104 con el cebador directo anidado b y el cebador inverso anidado a hibridados. 106 denota el producto final de la PCR. En algunas realizaciones se pueden utilizar dos conjuntos de cebadores completos. Los experimentos utilizando un protocolo completo de mini-PCR anidada se emplearon para realizar una amplificación de 146-plex con una única célula y tres células sin el paso 102 de añadir adaptadores de unión universal y amplificación.

*Mini-PCR hemi-anidada*: (ver la Figura 4) Se puede utilizar ADN diana que tiene adaptadores en los extremos de los fragmentos. La STA realizada comprende un conjunto multiplexado de cebadores directos (B) y uno o unos cuantos cebadores inversos específicos de etiqueta (A). Se puede utilizar una segunda STA que emplea un cebador directo específico de etiqueta universal y un cebador inverso específico diana. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con el cebador inverso A hibridado. 104 denota el producto de la PCR de 103 que ha sido amplificado utilizando el cebador inverso

A y el cebador de la etiqueta del adaptador de unión LT. 105 denota el producto de 104 con el cebador directo B hibridado. 106 denota el producto final de la PCR. En este flujo de trabajo, los cebadores directos e inversos específicos diana se utilizan en reacciones separadas, reduciendo así la complejidad de la reacción y evitando la formación de dímeros en los cebadores directos e inversos. Cabe señalar que en este ejemplo los cebadores A y B se pueden considerar los primeros cebadores y los cebadores "a" y "b" se pueden considerar los cebadores internos. Este método representa una gran mejora con respecto a la PCR directa, porque es igual de válido que esta pero evita los dímeros de los cebadores. Tras una primera ronda del protocolo hemi-anidado, por lo general se observa aproximadamente un 99% de ADN no focalizado; sin embargo, tras una segunda ronda se observa por lo general una gran mejora.

*Mini-PCR hemi-anidada por triplicado*: (ver la Figura 5) Se puede utilizar ADN diana que tiene un adaptador en los extremos de los fragmentos. La STA realizada comprende un conjunto multiplexado de cebadores directos (B) y uno o unos cuantos cebadores inversos específicos de etiqueta (A) y (a). Se puede utilizar una segunda STA que emplea un cebador directo específico de etiqueta universal y un cebador inverso específico diana. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con el cebador inverso A hibridado. 104 denota el producto de la PCR de 103 que ha sido amplificado utilizando el cebador inverso A y el cebador de la etiqueta del adaptador de unión LT. 105 denota el producto de 104 con el cebador directo B hibridado. 106 denota el producto de la PCR de 105 que se ha amplificado utilizando el cebador inverso A y el cebador directo B. 107 denota el producto de 106 con un cebador inverso "a" hibridado. 108 denota el producto final de la PCR. Cabe señalar que en este ejemplo los cebadores "a" y B se pueden considerar los cebadores internos y A se puede considerar un primer cebador. Opcionalmente, tanto A como B se pueden considerar los primeros cebadores y "a" se puede considerar un cebador interno. La designación de cebadores inversos y directos se puede intercambiar. En este flujo de trabajo, los cebadores directos e inversos específicos diana se utilizan en reacciones separadas, reduciendo así la complejidad de la reacción y evitando la formación de dímeros en los cebadores directos e inversos. Este método representa una importante mejora con respecto a la PCR directa, porque es igual de válido que esta pero evita los dímeros de los cebadores.

Tras una primera ronda del protocolo hemi-anidado, por lo general se observa aproximadamente un 99% de ADN no focalizado; sin embargo, tras una segunda ronda se observa por lo general una importante mejora.

*Mini-PCR anidada unilateral*: (ver la Figura 6) Se puede utilizar ADN diana que tiene un adaptador en los extremos de los fragmentos. La STA también se puede realizar con un conjunto multiplexado de cebadores directos anidados y utilizando la etiqueta del adaptador de unión como cebador inverso. A continuación, se puede realizar una segunda STA que emplea un conjunto de cebadores directos anidados y un cebador inverso universal. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple que ha sido universalmente amplificado con el cebador directo A hibridado. 104 denota el producto de la PCR de 103 que ha sido amplificado utilizando el cebador directo A y el cebador inverso de la etiqueta del adaptador de unión LT. 105 denota el producto de 104 con el cebador directo anidado a hibridado. 106 denota el producto final de la PCR. Este método puede detectar secuencias diana más cortas que la PCR estándar utilizando cebadores que se solapan en las primeras y las segundas STA. El método se realiza típicamente con una muestra de ADN que ya se ha sometido al paso 1 de la STA anterior, es decir la adición de etiquetas universales y amplificación; los dos cebadores anidados se encuentran únicamente a un lado y el otro lado utiliza la etiqueta de la biblioteca. El método se realizó con bibliotecas de supernatantes apoptópticos y plasma del embarazo. Con este flujo de trabajo, aproximadamente el 60% de las secuencias se correspondían con las dianas previstas. Cabe señalar que las lecturas que contenían la secuencia del adaptador inverso no se correspondían, por lo que se espera que esta cifra sea mayor si las lecturas que contienen la secuencia del adaptador inverso se corresponden.

*Mini-PCR unilateral:* Se puede utilizar ADN diana que tiene un adaptador en los extremos de los fragmentos (ver la Figura 7). Se puede realizar la STA con un conjunto multiplexado de cebadores directos y uno o unos cuantos cebadores inversos específicos de etiqueta. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple con el cebador directo A hibridado. 104 denota el producto de la PCR de 103 que ha sido amplificado utilizando el cebador directo A y el cebador inverso de la etiqueta del adaptador de unión LT, y que es el producto final de la PCR. Este método puede detectar secuencias diana más cortas que la PCR estándar. Sin embargo, puede resultar relativamente inespecífico, dado que solo se utiliza un cebador específico diana. Este protocolo es efectivamente la mitad de una mini-PCR anidada unilateral.

*Mini-PCR semi-anidada inversa:* Se puede utilizar ADN diana que tiene un adaptador en los extremos de los fragmentos (ver la Figura 8). Se puede realizar la STA con un conjunto multiplexado de cebadores directos y uno o unos cuantos cebadores inversos específicos de etiqueta. 101 denota ADN de doble cadena con un locus polimórfico de interés en X. 102 denota el ADN de doble cadena con adaptadores de unión añadidos para la amplificación universal. 103 denota el ADN de cadena simple con el cebador inverso B hibridado. 104 denota el producto de la PCR de 103 que ha sido amplificado utilizando el cebador inverso B y el cebador directo de la etiqueta del adaptador de unión LT. 105 denota el producto de la PCR 104 con el cebador directo hibridado A y el cebador inverso interno "b". 106 denota el producto de la PCR que ha sido amplificado a partir de 105 utilizando el

cebador directo A y el cebador inverso "b" y que es el producto final de la PCR. Este método puede detectar secuencias diana más cortas que la PCR estándar.

También existen más variantes que son simplemente iteraciones o combinaciones de los métodos anteriores, como la PCR doblemente anidada, donde se utilizan tres conjuntos de cebadores. Otra variante es la mini-PCR anidada sobre un lado y medio de la región a amplificar (*one-and-a-half-sided*), donde la STA también se puede realizar con un conjunto multiplexado de cebadores directos anidados y uno o unos cuantos cebadores inversos específicos diana.

Cabe señalar que, en todas estas variantes, la identidad del cebador directo y del cebador inverso no se puede intercambiar. Cabe señalar que, en algunas realizaciones, la variante anidada también se puede realizar perfectamente sin la preparación de la biblioteca inicial que comprende la adición de etiquetas de los adaptadores y un paso de amplificación universal. Cabe señalar que, en algunas de estas realizaciones, se pueden incluir rondas adicionales de PCR, con cebadores directos y/o inversos y pasos de amplificación adicionales; estos pasos adicionales pueden resultar particularmente útiles si resulta recomendable aumentar todavía más el porcentaje de moléculas de ADN que corresponden a los loci focalizados.

*Flujos de trabajo de anidado*

Existen muchas maneras de realizar la amplificación, con diferentes grados de anidado y con diferentes grados de multiplexado. En la Figura 9 se incluye un diagrama de flujo con algunos de los flujos de trabajo posibles. Cabe señalar que el uso de la PCR de 10 000-plex solo se ofrece a modo de ejemplo; estos diagramas de flujo funcionarían perfectamente también con otros grados de multiplexado.

*Adaptadores de unión en bucle*

Cuando se añaden adaptadores etiquetados universales, por ejemplo, al objeto de elaborar una biblioteca para la secuenciación, existen diversas maneras de ligar los adaptadores. Una consiste en someter el ADN a un proceso para generar extremos romos, añadir una cola A y ligar los adaptadores que tienen una proyección T. Hay varios otros métodos para ligar los adaptadores También hay varios adaptadores que se pueden ligar. Por ejemplo, se puede utilizar un adaptador Y donde el adaptador se compone de dos cadenas de ADN donde una cadena tiene una región de doble cadena, y una región especificada por una región de un cebador directo, y donde la otra cadena especificada por una región de doble cadena es complementaria a la región de doble cadena de la primera cadena, y una región con un cebador inverso. La región de doble cadena, cuando se hibrida, puede contener una proyección T al objeto de unirse al ADN de doble cadena con una proyección A.

En una realización, el adaptador puede ser un bucle de ADN donde las regiones terminales son complementarias y donde la región del bucle contiene una región etiquetada del cebador directo (LFT), una región etiquetada del cebador inverso (LRT) y un punto de clivaje entre ambas (ver la Figura 10). 101 se refiere al ADN diana de extremos romos y doble cadena. 102 se refiere al ADN diana de cola A. 103 se refiere al adaptador de unión en bucle con proyección "T" y el punto de clivaje "Z". 104 se refiere al ADN diana con adaptadores de unión en bucle unidos. 105 se refiere al ADN diana con adaptadores de unión unidos clivados en el punto de clivaje. LFT se refiere a la etiqueta directa del adaptador de unión y LRT se refiere a la etiqueta inversa del adaptador de unión. La región complementaria puede terminar en una proyección T u otra característica que se puede utilizar para la unión del ADN diana. El punto de clivaje puede ser una serie de uracilos para el clivaje mediante UNG, o una secuencia que puede ser reconocida o clivada por una enzima de restricción u otro método de clivaje o simplemente una amplificación básica. Estos adaptadores se pueden utilizar para la elaboración de una biblioteca, por ejemplo, para la secuenciación. Estos adaptadores se pueden utilizar en combinación con cualquiera de los otros métodos descritos en el presente documento, por ejemplo, los métodos de amplificación por mini-PCR.

*Cebadores etiquetados internamente*

Cuando se utiliza la secuenciación para determinar el alelo presente en un determinado locus polimórfico, la lectura de la secuencia normalmente comienza en sentido ascendente del punto de unión del cebador (a) y hacia el punto polimórfico (X). Las etiquetas están configuradas típicamente como se muestra en la Figura 11 (izquierda). 101 se refiere al ADN diana de cadena simple con el locus polimórfico de interés "X" y el cebador "a" con la etiqueta "b" unida. A fin de evitar la hibridación no específica, el punto de unión del cebador (región de ADN diana complementaria de "a") tiene típicamente de 18 a 30 pb de longitud. La etiqueta de la secuencia "b" tiene típicamente 20 pb; en teoría estas pueden tener cualquier longitud superior a 15 pb, aunque muchas personas utilizan secuencias de cebadores comercializadas por la empresa de la plataforma de secuenciación. La distancia "d" entre "a" y "X" puede ser al menos de 2 pb para evitar un sesgo alélico. Si se realiza la amplificación por PCR multiplexada utilizando los métodos descritos en el presente documento u otros métodos, cuando es necesario prestar atención al diseño del cebador para evitar una interacción excesiva entre cebadores, la ventana de la distancia aceptable "d" entre "a" y "X" puede variar bastante; de 2 a 10 pb, de 2 a 20 pb, de 2 a 30 pb o incluso de 2 a más de 30 pb. Por tanto, cuando se utiliza la configuración del cebador mostrada en la Figura 11 (izquierda), las lecturas de secuencia deben tener como mínimo 40 pb para obtener lecturas lo suficientemente largas como para medir el locus polimórfico, y dependiendo de las longitudes de "a" y "d" puede que las lecturas de secuencia tengan que tener hasta 60 o 75 pb. Por lo general, cuanto más largas son las lecturas de secuencia, mayor es el coste y el tiempo de la secuenciación de un determinado número de lecturas; por tanto, al minimizar la longitud de lectura necesaria se puede ahorrar tanto tiempo como dinero. Por otra parte, dado que, de media, las bases leídas con

anterioridad en la lectura son leídas de forma más precisa que las que se leen más avanzada la lectura, al reducir la longitud de la lectura de secuencia necesaria también se aumenta la precisión de las mediciones de la región polimórfica.

En una realización, denominada cebadores etiquetados internamente, el punto de unión del cebador (a) se divide en diversos segmentos (a', a'', a'''...) y la etiqueta de la secuencia (b) se encuentra en un segmento del ADN que se sitúa en el centro de dos puntos de unión de cebadores, tal y como se muestra en la Figura 11 (103). Esta configuración permite al secuenciador realizar lecturas de secuencia más cortas. En una realización, a' + a'' deben tener al menos 18 pb y pueden tener hasta 30, 40, 50, 60, 80, 100 o más de 100 pb. En una realización, a'' debe tener al menos 6 pb y en una realización tiene entre unos 8 y 16 pb. Manteniéndose todos los demás factores sin cambios, el uso de cebadores etiquetados internamente puede recortar la longitud de las lecturas de secuencia necesarias al menos en 6 pb, como hasta 8 pb, 10 pb, 12 pb, 15 pb, e incluso hasta 20 pb o 30 pb. Esto puede ofrecer grandes ventajas en términos de tiempo, dinero y precisión. Un ejemplo de cebadores etiquetados internamente se recoge en la Figura 12.

*Cebadores con región de unión del adaptador de unión*

Un problema del ADN fragmentado es, dado que tiene una longitud corta, la probabilidad de que un polimorfismo se encuentre cerca del extremo de una cadena de ADN es mayor que en el caso de una cadena larga (por ejemplo,101, Figura 10). Dado que la captura de un polimorfismo requiere un punto de unión del cebador de una longitud adecuada a ambos lados del polimorfismo, un número significativo de cadenas de ADN con el polimorfismo focalizado se perderán debido a un solapamiento insuficiente entre el cebador y el punto de unión focalizado. En una realización, el ADN diana 101 puede tener adaptadores de unión unidos 102 y el cebador diana 103 puede tener una región (cr) que es complementaria a la etiqueta del adaptador de unión (lt) unida en sentido ascendente de la región de unión diseñada (a) (ver la Figura 13); por tanto, en los casos en los que la región de unión (región 101 complementaria de a) tiene menos de los 18 pb típicamente necesarios para la hibridación, la región (cr) del cebador que es complementaria a la etiqueta de la biblioteca puede aumentar la energía de unión a un punto en el que se puede producir la PCR. Cabe señalar que cualquier especificidad que se pierde debido a una región de unión más corta se puede producir a través de otros cebadores para la PCR con regiones de unión diana con una longitud adecuada. Cabe señalar que esta realización se puede utilizar en combinación con la PCR directa, o cualesquiera otros métodos descritos en el presente documento, como la PCR anidada, la PCR semi-anidada, la PCR hemi-anidada, la PCR unilateral o semi o hemi-anidada, u otros protocolos de PCR.

Cuando se utilizan los datos de secuencia para determinar el estado de ploidía en combinación con un método analítico que implica la comparación de los datos alélicos observados respecto de las distribuciones alélicas previstas para diversas hipótesis, cada lectura adicional de los alelos con una profundidad de lectura baja producirá más información que una lectura de un alelo con una profundidad de lectura elevada. Por tanto, idealmente, se pretende obtener una profundidad de lectura uniforme (DOR) en la que cada locus tenga un número similar de lecturas de secuencia representativas. Así pues, resulta recomendable minimizar la varianza de DOR. En una realización, resulta posible reducir el coeficiente de varianza de la DOR (que se puede definir como la desviación estándar de la DOR/la DOR media) aumentando los tiempos de hibridación. En algunas realizaciones, los tiempos de hibridación pueden ser de más de 2 minutos, más de 4 minutos, más de 10 minutos, más de 30 minutos, más de una hora o incluso más. Dado que la hibridación es un proceso de equilibrio, no existe límite a la mejora de la varianza de la DOR con el aumento de los tiempos de hibridación. En una realización, al aumentar la concentración del cebador se puede reducir la varianza de la DOR.

*Caja de diagnóstico*

En una realización, la presente divulgación comprende una caja de diagnóstico que es capaz de realizar parcial o completamente cualquiera de los métodos descritos en esta divulgación. En una realización, la caja de diagnóstico puede estar ubicada en la consulta de un médico, el laboratorio de un hospital o cualquier lugar adecuado y relativamente próximo al punto de atención del paciente. La caja puede ser capaz de ejecutar el método completo de forma totalmente automatizada o puede que un técnico tenga que completar uno o varios pasos manualmente. En una realización, la caja puede ser capaz de analizar al menos los datos genotípicos medidos en el plasma materno. En una realización, la caja puede estar conectada a medios que transmiten los datos genotípicos medidos en la caja de diagnóstico a una instalación de computación externa que puede posteriormente analizar los datos genotípicos y posiblemente también generar un informe. La caja de diagnóstico puede incluir una unidad robótica capaz de transferir muestras acuosas o líquidas de un recipiente a otro. Puede comprender una serie de reactivos, tanto sólidos como líquidos. Puede comprender un secuenciador de alto rendimiento. Puede comprender un ordenador.

*Kit para cebadores*

En algunas realizaciones de la divulgación, se puede formular un kit que comprenda una pluralidad de cebadores diseñados para conseguir los métodos descritos en la presente divulgación. Los cebadores pueden ser cebadores externos directos e inversos, cebadores internos directos e inversos como los divulgados en el presente documento; podrían ser cebadores diseñados para presentar una escasa afinidad de unión con otros cebadores del kit como se divulga en la sección sobre el diseño de los cebadores; podrían ser sondas de captura híbrida o sondas precircularizadas como las descritas en las correspondientes secciones, o una combinación de estos. En una realización, se puede formular un kit para determinar el estado de ploidía de un cromosoma diana en un feto en

gestación diseñado para ser utilizado con los métodos divulgados en el presente documento, donde el kit comprende una pluralidad de cebadores internos directos y opcionalmente la pluralidad de cebadores internos inversos, y opcionalmente cebadores externos directos y cebadores externos inversos, donde cada uno de los cebadores está diseñado para hibridarse con la región de ADN inmediatamente en sentido ascendente y/o en sentido descendente de uno de los puntos polimórficos del cromosoma diana, y opcionalmente cromosomas adicionales. En una realización, el kit para cebadores se puede utilizar en combinación con la caja de diagnóstico descrita en otro apartado de este documento.

*Composiciones de ADN*

Cuando se realiza un análisis informático de los datos de secuencia medidos en una mezcla de sangre materna y fetal para determinar la información genómica correspondiente al feto, por ejemplo, el estado de ploidía del feto, puede resultar recomendable medir las distribuciones alélicas de un conjunto de alelos. Lamentablemente, en muchos casos, por ejemplo, cuando se intenta determinar el estado de ploidía de un feto a partir de la mezcla de ADN que se encuentra en el plasma de una muestra de sangre materna, la cantidad de ADN disponible no es suficiente como para medir directamente las distribuciones alélicas de una forma fiable en la muestra. En estos casos, la amplificación de la mezcla de ADN proporcionará un número suficiente de moléculas de ADN como para medir de forma fiable las distribuciones alélicas deseadas. Sin embargo, los métodos actuales de amplificación que se utilizan típicamente en la amplificación de ADN para la secuenciación a menudo presentan sesgos importantes, lo que significa que no amplifican los dos alelos de un locus polimórfico de forma homogénea. Una amplificación sesgada puede producir distribuciones alélicas bastante diferentes de las presentes en la mezcla original. Para la mayoría de fines, no son necesarias mediciones altamente precisas de las cantidades relativas de alelos presentes en los loci polimórficos. Sin embargo, en una realización de la presente divulgación, los métodos de amplificación o enriquecimiento que enriquecen de forma específica los alelos polimórficos y preservan los ratios de alelos resultan ventajosos.

En el presente documento se describen diversos métodos que se pueden utilizar para enriquecer preferentemente una muestra de ADN en diversos loci de forma que se minimice el sesgo alélico. Algunos ejemplos incluyen el uso de sondas de circularización para focalizar una pluralidad de loci donde los extremos 3' y 5' de la sonda precircularizada están diseñados para hibridarse con bases que se encuentran a una o unas cuantas posiciones de distancia de los puntos polimórficos del alelo focalizado. Otro ejemplo consiste en utilizar sondas para la PCR donde la sonda para la PCR del extremo 3' está diseñada para hibridarse a bases que se encuentran a una o unas cuantas posiciones de distancia de los puntos polimórficos del alelo focalizado. Otro ejemplo consiste en utilizar un método de división y reagrupación para crear mezclas de ADN en las que los loci enriquecidos preferentemente son enriquecidos con un escaso sesgo alélico sin las desventajas del multiplexado directo. Otro ejemplo es el uso de un método de captura híbrida, donde las sondas de captura están diseñadas de forma que la región de la sonda de captura que está diseñada para hibridarse con el ADN que flanquea el punto polimórfico de la diana está separada del punto polimórfico por uno o un número reducido de bases.

Cuando las distribuciones alélicas medidas en un conjunto de loci polimórficos se utilizan para determinar el estado de ploidía de un individuo, resulta recomendable preservar las cantidades relativas de alelos de una muestra de ADN mientras se prepara para las mediciones genéticas. Esta preparación puede implicar la amplificación por WGA, la amplificación focalizada, técnicas de enriquecimiento selectivo, técnicas de captura híbrida, sondas de circularización u otros métodos diseñados para amplificar la cantidad de ADN y/o mejorar selectivamente la presencia de moléculas de ADN correspondientes a determinados alelos.

En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas de ADN diseñadas para loci diana, donde los loci tienen unas frecuencias alélicas menores máximas. En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas que están diseñadas para una diana donde los loci tienen la probabilidad máxima de que el feto tenga un SNP altamente informativo en esos loci. En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas diseñadas para loci diana, donde las sondas están optimizadas para un subgrupo de población determinado. En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas diseñadas para loci diana, donde las sondas están optimizadas para una mezcla determinada de subgrupos de población. En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas diseñadas para loci diana, donde las sondas están optimizadas para una determinada pareja de progenitores que pertenecen a subgrupos de población diferentes que tienen perfiles de frecuencia alélica menor diferentes. En algunas realizaciones de la presente divulgación, se presenta una cadena circularizada de ADN que comprende al menos un par de bases que se hibrida a una porción de ADN que es de origen fetal. En algunas realizaciones de la presente divulgación, se presenta una cadena circularizada de ADN que comprende al menos un par de bases que se hibrida a una porción de ADN que es de origen placentario. En algunas realizaciones de la presente divulgación, se presenta una cadena circularizada de ADN que se circularizó mientras que al menos algunos de los nucleótidos se hibridaban con ADN de origen fetal. En algunas realizaciones de la presente divulgación, se presenta una cadena circularizada de ADN que se circularizó mientras que al menos algunos de los nucleótidos se hibridaban con ADN de origen placentario. En algunas realizaciones de la presente divulgación, se presenta un conjunto de sondas donde algunas de las sondas están dirigidas a repeticiones en tándem simples y algunas de las sondas están dirigidas a polimorfismos de un solo nucleótido. En algunas realizaciones, los loci se seleccionan para fines del diagnóstico prenatal no invasivo. En algunas realizaciones, las sondas se utilizan para fines del diagnóstico prenatal no invasivo. En algunas realizaciones, los loci se focalizan utilizando un método que

podría incluir sondas de circularización, MIP, sondas de captura mediante hibridación, sondas de un array de SNP o combinaciones de estas. En algunas realizaciones, las sondas se utilizan como sondas de circularización, MIP, sondas de captura mediante hibridación, sondas de un array de SNP o combinaciones de estas. En algunas realizaciones, los loci se secuencian para los fines del diagnóstico prenatal no invasivo.

Cuando la capacidad de información relativa de una secuencia es mayor si se combina con los contextos parentales relevantes, al maximizar el número de lecturas de secuencia que contienen un SNP para el que se conoce el contexto parental se puede maximizar la capacidad de información del conjunto de lecturas de secuencia de la muestra combinada. En una realización, el número de lecturas de secuencia que contienen un SNP para el que se conocen los contextos parentales se puede aumentar utilizando la qPCR para amplificar preferentemente secuencias específicas. En una realización, el número de lecturas de secuencia que contienen un SNP para el que se conocen los contextos parentales se puede aumentar utilizando sondas de circularización (por ejemplo, MIP) para amplificar preferentemente secuencias específicas. En una realización, el número de lecturas de secuencia que contienen un SNP para el que se conocen los contextos parentales se puede aumentar utilizando un método de captura por hibridación (por ejemplo, SURSELECT) para amplificar preferentemente secuencias específicas. Se pueden utilizar métodos diferentes para aumentar el número de lecturas de secuencia que contienen un SNP para el que se conocen los contextos parentales. En una realización, la focalización se puede realizar mediante unión con extensión, unión sin extensión, captura por hibridación o PCR.

En una muestra de ADN genómico fragmentado, una parte de las secuencias de ADN corresponde de forma exclusiva a cromosomas individuales; otras secuencias de ADN se pueden encontrar en diferentes cromosomas. Cabe señalar que el ADN que se encuentra en el plasma, sea de origen materno o fetal, está típicamente fragmentado, a menudo con longitudes inferiores a 500 pb. En una muestra genómica típica, aproximadamente el 3,3% de las secuencias correlacionables se correlacionarán con el cromosoma 13; el 2,2% de las secuencias correlacionables se correlacionarán con el cromosoma 18; el 1,35% de las secuencias correlacionables se correlacionarán con el cromosoma 21; el 4,5% de las secuencias correlacionables se correlacionarán con el cromosoma X en una mujer; el 2,25% de las secuencias correlacionables se correlacionarán con el cromosoma X (en un hombre); y 0,73% de las secuencias correlacionables se correlacionarán con el cromosoma Y (en un hombre). Estos son los cromosomas con más probabilidades de presentar estado de aneuploidía en un feto. Por otra parte, entre las secuencias cortas, aproximadamente 1 de 20 secuencias contendrá un SNP, utilizando los SNP contenidos en dbSNP. La proporción podrá ser superior dado que hay muchos SNP que no se han descubierto.

En una realización de la presente divulgación, se pueden utilizar métodos de focalización para mejorar la fracción de ADN de una muestra de ADN que corresponde a un cromosoma dado, de forma que la fracción supere notablemente los porcentajes anteriormente indicados que son típicos para las muestras genómicas. En una realización de la presente divulgación, se pueden utilizar métodos de focalización para mejorar la fracción de ADN de una muestra de ADN de forma que el porcentaje de secuencias que contienen SNP sea notablemente mayor que los porcentajes típicos para las muestras genómicas. En una realización de la presente divulgación, se pueden utilizar métodos de focalización para focalizar ADN de un cromosoma o de un conjunto de SNP en una mezcla de ADN materno y fetal para los fines del diagnóstico prenatal.

Cabe señalar que se ha documentado un método (Patente USA 7 888 017) para determinar la aneuploidía fetal realizando un recuento del número de lecturas que corresponden a un cromosoma sospechoso y comparándolo con el número de lecturas que corresponden a un cromosoma de referencia, y utilizando el supuesto de que una sobreabundancia de lecturas en el cromosoma sospechoso se corresponde a un estado de triploidía en el feto en ese cromosoma. Estos métodos para el diagnóstico prenatal no harían uso de ningún tipo de focalización ni describen el uso de la focalización para el diagnóstico prenatal.

Haciendo uso de métodos de focalización en la secuenciación de la muestra combinada, se puede conseguir un determinado nivel de precisión con menos lecturas de secuencia. La precisión se puede referir a la sensibilidad, se puede referir a la especificidad o se puede referir a una combinación de estas. El nivel deseado de precisión puede ser entre 90 y 95%; puede ser entre 95% y 98%; puede ser entre 98% y 99%; puede ser entre 99% y 99,5%; puede ser entre 99,5% y 99,9%; puede ser entre 99,9% y 99,99%; puede ser entre 99,99% y 99,999%; puede ser entre 99,999% y 100%. Los niveles de precisión superiores al 95% se pueden denominar de alta precisión.

Existen diversos métodos publicados en la técnica que demuestran cómo se puede determinar el estado de ploidía de un feto a partir de una muestra combinada de ADN materno y fetal, por ejemplo: G.J. W. Liao et al. Clinical Chemistry 2011; 57(1) pp. 92-101. Estos métodos se centran en miles de ubicaciones a lo largo de cada cromosoma. El número de ubicaciones a lo largo de un cromosoma que se pueden focalizar al tiempo que se consigue una determinación de alta precisión del estado de ploidía en un feto, para un determinado número de lecturas de secuencia con una muestra de ADN mezclado, es inesperadamente limitado. En una realización de la presente divulgación, una determinación precisa del estado de ploidía se puede realizar utilizando secuenciación focalizada, utilizando cualquier método de focalización, por ejemplo, qPCR, PCR mediada por ligando, otros métodos de PCR, captura por hibridación, o sondas de circularización, donde el número de loci a lo largo de un cromosoma que necesita ser focalizado puede ser de entre 5000 y 2000 loci; puede ser de entra 2000 y 1000 loci; puede ser de entre 1000 y 500 loci; puede ser de entre 500 y 300 loci; puede ser de entre 300 y 200 loci; puede ser de entre 200 y 150 loci; puede ser de entre 150 y 100 loci; puede ser de entre 100 y 50 loci; puede ser de entre 50 y 20 loci; puede ser de entre 20 y 10 loci. Óptimamente, puede ser de entre 100 y 500 loci. El elevado nivel de

precisión se puede conseguir focalizando un pequeño número de loci y ejecutando un número inesperadamente reducido de lecturas de secuencia. El número de lecturas puede ser entre 100 millones y 50 millones de lecturas; el número de lecturas puede ser entre 50 millones y 20 millones de lecturas; el número de lecturas puede ser entre 20 millones y 10 millones de lecturas; el número de lecturas puede ser entre 10 millones y 5 millones de lecturas; el número de lecturas puede ser entre 5 millones y 2 millones de lecturas; el número de lecturas puede ser entre 2 millones y 1 millón de lecturas; el número de lecturas puede ser entre 1 millón y 500 000 lecturas; el número de lecturas puede ser entre 500 000 y 200 000 lecturas; el número de lecturas puede ser entre 20 000 y 10 000 lecturas; el número de lecturas puede ser inferior a 10 000. Cuanto mayor es la cantidad de ADN disponible, menos lecturas se necesitan.

En algunas realizaciones de la divulgación, se presenta una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponden exclusivamente al cromosoma 13 es superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, superior al 20%, superior al 25%, o superior al 30%. En algunas realizaciones de la presente divulgación, se presenta una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponden exclusivamente al cromosoma 18 es superior al 3%, superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, superior al 20%, superior al 25%, o superior al 30%. En algunas realizaciones de la presente divulgación, se presenta una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponden exclusivamente al cromosoma 21 es superior al 2%, superior al 3%, superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, superior al 20%, superior al 25%, o superior al 30%. En algunas realizaciones de la divulgación, se presenta una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponden exclusivamente al cromosoma X es superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, superior al 20%, superior al 25%, o superior al 30%. En algunas realizaciones de la presente divulgación, se presenta una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponden exclusivamente al cromosoma Y es superior al 1%, superior al 2%, superior al 3%, superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, superior al 20%, superior al 25%, o superior al 30%.

En algunas realizaciones de la divulgación, se describe una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponde exclusivamente a un cromosoma y que contiene al menos un polimorfismo de un único nucleótido es superior al 0,2%, superior al 0,3%, superior al 0,4%, superior al 0,5%, superior al 0,6%, superior al 0,7%, superior al 0,8%, superior al 0,9%, superior al 1%, superior al 1,2%, superior al 1,4%, superior al 1,6%, superior al 1,8%, superior al 2%, superior al 2,5%, superior al 3%, superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, o superior al 20%, y donde el cromosoma se toma del grupo 13,18,21, X, o Y. En algunas realizaciones de la presente divulgación, se describe una composición que comprende una mezcla de ADN de origen fetal y ADN de origen materno, donde el porcentaje de secuencias que corresponde exclusivamente a un cromosoma y que contiene al menos un polimorfismo de un único nucleótido de un conjunto de polimorfismos de un único nucleótido es superior al 0,15%, superior al 0,2%, superior al 0,3%, superior al 0,4%, superior al 0,5%, superior al 0,6%, superior al 0,7%, superior al 0,8%, superior al 0,9%, superior al 1%, superior al 1,2%, superior al 1,4%, superior al 1,6%, superior al 1,8%, superior al 2%, superior al 2,5%, superior al 3%, superior al 4%, superior al 5%, superior al 6%, superior al 7%, superior al 8%, superior al 9%, superior al 10%, superior al 12%, superior al 15%, o superior al 20%, donde el cromosoma se toma del grupo de cromosomas 13, 18, 21, X e Y, y donde el número de polimorfismos de un único nucleótido del conjunto de polimorfismos de un único nucleótido de entre 1 y 10, entre 10 y 20, entre 20 y 50, entre 50 y 100, entre 100 y 200, entre 200 y 500, entre 500 y 1000, entre 1000 y 2000, entre 2000 y 5000, entre 5000 y 10 000, entre 10 000 y 20 000, entre 20 000 y 50 000, y entre 50 000 y 100 000.

En teoría, cada ciclo de la amplificación duplica la cantidad de ADN presente; sin embargo, en la práctica el grado de amplificación es ligeramente inferior a dos. En teoría, la amplificación, incluyendo la amplificación focalizada, producirá una amplificación libre de sesgos de una mezcla de ADN; sin embargo, en la práctica los diferentes alelos tienden a amplificarse en diferente medida que otros alelos. Cuando se amplifica ADN, normalmente el grado de sesgo alélico aumenta con el número de pasos de amplificación. En algunas realizaciones, los métodos descritos en el presente documento implican la amplificación de ADN con un bajo nivel de sesgo alélico. Dado que el sesgo alélico aumenta con cada ciclo adicional, se puede determinar el sesgo alélico por ciclo calculando la raíz n-ésima del sesgo total donde n es el logaritmo en base 2 del grado de enriquecimiento. En algunas realizaciones, se presenta una composición que comprende una segunda mezcla de ADN, donde la segunda mezcla de ADN ha sido enriquecida preferentemente en diversos loci polimórficos de una primera mezcla de ADN, donde el grado de enriquecimiento es al menos de 10, al menos de 100, al menos de 1000, al menos de 10 000, al menos de 100 000 o al menos de 1 000 000, y donde el ratio de los alelos de la segunda mezcla de ADN en cada locus difiere del ratio de alelos en ese locus en la primera mezcla de ADN en un factor que es, de media, inferior al 1 000%, 500%, 200%, 100%, 50%, 20%, 10%, 5%, 2%, 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,02%, o 0,01%. En algunas realizaciones, se presenta una composición que comprende una segunda mezcla de ADN, donde la segunda mezcla de ADN ha sido enriquecida preferentemente en diversos loci polimórficos de una primera mezcla de ADN, donde el sesgo alélico por

ciclo para la pluralidad de loci polimórficos es, de media, inferior al 10%, 5%, 2%, 1%, 0,5%, 0,2%, 0,1%, 0,05%, o 0,02%. En algunas realizaciones, la pluralidad de loci polimórficos comprende al menos 10 loci, al menos 20 loci, al menos 50 loci, al menos 100 loci, al menos 200 loci, al menos 500 loci, al menos 1000 loci, al menos 2000 loci, al menos 5000 loci, al menos 10 000 loci, al menos 20 000 loci o al menos 50 000 loci.

*Cálculos de máxima probabilidad*

La mayoría de los métodos conocidos en la técnica para detectar la presencia o ausencia de fenómenos biológicos o condiciones médicas implican el uso de una prueba de rechazo de una única hipótesis, donde se mide un parámetro que está correlacionado con la condición, y si el parámetro se encuentra a un lado de un determinado umbral, la condición está presente, mientras que si se encuentra al otro lado del umbral está ausente. Una prueba de rechazo de una única hipótesis solo analiza la distribución nula al decidir entre las hipótesis nulas y alternativas. Sin tener en cuenta la distribución alternativa, no se puede estimar la probabilidad de cada hipótesis dados los datos observados y, por tanto, no se puede calcular la certeza de la determinación. Por tanto, con una prueba de rechazo de una única hipótesis, se obtiene una respuesta positiva o negativa sin una indicación de la certeza asociada al caso concreto.

En algunas realizaciones, los métodos divulgados en el presente documento pueden detectar la presencia o ausencia de fenómenos biológicos o condiciones médicas utilizando un método de probabilidad máxima. Esto supone una mejora sustancial con respecto a un método que utiliza una técnica de rechazo de una única hipótesis, dado que el umbral para determinar la ausencia o presencia de la condición se puede ajustar como corresponda para cada caso. Esto es particularmente relevante para las técnicas de diagnóstico que pretenden determinar la presencia o ausencia de aneuploidía en un feto en gestación con los datos genéticos disponibles de la mezcla del ADN fetal y materno presente en el ADN flotante libre que se encuentra en el plasma materno. Esto se debe a que cuando cambia la fracción de ADN fetal de la fracción obtenida del plasma, varía el umbral óptimo para determinar la aneuploidía frente a la euploidía. Cuando cae la fracción fetal, la distribución de los datos asociados con una aneuploidía es cada vez más similar a la distribución de los datos asociados a la euploidía.

El método de la estimación de la probabilidad máxima utiliza las distribuciones asociadas con cada hipótesis para estimar la probabilidad de los datos condicionados de cada hipótesis. Estas probabilidades condicionales se pueden convertir en la determinación de una hipótesis y de la certeza. Del mismo modo, un método de estimación de la máxima a posteriori utiliza las mismas probabilidades condicionales que la estimación de la probabilidad máxima, pero también incorpora datos previos de la población a la hora de elegir la mejor hipótesis y determinar la certeza.

Por tanto, el uso de una técnica de estimación de la probabilidad máxima (MLE), o la técnica estrechamente relacionada de la máxima a posteriori (MAP) ofrece dos ventajas: la primera que aumenta la probabilidad de una determinación correcta y la segunda que también permite un cálculo de la certeza para cada determinación. En una realización, la selección del estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada se realiza utilizando estimaciones de probabilidad máxima o estimaciones de la máxima a posteriori. En una realización, se divulga un método para determinar el estado de ploidía de un feto en gestación que implica la utilización de cualquier método actualmente conocido en la técnica que utiliza una técnica de rechazo de una única hipótesis y la reformulación de dicho método para usar una técnica de MLE o MAP. Algunos ejemplos de los métodos que se pueden mejorar de forma significativa aplicando estas técnicas se pueden encontrar en la Patente USA 8 008 018, Patente USA 7 888 017 o Patente USA 7 332 277.

En una realización, se divulga un método para determinar la presencia o ausencia de aneuploidía fetal en una muestra de plasma materno que comprende ADN genómico fetal y materno, donde el método consiste en lo siguiente: obtener una muestra de plasma materno; medir los fragmentos de ADN que se encuentran en la muestra de plasma con un secuenciador de alto rendimiento; correlacionar las secuencias con el cromosoma y determinar el número de lecturas de secuencia que corresponden a cada cromosoma; calcular la fracción de ADN fetal en la muestra de plasma; calcular una distribución prevista de la cantidad de un cromosoma diana que cabría esperar que estuviese presente si el segundo cromosoma diana fuese euploide y uno o una pluralidad de distribuciones previstas que cabría esperar si ese cromosoma fuese aneuploide, utilizando la fracción fetal y el número de lecturas de secuencia que corresponden a uno o una pluralidad de cromosomas de referencia que se espera que sean euploides; y utilizando una MLE o MAP para determinar cuáles de las distribuciones es más probable que sean correctas, indicando así la presencia o ausencia de aneuploidía fetal. En una realización, la medición de ADN del plasma puede implicar la realización de una secuenciación por fuerza masivamente paralela. En una realización, la medición del ADN de la muestra de plasma puede implicar la secuenciación de ADN que se ha enriquecido preferentemente, por ejemplo, mediante amplificación focalizada, en diversos loci polimórficos o no polimórficos. La pluralidad de loci puede estar diseñada para focalizar uno o un pequeño número de cromosomas que se sospecha que son aneuploides y uno o un pequeño número de cromosomas de referencia. El propósito del enriquecimiento preferente consiste en aumentar el número de lecturas de secuencia que son informativas para la determinación de la ploidía.

*Métodos informáticos para la determinación de la ploidía*

En el presente documento se divulga un método para determinar el estado de ploidía de los datos de una secuencia dada de un feto. En algunas realizaciones, los datos de esta secuencia se pueden medir en un secuenciador de alto rendimiento. En algunas realizaciones, los datos de la secuencia se pueden medir en ADN procedentes del ADN flotante libre aislado de la sangre materna, donde el ADN flotante libre comprende ADN de origen materno y ADN de

origen fetal/placentario. Esta sección describirá una realización de la presente divulgación en la que el estado de ploidía del feto se determina asumiendo que la fracción de ADN fetal de la mezcla que ha sido analizada se desconoce y se calculará a partir de los datos. También se describirá una realización en la que la fracción de ADN fetal ("fracción fetal") o el porcentaje de ADN fetal de la mezcla se puede medir mediante otro método y se supone que es conocida para determinar el estado de ploidía del feto. En algunas realizaciones, la fracción fetal se puede calcular utilizando solo mediciones del genotipo realizadas con la muestra de la sangre materna, que es una mezcla de ADN fetal y materno. En algunas realizaciones, la fracción se puede calcular utilizando también el genotipo medido o conocido de otro modo de la madre y/o el genotipo medido o conocido de otro modo del padre. En otra realización, el estado de ploidía del feto se puede determinar basado exclusivamente en la fracción calculada de ADN fetal para el cromosoma en cuestión en comparación con la fracción calculada de ADN fetal para el cromosoma de referencia que se considera disómico.

En la realización preferible de la divulgación, se supone que, para un cromosoma concreto, observamos y analizamos N SNP, para los que tenemos:

Un conjunto de mediciones de la secuencia de ADN flotante libre NR S=(s1,...,SNR). Dado que este método utiliza las mediciones de SNP, todos los datos de la secuencia que corresponden a loci no polimórficos se pueden ignorar. En una versión simplificada, donde tenemos recuentos (A,B) en cada SNP, donde A y B corresponden a los dos alelos presentes en un determinado locus, S se puede escribir como S=((a1,b1),...,(aN, bn)), donde a; es el recuento de A en el SNP i, bi es el recuento de B en el SNP i, y $\sum_{i=1:N}(a_i + b_i) = NR$

Los datos parentales se componen de

- o genotipos de un microarray de SNP u otra plataforma de determinación del genotipo basada en la intensidad: madre M=(m1,...,nN), padre F=(f1,..., fN), donde mi, fi ϵ (AA,AB, BB).

- o Y/O mediciones de los datos de la secuencia: NRM mediciones de la madre SM=(sm1,...,smnrm), NRF mediciones del padre SF=(sf1,...,sfnrf). De forma similar a la simplificación anterior, si tenemos recuentos (A,B) de cada SNP SM=((am1,bm1),...,(am1, bm1)), SF=((af1,bf1),...,(afN, bfN))

Colectivamente, los datos de la madre, del padre y del niño se denotan como D = (M,F,SM,SF,S). Cabe señalar que los datos parentales son deseables y aumentan la precisión del algoritmo, aunque NO son necesarios, en especial los datos del padre. Esto significa que incluso en ausencia de datos de la madre y/o el padre, es posible obtener resultados muy precisos del número de copias.

Se puede obtener la mejor estimación del número de copias (H*) maximizando la probabilidad del logaritmo de datos LIK(D|H) con respecto a todas las hipótesis (H) consideradas. En particular, se puede determinar la probabilidad relativa de cada una de las hipótesis de ploidía utilizando el modelo de distribución conjunto y los recuentos de alelos medidos en la muestra preparada, y empleando esas probabilidades relativas para determinar la hipótesis que es más probable que sea correcta como sigue:

$$H^* = \underset{H}{\operatorname{argmax}} \, LIK(D|H)$$

De forma similar, la probabilidad de las hipótesis a posteriori en función de los datos se pueden escribir como sigue:

$$H^* = \underset{H}{\operatorname{argmax}} \, LIK(D|H) * priorprob(H)$$

Donde priorprob(H) es la probabilidad previa asignada a cada hipótesis H, basada en el diseño del modelo y en el conocimiento previo.

También se pueden utilizar datos previos para hallar la estimación de la máxima a posteriori:

$$H_{MA} = \underset{H}{\operatorname{argmax}} \, LIK(D|H)$$

En una realización, las hipótesis del número de copias que se pueden considerar son:

- Monosomía

  - o materna H10 (una copia de la madre)

  - o paterna H10 (una copia del padre)

- Disomía: H11 (una copia de la madre y una copia del padre)

- Trisomía simple (sin considerar los cruces):

  - o Materna: H21_matched (dos copias idénticas de la madre, una copia del padre), H21_unmatched (AMBAS copias de la madre, una copia del padre)

o Paterna: H12_matched (una copia de la madre, dos copias idénticas del padre), H12_unmatched (una copia de la madre, ambas copias del padre)

• Trisomía compuesta, permitiendo cruces (utilizando un modelo de distribución conjunto):

o materna H21 (dos copias de la madre, una del padre),

o paterna HI2 (una copia de la madre, dos copias del padre)

En otras realizaciones, se pueden considerar otros estados de ploidía, como la nulisomía (H00), disomía uniparental (H20 y H02), y tetrasomía (H04, HI3, H22, H31 y H40).

Si no hay cruces, cada trisomía, con independencia de que el origen haya sido la mitosis, meiosis I o meiosis II, sería una de las trisomías emparejadas o no emparejadas. Debido a los cruces, la verdadera trisomía suele ser una combinación de ambas. En primer lugar, se describe un método para obtener probabilidades de hipótesis para las hipótesis simples. A continuación, se describe un método para obtener probabilidades de hipótesis para las hipótesis compuestas, que combinan la probabilidad individual del SNP con los cruces.

LIK(D\H) para una hipótesis simple

En una realización de la divulgación, se puede determinar LIK(D|H) para las hipótesis simples como sigue. Para las hipótesis simples H, LIK(H), la probabilidad logarítmica de la hipótesis H en un cromosoma completo, se puede calcular como la suma de las probabilidades logarítmicas de los SNP individuales, asumiendo la fracción cf conocida u obtenida del niño. En una realización se puede obtener cf a partir de los datos.

$$\mathrm{LIK(D|H)} = \sum_{\mathrm{i}} \mathrm{LIK(D|H, cf, i)}$$

Esta hipótesis no asume ningún enlace entre los SNP y, por tanto, no utiliza un modelo de distribución conjunto.

En algunas realizaciones de la divulgación, la probabilidad logarítmica se puede determinar para cada SNP. En un SNP i concreto, asumiendo la hipótesis de ploidía fetal H y un porcentaje de ADN fetal cf, la probabilidad logarítmica de los datos observados D se define como:

$$\mathrm{LIK(D|H, i)} = \log \mathrm{P(D|H, cf, i)} = \log\left(\sum_{\mathrm{m,f,c}} \mathrm{P(D|m, f, c, H, cf, i)P(c|m, f, H)P(m|i)P(f|i)}\right)$$

donde m son posibles genotipos verdaderos de la madre, f son posibles genotipos verdaderos del padre, donde m,f {AA,AB,BB}, y c son posibles genotipos del niño dada la hipótesis H. En concreto, para la monosomía c ϵ {A, B}, para la disomía c ϵ {AA,AB, BB}, para la trisomía c ϵ {AAA, AAB, ABB, BBB}.

Frecuencia previa del genotipo: p(m|i) es la probabilidad previa general del genotipo de la madre m en el SNP i, basado en la frecuencia de la población conocida en SNP I, denotada como pAi. En concreto

$$p(AA|pA_i) = (pA_i)^2, p(AB|pA_i) = 2(pA_i) * (1 - pA_i), p(BB|pA_i) = (1 - pA_i)^2$$

La probabilidad del genotipo del padre, p(f|i), se puede determinar de forma análoga.

Probabilidad verdadera del niño: p(c\m,f,H) es la probabilidad de obtener el genotipo verdadero del niño = c, dados los progenitores m, f, y asumiendo la hipótesis H, que se puede calcular fácilmente. Por ejemplo, para H11, H21 emparejado y H21 no emparejado, p(c|m,f,H) se proporciona a continuación.

| p(c\|m,f,H) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | H11 | | | H21 emparejado | | | | H21 no emparejado | | | |
| m | f | AA | AB | BB | AAA | AAB | ABB | BBB | AAA | AAB | AAB | BBB |
| AA | AA | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| AB | AA | 0,5 | ,05 | 0 | 0,5 | 0 | 0,5 | 0 | 0 | 1 | 0 | 0 |
| BB | AA | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| AA | AB | 0,5 | 0,5 | 0 | 0,5 | 0,5 | 0. | 0 | 0,5 | 0,5 | 0 | 0 |
| AB | AB | 0,25 | 0,5 | 0,25 | 0,25 | 0,25 | 0,25 | 0,25 | 0 | 0,5 | 0,5 | 0 |
| BB | AB | 0 | 0,5 | 0,5 | 0 | 0 | 0,5 | 0,5 | 0 | 0 | 0,5 | 0,5 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | BB | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| AB | BB | 0 | 0,5 | 0,5 | 0 | 0,5 | 0 | 0,5 | 0 | 0 | 1 | 0 |
| BB | BB | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |

Probabilidad de los datos: P(D|m, f, c, H,i, cf) es la probabilidad de los datos dados D en el SNP i, dado el genotipo verdadero de la madre m, el genotipo verdadero del padre f, el genotipo verdadero del niño c, la hipótesis H y la fracción del niño cf. Se pueden dividir en los datos de probabilidad de la madre, del padre y del niño como sigue:

$$P(D|m, f, c, H, cf, i) = P(SM|m, i)P(M|m, i)P(SF|f, i)P(F|f, i)P(S|m, c, H, cf, i)$$

Probabilidad de los datos del array del SNP de la madre: La probabilidad de los datos del genotipo del array del SNP de la madre Hi} en SNP i en comparación con el genotipo verdadero m, asumiendo que los genotipos del array de SNP son correctos, es simplemente

$$P(M|m, i) = \begin{cases} 1 & m_i = m \\ 0 & m_i \neq m \end{cases}$$

Probabilidad de los datos de la secuencia de la madre: la probabilidad de los datos de la secuencia de la madre en SNP i, en el caso de los recuentos Si=(ami,bmi), sin ningún ruido o sesgo adicional implicado, es la probabilidad binomial definida como P(SM|m,i)=Px|m(ami) donde X|m~Binom(pm(A), ami+bmi) con pm(A) se define como

| m | AA | AB | BB | A | B | Sin determinar |
|---|---|---|---|---|---|---|
| P(A) | 1 | 0,5 | 0 | 1 | 0 | 0,5 |

Probabilidad de los datos del padre: una ecuación similar se aplica a la probabilidad de los datos del padre.

Cabe señalar que se puede determinar el genotipo del hijo sin los datos parentales, especialmente los datos del padre. Por ejemplo, si no se dispone de los datos del genotipo del padre F, se puede utilizar simplemente P(F|f, i) = 1. Si no se dispone de los datos de la secuencia del padre, se puede utilizar simplemente P(SF|f,i)=1.

En algunas realizaciones de la divulgación, el método implica la elaboración de un modelo de distribución conjunto para los recuentos de alelos previstos en diversos loci polimórficos del cromosoma para cada una de las hipótesis de ploidía; en el presente documento se describe un método para ello. Probabilidad de los datos del ADN fetal libre: P(S|m, c, H, cf, i) es la probabilidad de los datos de la secuencia de ADN fetal libre en SNP i, dado el genotipo verdadero de la madre m, el genotipo verdadero del hijo c, la hipótesis del número de copias del hijo H, y asumiendo una fracción del hijo cf. De hecho, se trata de la probabilidad de los datos de la secuencia S en SNP I, dada la probabilidad verdadera de contenido A en SNP i μ (m, c, cf, H)

$$P(S|m, c, H, cf, i) = P(S|\mu(m, c, cf, H), i)$$

Para los recuentos, donde Si=(ai,bi), sin ningún ruido ni sesgo adicional de los datos,

$$P(S|\mu(m, c, cf, H), i) = P_x(a_i)$$

donde X~Binom(p(A), ai+bi) con p(A)= μ(m, c, cf, H). En un caso más complejo donde la alineación exacta y los recuentos (A,B) por SNP se desconocen, P(S|μ(m, c, cf, H), i) es una combinación de binomios integrados.

Probabilidad verdadera de contenido A: μ(m, c, cf, H), la probabilidad verdadera de contenido A en SNP i de esta mezcla madre/hijo, asumiendo el genotipo verdadero de la madre = m, el genotipo verdadero del niño = c, y la fracción del niño total = cf, se define como

$$\mu(m, c, cf, H) = \frac{\#A(m) * (1 - cf) + \#A(c) * cf}{n_m * (1 - cf) + n_c * cf}$$

donde #A(g) = número de A en el genotipo g, nm = 2 es la somía de la madre y nc es la ploidía del hijo bajo la hipótesis H (1 para monosomía, 1 para disomía, 3 para trisomía).

*Utilización de un modelo de distribución conjunto: LIK(D\H) para una hipótesis compuesta*

En algunas realizaciones de la divulgación, el método implica la elaboración de un modelo de distribución conjunto para los recuentos de alelos previstos en diversos loci polimórficos del cromosoma para cada una de las hipótesis de ploidía; en el presente documento se describe un método para ello. En muchos casos, habitualmente la trisomía no

está puramente emparejada o no emparejada, debido a los cruces, por lo que en esta sección se obtienen los resultados de las hipótesis compuestas H21 (trisomía materna) y H12 (trisomía paterna), que se combinan con la trisomía emparejada y no emparejada, teniendo en cuenta posibles cruces.

En el caso de la trisomía, si no hubiese cruces, la trisomía sería simplemente una trisomía emparejada o no emparejada. La trisomía emparejada es cuando el niño hereda dos copias del segmento del cromosoma idéntico de uno de los progenitores. La trisomía no emparejada es cuando el niño hereda una copia de cada segmento del cromosoma homólogo del progenitor. Debido a los cruces, algunos segmentos de un cromosoma pueden tener trisomía emparejada y otras partes presentar una trisomía no emparejada. En esta sección se describe cómo elaborar un modelo de distribución conjunto para los índices de heterocigosidad de un conjunto de alelos; es decir, para los recuentos de alelos previstos en una serie de loci para una o más hipótesis.

Suponiendo que SNP i, LIK(D|Hm,i) corresponde a la hipótesis Hm, y LIK(D|Hu,i) corresponde a la hipótesis Hu, y pc(i) = probabilidad de cruce entre SNPs i-1 andi. A continuación se puede calcular la probabilidad completa como:

$$\mathrm{LIK(D|H)} = \sum_{\mathrm{E}} \mathrm{LIK(D|E, 1{:}N)}$$

donde LIK(D|E,1:N) es la probabilidad de acabar en la hipótesis E, para SNP 1:N. E = hipótesis del último SNP, E ϵ (Hm, Hu). Repetidamente, se puede calcular:

$$\mathrm{LIK(D|E, 1{:}i)} = \mathrm{LIK(D|E, i)} + \log(\exp(\mathrm{LIK(D|E, 1{:}i-1)}) * \big(1 - \mathrm{pc(i)}\big) + \exp(\mathrm{LIK(D|{\sim}E, 1{:}i-1)}) * \mathrm{pc(i)})$$

donde ~E es la hipótesis distinta de E (no E), donde las hipótesis consideradas son Hm y Hu. En particular, se puede calcular la probabilidad de l:i SNP, basado en la probabilidad de1 a (i-1) SNP con la misma hipótesis y ningún cruce, o la hipótesis opuesta y un cruce, multiplicada por la probabilidad del SNP i

Para SNP 1, i=1, LIK(D|E, 1:1) = LIK(D|E, 1).

Para SNP 2, i=2,

$$\mathrm{LIK(D|E, 1{:}2)} = \mathrm{LIK(D|E, 2)} + \log(\exp(\mathrm{LIK(D|E, 1)}) * \big(1 - \mathrm{pc(2)}\big) + \exp(\mathrm{LIK(D|{\sim}E, 1)}) * \mathrm{pc(2)}),$$

y así sucesivamente para i=3 :N.

En algunas realizaciones de la divulgación se puede determinar la fracción del niño. La fracción del niño se puede referir a la proporción de secuencias en una mezcla de ADN procedente del niño. En el contexto del diagnóstico prenatal no invasivo, la fracción del niño se puede referir a la proporción de secuencias del plasma materno que proceden del feto o la proporción de la placenta con genotipo fetal. Se puede referir a la fracción del niño en una muestra de ADN que se ha preparado con plasma materno y que puede ser enriquecida en ADN fetal. Un propósito de la determinación de la fracción del niño en la muestra de ADN es el uso de un algoritmo que puede realizar determinaciones del estado de ploidía del feto; por tanto, la fracción del niño se podría referir a cualquier muestra de ADN analizada por secuenciación para los fines del diagnóstico prenatal no invasivo.

Algunos de los algoritmos presentados en esta divulgación que forman parte de un diagnóstico de aneuploidía prenatal no invasivo asumen una fracción del niño conocida, que no siempre es el caso. En una realización, se puede hallar la fracción del niño más probable maximizando la probabilidad para la disomía en cromosomas seleccionados, con o sin la presencia de los datos parentales.

En concreto, suponiendo que LIK(D| H11, cf, chr) = probabilidad logarítmica anteriormente descrita, para las hipótesis de disomía, y para la fracción del niño cf en el cromosoma chr. Para cromosomas seleccionados en Cset (normalmente 1:16), que se supone que son euploides, la probabilidad completa es:

$$\mathrm{LIK(cf)} = \sum_{\mathrm{chr} \in \mathrm{Cset}} \mathrm{Lik(D|\ H11, cf, chr)}$$

La fracción del niño más probable (cf*) se obtiene como cf* = argmaxcf LIK(cf).

Se puede utilizar cualquier conjunto de cromosomas. También se puede obtener la fracción del niño sin asumir la euploidía de los cromosomas de referencia. Utilizando este método, se puede determinar la fracción del niño para cualquiera de las situaciones siguientes: (1) se dispone de datos del array de los progenitores y datos de la secuenciación por fuerza bruta del plasma materno; (2) se dispone de datos del array de los progenitores y datos de la secuenciación focalizada del plasma materno; (3) se dispone de datos de la secuenciación focalizada de ambos progenitores y del plasma materno; (4) se dispone datos de la secuenciación focalizada de la madre y de la fracción de plasma materno; (5) se dispone de datos de la secuenciación focalizada de la fracción del plasma materno; (6) otras combinaciones de mediciones de la fracción parental y del niño.

En algunas realizaciones, el método informático puede incorporar pérdidas de datos; esto puede proporcionar determinaciones de la ploidía con una mayor precisión. En otro apartado de esta divulgación se ha asumido que la probabilidad de obtener una A es una función directa del genotipo verdadero de la madre, el genotipo verdadero del

niño, la fracción del niño en la mezcla y el número de copias del niño. También es posible que se produzca una pérdida de alelos de la madre o del niño, por ejemplo, en lugar de medir el verdadero AB del niño en la mezcla, se puede dar el caso de que solo se midan las secuencias correspondientes al alelo A. Se puede denotar la tasa de pérdida parental para los datos genómicos de Illumina $d_{pg}$, la tasa de pérdida parental para los datos de la secuencia $d_{ps}$ y la tasa de pérdida del niño para los datos de la secuencia $d_{cs}$. En algunas realizaciones, la tasa de pérdida de la madre se puede asumir que es cero y las tasas de pérdida del niño son relativamente bajas; en este caso, los resultados no se ven gravemente afectados por las pérdidas. En algunas realizaciones, la posibilidad de pérdidas de alelos puede ser suficientemente grande como para que produzca un efecto significativo sobre la determinación del estado de ploidía previsto. En este caso, las pérdidas de alelos se han incorporado al algoritmo como sigue:

Pérdidas de los datos del array del SNP parental: Para los datos genómicos de la madre M, se supone que el genotipo tras la pérdida es md, por tanto

$$P(M|m,i) = \sum_{m_d} P(M|m_d,i)P(m_d|m)$$

donde $P(M|m_d,i) = \begin{cases} 1 & m_i = m_d \\ 0 & m_i \neq m_d \end{cases}$ como antes, y P(md|m) es la probabilidad del genotipo $m_d$ tras la posible pérdida dado el verdadero genotipo m, definido más abajo, para la tasa de pérdida d

| | | | | md | | |
|---|---|---|---|---|---|---|
| m | AA | AB | BB | A | B | Sin determinar |
| AA | (1-d)∧2 | 0 | 0 | 2d(1-d) | 0 | d∧2 |
| AB | 0 | (1-d) ∧2 | 0 | d(1-d) | d(1-d) | d∧2 |
| BB | 0 | 0 | 0 (1-d) ∧ 2 | 0 | 2d(1-d) | d∧2 |

Una ecuación similar se aplica a los datos del array del SNP del padre.

Pérdidas de datos de la secuencia parental: Para los datos de la secuencia de la madre SM

$$P(SM|m,i) = \sum_{m_d} P_{X|m_d}(am_i)P(m_d|m)$$

donde $P(m_d|m)$ se define como en la sección anterior y la probabilidad $Px|m_d(ami)$ de una distribución binomial se define como anteriormente en la sección de la probabilidad de los datos parentales. Una ecuación similar se aplica a los datos de la secuencia paterna.

Pérdida de datos de la secuencia de ADN flotante libre:

$$P(S|m,c,H,cf,i) = \sum_{m_d,c_d} P(S|\mu(m_d,c_d,cf,H),i)P(m_d|m)P(c_d|c)$$

donde P(S|μ,(md, cd, cf, H), i) es como se define en la sección sobre probabilidad de los datos del ADN flotante libre.

En una realización de la divulgación, $p(m_d\backslash m)$ es la probabilidad del genotipo observado de la madre $m_d$, dado el genotipo verdadero de la madre m, asumiendo una tasa de pérdida $d_{ps}$, y $p(c_d|c)$ es la probabilidad del genotipo observado del niño cd, dado el genotipo verdadero del niño c, asumiendo una tasa de pérdida $d_{cs}$. Si $nA_T$ = número de alelos A en el genotipo verdadero c, nAD = número de alelos A en el genotipo observado cd, donde nAT > nAD, y de forma similar $nB_T$ = número de alelos B en el genotipo verdadero c, $nB_D$ = número de alelos B en el genotipo observado cd, donde $nB_T > nB_D$ y d = tasa de pérdida, entonces

$$\boldsymbol{p(c_d|c) = \binom{nA_T}{nA_D} * d^{nA_T - nA_D} * (1-d)^{nA_D} * \binom{nB_T}{nB_D} * d^{nB_T - nB_D} * (1-d)^{nB_D}}$$

En una realización de la divulgación, el método informático puede incorporar un sesgo aleatorio y constante. En una situación ideal no existe ningún ruido aleatorio ni sesgo de muestreo constante por SNP (además de la variación de la distribución binomial) en el número de recuentos de secuencias. En concreto, con respecto a SNP i, para el genotipo de la madre m, el genotipo verdadero del niño c y la fracción del niño cf, y X = el número de A en el conjunto de lecturas de (A+B) en SNP i, X actúa como un X~Binómico(p, A+B), donde p = μ(m, c, cf, H) = probabilidad verdadera de contenido de A.

En una realización, el método informático puede incorporar un sesgo aleatorio. Como suele ocurrir, se supone que existe un sesgo en las mediciones, de forma que la probabilidad de obtener una A en este SNP es igual a q, que es ligeramente diferente de p anteriormente descrito. El punto hasta el que p es diferente de q depende de la precisión del proceso de medición y de varios otros factores y se puede cuantificar a través de las desviaciones estándar de q con respecto a p. En una realización, se puede elaborar un modelo de q que tiene una distribución beta, con los parámetros a, β dependientes de la media de esa distribución centrada en p, y cierta desviación estándar especificada s. En concreto, esto da X\q~Bin(q,Di), donde q~Beta(a,β). Si E (q) =

P, V(q) = s2, y los parámetros a, β se pueden obtener como a = pN, β = (1— p)N, donde

$$N = \frac{p(1-p)}{s^2} - 1.$$

Esta es la definición de una distribución beta-binomial, donde se realiza un muestreo de una distribución binomial con un parámetro variable q, donde q sigue una distribución beta como una media p. Por tanto, en una configuración sin sesgo, con respecto a SNP i, la probabilidad de los datos de la secuencia paterna (SM) asumiendo el genotipo verdadero de la madre (m), dado el recuento de A de la secuencia de la madre en SNP i ($am_i$) y el recuento de B de la secuencia de la madre en SNP i ($bm_i$) se puede calcular como:

$$P(SM|m,i)=P_{X|m}(am_i) \text{ where } X|m\sim Binom(p_m(A), am_i+bm_i)$$

Ahora, incluyendo un sesgo aleatorio con una desviación estándar s, esto resulta:

$$X|m\sim BetaBinom(p_m(A), am_i+bm_i,s)$$

En caso de ausencia de sesgo, la probabilidad de los datos de la secuencia de ADN de plasma materno (S), asumiendo el genotipo verdadero de la madre (m), el genotipo verdadero del niño (c) y la fracción del niño (cf), asumiendo la hipótesis del niño H, dado un recuento de A de la secuencia de ADN flotante libre en SNP i (ai) y un recuento de B en la secuencia de ADN flotante libre en SNP i (bi), se puede calcular como sigue:

$$P(S|m, c, cf, H, i) = P_x(a_i)$$

donde X~Binom(p(A), ai+bi) con p(A)= μ(m, c, cf, H).

En una realización, incluyendo un sesgo aleatorio con una desviación estándar s, esto resulta X~BetaBinom(p(A),ai+bi,s), donde la cantidad de variación adicional viene determinada por el parámetro de desviación s, o equivalentemente N. Cuanto menor es el valor de s (o mayor el valor de N) más se aproximará esta distribución a la distribución binomial ordinaria. Se puede estimar la cantidad de sesgo, es decir estimar la N anterior, a partir de contextos no ambiguos AA|AA, BB|BB, AA|BB, BB|AA y utilizar el valor de N estimado en la probabilidad anterior. Dependiendo del comportamiento de los datos, N puede ser una constante con independencia de la profundidad de lectura ai+bi, o una función de ai+bi, haciendo el sesgo menor para las profundidades de lectura mayores.

En una realización de la divulgación, el método informático puede incorporar un sesgo constante por SNP. Debido a los artefactos del proceso de secuenciación, algunos SNP pueden tener unos recuentos más bajos o más altos con independencia de la cantidad verdadera de contenido de A. Supongamos que SNP i añade de forma constante un sesgo de un porcentaje $w_i$ al número de recuentos A. En algunas realizaciones, este sesgo se puede estimar a partir del conjunto de datos de formación obtenidos en las mismas condiciones, y añadirse a la estimación de los datos de la secuencia de los progenitores como sigue:

$$P(SM|m,i)=P_{X|m}(am_i) \text{ where } X|m\sim BetaBinom(p_m(A)+ w_i, am_i+bm_i,s)$$

y con la estimación de la probabilidad de los datos de la secuencia de ADN flotante libre como:

$P(S|m,c, cf, H, i) = P_x(a_i)$ donde $X\sim BetaBinom(p(A)+ w_i,a_i+b_i, s)$

En algunas realizaciones de la divulgación, el método puede estar escrito para tener en cuenta específicamente el ruido adicional, la calidad de la muestra diferencial, la calidad del SNP diferencial y el sesgo de muestreo aleatorio. Un ejemplo de esto se proporciona en el presente documento. Este método ha demostrado ser particularmente útil en el contexto de los datos generados utilizando el protocolo de mini-PCR masivamente multiplexada, y se utilizó en los Experimentos 7 a 13. El método implica varios pasos que introducen cada uno diferentes tipos de ruido y/o sesgo al modelo final:

(1) Se asume una primera muestra que contiene una mezcla de ADN materno y fetal que comprende una cantidad original de ADN de tamaño=N0 moléculas, ha= %refs verdaderas

(2) En la amplificación utilizando adaptadores de unión universales, se asume que se someten a muestreo moléculas Ni, normalmente moléculas N1 ~ N0/2 y el sesgo de muestreo aleatorio se introduce debido al muestreo. La muestra amplificada puede contener un número de moléculas N2 donde N2 >> N1. Suponiendo que X1 representa la cantidad de loci de referencia (por SNP) de moléculas N1 sometidas a muestreo, con una variación en p1=X1/N1 que introduce un sesgo de muestreo aleatorio durante el resto del protocolo. Este sesgo de muestreo se incluye en el modelo utilizando una distribución Beta-binomial (BB) en lugar de utilizar un modelo de distribución binomial simple. El parámetro N de la distribución Beta-binomial se puede estimar posteriormente para cada muestra a partir de los datos de formación, tras haber realizado el correspondiente ajuste por el sesgo de pérdida y amplificación, en los SNP con 0<p<1. La pérdida es la tendencia de un SNP a ser leído incorrectamente.

(3) El paso de amplificación amplificará cualquier sesgo alélico y, por tanto, el sesgo de amplificación introducido debido a una posible amplificación heterogénea. Asumiendo que un alelo de un locus sea amplificado f veces, otro alelo de ese locus sea amplificado g veces, donde $f=ge^b$, donde b=0 indica ausencia de sesgo. El parámetro del sesgo, b, está centrado en 0 e indica en qué grado mayor o mayor se ha amplificado el alelo A en comparación con el alelo B en un SNP concreto. El parámetro b puede diferir de un SNP a otro SNP. El parámetro del sesgo b se puede estimar por cada SNP, por ejemplo, a partir de los datos de formación.

(4) El paso de secuenciación implica la secuenciación de una muestra de moléculas amplificadas. En este paso puede haber una pérdida, donde la pérdida es la situación en la que un SNP es leído de forma incorrecta. La pérdida puede producirse por cualquier número de problemas y puede provocar que un SNP no sea leído como el alelo A correcto, sino como otro alelo B que se encuentra en ese locus o como un alelo C o D que no se encuentra típicamente en ese locus. Suponiendo que la secuenciación mide los datos de la secuencia de una serie de moléculas de ADN de una muestra amplificada de tamaño N2, donde N3 < N2. En algunas realizaciones, N3 se puede encontrar en el rango de 20 000 a 100 000; 100 000 a 500 000; 500 000 a 4 000 000; 4 000 000 a 20 000 000; o 20 000 000 a 100 000 000. Cada molécula sometida a muestreo tiene una probabilidad $p_g$ de ser leída correctamente, en cuyo caso aparecerá correctamente como alelo A. La muestra será leída incorrectamente como un alelo no relacionado con la molécula original con probabilidad $1-p_g$, y aparecerá como un alelo A con probabilidad $p_r$, un alelo B con probabilidad $p_m$ o un alelo C o alelo D con probabilidad $p_0$, donde $p_r+p_m+p_o=1$. Los parámetros $p_g$, $p_r$, $p_m$, $p_o$ se estiman para cada SNP a partir de los datos de formación.

Diferentes protocolos pueden implicar pasos similares con variaciones en los pasos de biología molecular que resultan en diferentes cantidades de muestreo aleatorio, diferentes niveles de amplificación y diferente sesgo de pérdidas. El siguiente modelo también se puede aplicar perfectamente a cada uno de estos casos. El modelo para la cantidad de ADN sometido a muestreo, por cada SNP, viene dado por:

X3~BetaBinómico(L(F(p,b),pr,Pg), N*H(p,b))

donde p = la cantidad verdadera de ADN de referencia, b = sesgo por SNP, y tal y como se ha descrito anteriormente, pg es la probabilidad de una lectura correcta, pr es la probabilidad de que una lectura sea leída de forma incorrecta pero accidentalmente parezca el alelo correcto, en caso de una mala lectura, tal y como se ha descrito anteriormente, y:

$$F(p,b)= pe^b/(pe^b+(1-p)),\ H(p,b) = (e^bp+(1-p))^2/e^b,\ L(p,p_r,p_g)=p*p_g+p_r*(1-p_g).$$

En algunas realizaciones de la divulgación, el método utiliza una distribución Beta-binomial en lugar de una distribución binomial simple El parámetro N de la distribución Beta-binomial es estimado para cada muestra o cuando resulta necesario. Utilizando una corrección del sesgo F(p,b), H(p,b), en lugar de tan solo p, se presta atención al sesgo de amplificación. El parámetro b del sesgo es estimado para cada SNP a partir de los datos de formación con antelación.

En algunas realizaciones de la divulgación, el método utiliza la corrección de pérdidas L(p,pr,pg), en lugar de tan solo p; de este modo se presta atención al sesgo de pérdida, es decir la variación del SNP y de la calidad de la muestra. En algunas realizaciones, los parámetros pg, pr, po se estiman para cada SNP a partir de los datos de formación con antelación. En algunas realizaciones, los parámetros pg, pr, po se pueden actualizar con la muestra actual sobre la marcha, para tomar en cuenta la variación de la calidad de la muestra.

El modelo descrito en el presente documento es bastante general y puede tener en cuenta tanto la calidad diferencial de la muestra como la calidad diferencial del SNP. Las diferentes muestras y SNP se tratan de forma diferente, tal y como se pone de manifiesto por el hecho de que algunas realizaciones utilizan distribuciones Beta-binomiales cuya media y varianza son una función de la cantidad original de ADN, así como la calidad de la muestra y del SNP.

**Elaboración de modelos de plataformas**

Se asume un único SNP en el que el ratio del alelo previsto presente en el plasma es r (basándose en los genotipos maternos y fetales). El ratio del alelo previsto se define como la fracción prevista de los alelos A en el ADN combinado materno y fetal. Para el genotipo materno gm y el genotipo del niño gc, el ratio del alelo previsto viene dado por la ecuación 1, asumiendo que los genotipos son representados también como ratios de alelos.

r = fgc + (1 - f)gm (1)

La observación en el SNP se compone del número de lecturas correlacionadas con cada alelo presente, na y nb, que se suman a la profundidad de lectura d. Se asume que los umbrales ya se han aplicado a las probabilidades de correlación y las puntuaciones phred de forma que las correlaciones y las observaciones de alelos se pueden considerar correctas. Una puntuación phred es una medida numérica que se refiere a la probabilidad de que una medición concreta en una base concreta sea errónea. En una realización, cuando la base ha sido medida mediante secuenciación, la puntuación phred se puede calcular a partir del ratio de la intensidad del colorante correspondiente a la base determinada para la intensidad del colorante de las demás bases. El modelo más sencillo para la probabilidad de observación es una distribución binomial que asume que cada una de las lecturas d se extrae independientemente de un amplio conjunto que tiene un ratio de alelos r. La ecuación 2 describe este modelo.

$$P(n_a,n_b|r) = p_{bino}(n_a;\ n_a + n_b,\ r) = \binom{n_a + n_b}{n_a} r^{n_a}(1-r)^{n_b} \quad (2)$$

El modelo binomial se puede ampliar de varias maneras. Cuando los genotipos materno y fetal son todos A o todos B, el ratio alélico previsto en el plasma será 0 o 1 y la probabilidad binomial no estará bien definida. En la práctica, en ocasiones se observan alelos imprevistos. En una realización, es posible utilizar un ratio alélico corregido r = l/(na + nb) para permitir un pequeño número de alelos imprevistos. En una realización, se pueden utilizar los datos de formación para elaborar un modelo de la tasa de alelos imprevistos que aparece en cada SNP y utilizar este modelo para corregir el ratio alélico previsto. Cuando el ratio alélico previsto no es 0 o 1, el ratio alélico observado puede no coincidir con una profundidad de lectura suficientemente elevada para el ratio alélico previsto, debido al sesgo de amplificación u otros fenómenos. A continuación, se puede elaborar un modelo del ratio alélico como una distribución beta centrada en el ratio alélico previsto, lo que conduce a una distribución beta-binomial para P(na, nb|r) que tiene una varianza más elevada que la binomial.

El modelo de la plataforma para la respuesta en un único SNP se definirá como F(a, b, gc, gm, f) (3), o la probabilidad de observar na = a y nb = b dados los genotipos materno y fetal, que también depende de la fracción fetal a través de la ecuación 1. La forma funcional de F puede ser una distribución binomial, una distribución beta-binomial o funciones similares como las anteriormente descritas.

$$F(a, b, g_c, g_m, f) = P(n_a = a, n_b = b|g_c, g_m, f) = P(n_a = a, n_b = b|r(g_c, g_m, f)) \quad (3)$$

En una realización de la divulgación, se puede determinar la fracción del niño como sigue. Una estimación de la probabilidad máxima de la fracción fetal f para una prueba prenatal se puede obtener sin el uso de información paterna. Esto puede resultar relevante cuando no se dispone de los datos genéticos paternos, por ejemplo, cuando el padre registrado no es realmente el padre genético del feto. La fracción fetal se estima a partir del conjunto de SNP en los que el genotipo materno es 0 o 1, lo que resulta en un conjunto de solamente dos genotipos fetales posibles. Se define $S_o$ como el conjunto de SNP con el genotipo materno 0 y S1 como el conjunto de SNP con el genotipo materno 1. Los genotipos fetales posibles en $S_0$ son 0 y 0,5, lo que resulta en un conjunto de ratios alélicos posibles R0(f) = {0,f/2}. De forma similar, R1(f) = {1-f/2,1}. Este método se puede ampliar de forma trivial para incluir los SNP donde el genotipo materno es 0,5, pero estos SNP serán menos informativos debido al conjunto mayor de ratios alélicos posibles.

Se define $N_{a0}$ y $N_{b0}$ como los vectores formados por $n_{as}$ y $n_{bs}$ para los SNP s en $S_0$, y $N_{a1}$ y $N_{b1}$ similarmente para S1. La estimación de la probabilidad máxima f de f se define por la ecuación 4.

$$\hat{f} = \arg\max_f P(N_{a0}, N_{b0}|f)\ P(N_{a1}, N_{b1}|f) \quad (4)$$

Asumiendo que los recuentos de alelos en cada SNP son independientes condicionados por el ratio alélico del plasma en los SNP, las probabilidades se puede expresar como productos sobre los SNP en cada conjunto (5).

$$P(N_{a0}, N_{b0}|f) = \prod_{s\epsilon S_0} P(n_{as}, n_{bs}|f) \quad (5)$$

$$P(N_{a1}, N_{b1}|f) = \prod_{s\epsilon S_1} P(n_{as}, n_{bs}|f)$$

La dependencia de f se produce a través de los posibles ratios alélicos $R_0(f)$ y $R_1(f)$. La probabilidad del SNP P (nas, nbS|f) se puede aproximar asumiendo el genotipo de la probabilidad máxima condicionado en f. A una fracción fetal y una profundidad de lectura relativamente elevadas, la selección del genotipo de la probabilidad máxima será de alta certeza. Por ejemplo, a una fracción fetal del 10 por ciento y una profundidad de lectura de 100, se considera un SNP donde la madre tiene el genotipo cero. Los ratios alélicos previstos son 0 y 5 por ciento, que serán fácilmente distinguibles a una profundidad de lectura suficientemente elevada. La sustitución del genotipo del niño estimado en la ecuación 5 resulta en la ecuación completa (6) para la estimación de la fracción fetal.

$$\hat{f} = \arg\max_f \left[\prod_{s\epsilon S_0}\left(\max_{r_s\epsilon R_0(f)} P(n_{as},\ n_{bs}|r_s)\right)\ \prod_{s\epsilon S_1}\ \left(\max_{r_s\epsilon R_1(f)} P(n_{as},\ n_{bs}|r_s)\right)\right] \quad (6)$$

La fracción fetal debe estar en el rango [0,1] y por tanto la optimización se puede implementar fácilmente a través de una búsqueda unidimensional limitada.

En presencia de una baja profundidad de lectura o un elevado nivel de ruido, puede ser preferible no asumir el genotipo de la probabilidad máxima, que puede proporcionar certezas artificialmente elevadas. Otro método consistiría en sumar los posibles genotipos de cada SNP, resultando en la siguiente expresión (7) para $P(n_a, n_b|f)$ para un SNP en $S_o$. La probabilidad previa P(r) se podría asumir uniforme en $R_o(f)$ o se podría basar en las frecuencias de la población. La extensión al grupo $S_i$ es insignificante.

$$P(n_a, n_b|f) = \sum_{r\epsilon R_0(f)} P(n_a, n_a|r)P(r) \quad (7)$$

En algunas realizaciones de la divulgación, se pueden determinar las probabilidades como sigue. Se puede calcular una certeza a partir de las probabilidades de los datos de las dos hipótesis $H_t$ y $H_f$. La probabilidad de cada hipótesis se obtiene basándose en el modelo de respuesta, la fracción fetal estimada, los genotipos maternos, las frecuencias alélicas de la población y los recuentos alélicos del plasma.

Se define la siguiente notación:

$G_m$, $G_e$ genotipos verdaderos materno y del niño

$G_{af}$, $G_{tf}$ genotipos verdaderos del supuesto padre y del verdadero padre

$G(g_c, g_m, g_{tf}) = P(GC = g_c|G_m = g_m, G_{tf} = g_{tf})$ probabilidades de herencia

$P(g) = P(G_{tf} = g)$ frecuencia de la población del genotipo g en un SNP concreto

Asumiendo que la observación en cada SNP está independiente condicionada en el ratio alélico del plasma, la probabilidad de una hipótesis de paternidad es el producto de las probabilidades de los SNP. Las siguientes ecuaciones calculan la probabilidad para un único SNP. La ecuación 8 es una expresión general para la probabilidad de cualquier hipótesis h, que posteriormente se dividirá en casos específicos de Ht y Hf.

$$\begin{aligned} P(n_a, n_b|h, G_m, G_{tf}, f) &= \sum_{g_c\epsilon(0,0.5,1)} P(n_a, n_b|G_c = g_c, G_m, G_{tf}, h, f)P(G_c = g_c, G_m, G_{tf}, h, f) \\ &= \sum_{g_c\epsilon(0,0.5,1)} P(n_a, n_b|G_c = g_c, G_m, f)P(G_c = g_c|G_m, G_{tf}, h) \\ &= \sum_{g_c\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)P(G_c = g_c|G_m, G_{tf}, h) \quad (8) \end{aligned}$$

En el caso de Ht, el supuesto padre es el verdadero padre y los genotipos fetales son heredados de los genotipos maternos y de los genotipos del supuesto padre en función de la ecuación 9.

$$\begin{aligned} P(n_a, n_b|H_t, G_m, G_{tf}, f) &= \sum_{g_c\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)P(G_c = g_c|G_m, G_{tf}, H_t) \quad (9) \\ &= \sum_{g_c\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)G(g_c, G_m, G_{tf}) \end{aligned}$$

En el caso de Hf, el supuesto padre no es el verdadero padre. La mejor estimación de los genotipos del verdadero padre vienen dados por las frecuencias de la población en cada SNP. Por tanto, las probabilidades de los genotipos del niño vienen determinadas por los genotipos de la madre conocidos y las frecuencias de la población, como en la ecuación 10.

$$\begin{aligned} P(n_a, n_b|H_t, G_m, G_{tf}, f) &= \sum_{g_c\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)P(G_c = g_c|G_m, G_{tf}, H_f) \\ &= \sum_{g_c\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)P(G_c = g_c|G_m) \\ &= \sum_{g_c\epsilon(0,0.5,1)} \sum_{g_{tf}\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)P(G_c = g_c|G_m, G_{tf} = g_{tf})P(G_{tf} = g_{tf}) \\ &= \sum_{g_c\epsilon(0,0.5,1)} \sum_{g_{tf}\epsilon(0,0.5,1)} F(n_a, n_b, g_c, g_m, f)G(g_c, G_m, g_{tf})P(g_{tf}) \end{aligned}$$

La certeza Cp de la paternidad correcta se calcula a partir del producto de los SNP de las dos probabilidades utilizando la regla de Bayes (11).

$$\mathrm{Cp} = \frac{\prod_s P(n_{as}, n_{bs} | H_t, G_{ms}, G_{tf}, f)}{\prod_s P(n_{as}, n_{bs} | H_t, G_{ms}, G_{tf}, f) + \prod_s P(n_{as}, n_{bs} | H_f, G_{ms}, G_{tf}, f)} \qquad (11)$$

*Modelo de la probabilidad máxima utilizando el porcentaje de la fracción fetal*

La determinación del estado de ploidía de un feto midiendo el ADN flotante libre contenido en suero materno o midiendo el material genotípico de cualquier muestra combinada, es un ejercicio no trivial. Existen diversos métodos, por ejemplo, para realizar un análisis del recuento de lecturas donde se presume que si el feto es trisómico en un determinado cromosoma, entonces la cantidad total de ADN de ese cromosoma que se encuentra en la sangre materna será elevada con respecto al cromosoma de referencia. Una forma de detectar la trisomía en estos fetos consiste en normalizar la cantidad de ADN prevista para cada cromosoma, por ejemplo, en función del número de SNP del conjunto de análisis que corresponde a un cromosoma dado o en función del número de porciones del cromosoma correlacionables de forma única. Una vez que se han normalizado las mediciones, se determina que cualesquiera cromosomas para los que la cantidad de ADN medida supere un determinado umbral son trisómicos. Este método se describe en Fan, et al. PNAS, 2008; 105(42); pp. 16266-16271, y también en Chiu et al. BMJ 2011;342:c7401. En el documento de Chiu et al., la normalización se realizó calculando una puntuación Z como sigue:

puntuación Z para el porcentaje del cromosoma 21 en el caso de ensayo = ((porcentaje del cromosoma en el caso de ensayo) - (porcentaje medio del cromosoma 21 en controles de referencia)) / (desviación estándar del porcentaje del cromosoma 21 en controles de referencia).

Estos métodos determinan el estado de ploidía del feto utilizando un método de rechazo de una única hipótesis. Sin embargo, sufren algunas carencias significativas. Puesto que estos métodos para la determinación del estado de ploidía en el feto son invariables en función del porcentaje de ADN fetal en la muestra, utilizan un valor de corte; el resultado de esto es que las precisiones de las determinaciones no son óptimas y aquellos casos en los que el porcentaje de ADN fetal en la mezcla es relativamente bajo sufrirán las peores precisiones.

En una realización, se utiliza un método de la presente divulgación para determinar el estado de ploidía del feto que implica tener en cuenta la fracción de ADN fetal de la muestra. En otra realización de la presente divulgación, el método implica el uso de estimaciones de la probabilidad máxima. En una realización, un método de la presente divulgación implica calcular el porcentaje de ADN en una muestra que es de origen fetal o placentario. En una realización, el umbral para la determinación del estado de aneuploidía se ajusta adaptativamente sobre el porcentaje de ADN fetal calculado. En algunas realizaciones, el método para estimar el porcentaje de ADN que es de origen fetal en una mezcla de ADN comprende la obtención de una muestra combinada que contiene material genético de la madre y material genético del feto, la obtención de una muestra genética del padre del feto, la medición del ADN en la muestra combinada, la medición del ADN en la muestra del padre, y el cálculo del porcentaje de ADN que es de origen fetal en la muestra combinada utilizando las mediciones de ADN de la muestra combinada y de la muestra del padre.

En una realización de la presenta divulgación, la fracción de ADN fetal o el porcentaje de ADN fetal en la muestra se puede medir. En algunas realizaciones, la fracción se puede calcular utilizando solo mediciones del genotipo realizadas con la muestra de plasma materno, que es una mezcla de ADN fetal y materno. En algunas realizaciones, la fracción se puede calcular utilizando también el genotipo medido o conocido de otro modo de la madre y/o el genotipo medido o conocido de otro modo del padre. En algunas realizaciones, el porcentaje de ADN fetal se puede calcular utilizando mediciones realizadas en la mezcla de ADN materno y fetal, junto con el conocimiento de los contextos parentales. En una realización, la fracción de ADN fetal se puede calcular utilizando frecuencias de la población para ajustar el modelo de la probabilidad a mediciones alélicas concretas.

En una realización de la presente divulgación, se puede calcular una certeza de la precisión de la determinación del estado de ploidía del feto. En una realización, la certeza de la hipótesis de mayor probabilidad (Hprincipal) se puede calcular como (1- Hprincipal) / Σ (todas las H). Se puede determinar la certeza de una hipótesis si se conocen las distribuciones de todas las hipótesis. Se puede determinar la distribución de todas las hipótesis, si se conoce la información del genotipo parental. Se puede calcular una certeza de la determinación del estado de ploidía si se conoce la distribución de los datos prevista para el feto euploide y la distribución de los datos prevista para el feto aneuploide. Se pueden calcular estas distribuciones previstas si se conocen los datos del genotipo parental. En una realización, se puede utilizar el conocimiento de la distribución de una estadística de ensayo con respecto a una hipótesis normal y con respecto a una hipótesis anómala para determinar tanto la fiabilidad de la ploidía como para reajustar el umbral y obtener una determinación más fiable. Esto resulta particularmente útil cuando la cantidad y/o el porcentaje de ADN fetal en la mezcla son bajos. Esto ayudará a evitar la situación de que un feto que sea realmente aneuploide se determine que es euploide porque una estadística de la prueba, como la estadística Z, no supera un umbral establecido en base al umbral optimizado para el caso en el que existe un porcentaje de ADN fetal superior.

En una realización, se puede utilizar un método divulgado en el presente documento para determinar una aneuploidía fetal estableciendo el número de copias de cromosomas diana maternos y fetales en una mezcla de material genético materno y fetal. Este método puede implicar la obtención de tejido materno que comprende tanto

material genético materno como fetal; en algunas realizaciones, este tejido materno puede ser plasma materno o un tejido aislado de sangre materna. Este método también puede implicar la obtención de una mezcla de material genético materno y fetal de dicho tejido materno mediante el procesamiento del mencionado tejido materno. Este método puede implicar la distribución del material genético obtenido en diversas mezclas de reacción para proporcionar aleatoriamente muestras de reacción individuales que comprenden una secuencia diana de un cromosoma diana y muestras de reacción individuales que no comprenden una secuencia diana de un cromosoma diana, por ejemplo, realizando una secuenciación de alto rendimiento de la muestra. Este método puede implicar el análisis de secuencias diana de material genético presente o ausente en dichas muestras de reacción individuales para proporcionar un primer número de resultados binarios que representan la presencia o ausencia de un cromosoma fetal presumiblemente euploide en las muestras de la reacción y un segundo número de resultados binarios que representan la presencia o ausencia de un cromosoma fetal posiblemente aneuploide en las muestras de la reacción. Cualquiera de las cifras de resultado binario se puede calcular, por ejemplo, con una técnica informática que hace el recuento de las lecturas de secuencia que corresponden a un determinado cromosoma, a una determinada región de un cromosoma, a un determinado locus o conjunto de loci. El método puede implicar la normalización del número de eventos binarios basándose en la longitud del cromosoma, la longitud de la región del cromosoma o el número de loci del conjunto. Este método puede implicar el cálculo de una distribución prevista del número de resultados binarios para un cromosoma fetal presumiblemente euploide en las muestras de reacción utilizando el primer número. Este método puede implicar el cálculo de una distribución prevista del número de resultados binarios para un cromosoma fetal presumiblemente aneuploide en las muestras de la reacción utilizando el primer número y una fracción estimada de ADN fetal que se encuentra en la muestra, por ejemplo, multiplicando la distribución del recuento leído previsto del número de resultados binarios para un cromosoma fetal presumiblemente euploide por (1 + n/2) donde n es la fracción fetal prevista. En algunas realizaciones, las lecturas de secuencia se pueden tratar en correspondencias probabilísticas más que resultados binarios; este método proporcionaría precisiones más elevadas, aunque requiere una mayor potencia de computación. La fracción fetal se puede estimar a través de una pluralidad de métodos, algunos de los cuales se describen en otro apartado de la presente divulgación. Este método puede implicar el uso de un método de la probabilidad máxima para determinar si el segundo número corresponde al cromosoma fetal posiblemente aneuploide que es euploide o aneuploide. Este método puede implicar que la determinación del estado de ploidía del feto sea el estado de ploidía que corresponde a la hipótesis con la probabilidad máxima de ser correcta dados los datos medidos.

Cabe señalar que se puede utilizar un modelo de probabilidad máxima para aumentar la precisión de cualquier método que determine el estado de ploidía de un feto. De forma similar, se puede calcular una certeza para cualquier método que determine el estado de ploidía del feto. El uso de un modelo de probabilidad máxima supondría una mejora de la precisión de cualquier método en el que se determine el estado de ploidía utilizando una técnica de rechazo de una única hipótesis. Se puede utilizar un modelo de probabilidad máxima para cualquier método en el que se pueda calcular una distribución de la probabilidad tanto para los casos normales como para los anómalos. El uso de un modelo de probabilidad máxima implica la capacidad de calcular una certeza para la determinación del estado de ploidía.

*Debate detallado del método*

En una realización, un método divulgado en el presente documento utiliza una medición cuantitativa del número de observaciones independientes de cada alelo en un locus polimórfico, donde esto no implica el cálculo del ratio de los alelos. Esto contrasta con otros métodos, como algunos métodos basados en microarrays, que proporcionan información sobre el ratio de dos alelos en un locus, pero no cuantifican el número de observaciones independientes de ninguno de los alelos. Algunos métodos conocidos en la técnica pueden proporcionar información cuantitativa sobre el número de observaciones independientes, pero los cálculos que conducen a la determinación del estado de ploidía utilizan únicamente los ratios de alelos y no utilizan la información cuantitativa. Para ilustrar la importancia de retener información sobre el número de observaciones independientes, tendremos en cuenta el locus de la muestra con dos alelos, A y B. En un primer experimento se observan 20 alelos A y 20 alelos B, y en un segundo experimento se observan 200 alelos A y 200 alelos B. En ambos experimentos el ratio (A/(A+B)) es igual a 0,5; sin embargo, el segundo experimento proporciona más información que el primero acerca de la certidumbre de la frecuencia del alelo A o B. El método instantáneo, en lugar de utilizar los ratios de los alelos, utiliza datos cuantitativos para elaborar un modelo más preciso de las frecuencias alélicas más probables en cada locus polimórfico.

En una realización de la divulgación, los métodos instantáneos crean un modelo genético para sumar las mediciones de múltiples loci polimórficos, a fin de distinguir mejor la trisomía de la disomía y también para determinar el tipo de trisomía. Por otra parte, el método instantáneo incorpora información del enlace genético para mejorar la precisión del método. Esto contrasta con algunos métodos conocidos en la técnica, donde se establece la media de los ratios de los alelos de todos los loci polimórficos de un cromosoma. El método divulgado en el presente documento establece un modelo explícitamente de las distribuciones de la frecuencia alélica prevista en la disomía, así como la trisomía resultante de la ausencia de disyunción durante la meiosis I, la ausencia de disyunción durante la meiosis II, y la ausencia de disyunción durante la mitosis temprana en el desarrollo fetal. Para ilustrar por qué esto es importante, si no se produjesen cruces, la ausencia de disyunción durante la meiosis I resultaría en una trisomía en la que dos homólogos diferentes se heredaron de un progenitor; la ausencia de disyunción durante la meiosis II o la mitosis temprana en el desarrollo fetal resultaría en dos copias del mismo homólogo de un progenitor. Cada

escenario resulta en diferentes frecuencias alélicas previstas en cada locus polimórfico y también en loci físicamente unidos (es decir, loci en el mismo cromosoma) considerados conjuntamente. Los cruces, que provocan el intercambio de material genético entre homólogos, hacen que el patrón de herencia sea más complejo; sin embargo, el método instantáneo tiene esto en cuenta utilizando información sobre la unión genética, es decir, información sobre la tasa de recombinación, además de la distancia física entre loci. Para distinguir mejor entre la ausencia de disyunción de la meiosis I y la ausencia de disyunción de la meiosis II o la mitosis, el método instantáneo incorpora en el modelo una probabilidad creciente de cruce dado que la distancia desde el centrómero aumenta. La ausencia de disyunción de la meiosis II y la mitosis se puede distinguir por el hecho de que la ausencia de disyunción mitótica típicamente resulta en copias idénticas o prácticamente idénticas de un homólogo, mientras que los dos homólogos presentes tras un evento de ausencia de disyunción de la meiosis II a menudo difieren debido a uno o más cruces durante la gametogénesis.

En una realización, un método de la presente divulgación no puede determinar los haplotipos de los padres si se asume una disomía. En una realización, en caso de trisomía, el método instantáneo puede realizar una determinación sobre los haplotipos de uno de los progenitores o ambos utilizando el hecho de que el plasma toma dos copias de un progenitor, y la información de fase del progenitor se puede determinar analizando qué dos copias se han heredado del progenitor en cuestión. En particular, un niño puede heredar dos de las mismas copias del progenitor (trisomía emparejada) o ambas copias del progenitor (trisomía no emparejada). En cada SNP se puede calcular la probabilidad de la trisomía emparejada y de la trisomía no emparejada. Un método para determinar el estado de ploidía que no utiliza el modelo de enlace que tiene en cuenta los cruces calcularía la probabilidad total de la trisomía como una media simple ponderada de las trisomías emparejadas y no emparejadas de todos los cromosomas. Sin embargo, debido a los mecanismos biológicos que provocan el error de disyunción y cruce, la trisomía puede cambiar de emparejada a no emparejada (y viceversa) en un cromosoma únicamente si se produce un cruce. El método instantáneo probabilísticamente tiene en cuenta la probabilidad de cruce, por lo que produce determinaciones del estado de ploidía más precisas que los métodos que no tienen esto en cuenta.

En una realización de la divulgación, se utiliza un cromosoma de referencia para determinar la fracción del niño y la cantidad del nivel de ruido o distribución de la probabilidad. En una realización, la fracción del niño, el nivel de ruido y/o la distribución de la probabilidad se determina utilizando únicamente la información genética disponible de los cromosomas cuyo estado de ploidía se está determinando. El método instantáneo funciona sin el cromosoma de referencia y también sin establecer la fracción del niño o el nivel de ruido concretos. Esto representa una mejora significativa y un punto de diferenciación con respecto a los métodos conocidos en la técnica, donde los datos genéticos de un cromosoma de referencia son necesarios para calibrar la fracción del niño y el comportamiento del cromosoma.

En una realización de la divulgación donde el cromosoma de referencia no es necesario para determinar la fracción fetal, la determinación de la hipótesis se realiza como sigue:

H* = argmaxH LIK(D|H)*priorprob(H)

Con el algoritmo con el cromosoma de referencia, normalmente se asume que el cromosoma de referencia es una disomía y entonces se puede (a) determinar la fracción del niño más probable y el nivel de ruido aleatorio N basándose en este supuesto y en los datos del cromosoma de referencia

$$[\mathrm{cfr}^*, \mathrm{N}^*] = \underset{\mathrm{cfr,N}}{\operatorname{argmax}} \mathrm{LIK}(\mathrm{D}(\mathrm{ref.chrom})|\mathrm{H11}, \mathrm{cfr}, \mathrm{N})$$

Y a continuación reducir

LIK(D|H) = LIK(D|H,cfr*,N*)

o (b) estimar la distribución de la fracción del niño y el nivel de ruido basándose en este supuesto y en los datos del cromosoma de referencia. En concreto, no se determinaría un solo valor para cfr y N, sino que se asignaría la probabilidad p(cfr, N) para el rango más amplio de posibles valores cfr, N:

p(cfr, N)~LIK(D(ref. chrom)|H11, cfr, N) * priorprob(cfr, N) donde priorprob(cfr, N) es la probabilidad previa de la fracción del niño y el nivel de ruido concretos, determinados por el conocimiento y los experimentos anteriores. Si se desea, solo uniformar con respecto al rango de cfr, N. Después se puede escribir:

$$\mathrm{LIK}(\mathrm{D}|\mathrm{H}) = \sum_{\mathrm{cfr,N}} \mathrm{LIK}(\mathrm{D}|\mathrm{H}, \mathrm{cfr}, \mathrm{N}) * \mathrm{p}(\mathrm{cfr}, \mathrm{N})$$

Los dos métodos anteriores dan buenos resultados.

Cabe señalar que en algunos casos el uso de un cromosoma de referencia no resulta recomendable, posible o factible. En este caso, se puede obtener la mejor determinación del estado de ploidía para cada cromosoma por separado. En concreto:

$$\mathrm{LIK(D|H)} = \sum_{\mathrm{cfr,N}} \mathrm{LIK(D|H, cfr, N)} * \mathrm{p(cfr, N|H)}$$

p(cfr,N|H) Se puede determinar, como se ha indicado anteriormente, para cada cromosoma por separado, asumiendo la hipótesis H, no solo para el cromosoma de referencia que asume la disomía. Utilizando este método se pueden mantener fijados los parámetros del ruido y la fracción del niño, se puede fijar uno de los dos parámetros o se pueden mantener ambos en forma probabilística para cada cromosoma y cada hipótesis.

Las mediciones de ADN son propensas a los ruidos y/o errores, en especial las mediciones cuando la cantidad de ADN es pequeña o cuando el ADN está mezclado con ADN contaminante. Este ruido genera datos genotípicos menos precisos y determinaciones del estado de ploidía menos precisas. En algunas realizaciones, los modelos de las plataformas o algún otro método de modelos de ruido se pueden utilizar para contrarrestar los efectos nocivos del ruido sobre la determinación del estado de ploidía. El método instantáneo utiliza un modelo conjunto de ambos canales, que tiene en cuenta el ruido aleatorio debido a la cantidad de ADN de entrada, la calidad del ADN y/o la calidad del protocolo.

Esto contrasta con algunos métodos conocidos en la técnica donde las determinaciones del estado de ploidía se realizan utilizando el ratio de intensidades alélicas en un locus. Este método excluye los modelos de ruido de SNP precisos. En concreto, por lo general los errores de las mediciones no dependen específicamente del ratio de intensidad del canal medido, lo que reduce el modelo al uso de la información unidimensional. La elaboración de modelos precisos del ruido, la calidad del canal y la interacción del canal requiere un modelo conjunto bidimensional, que no se puede conseguir utilizando ratios alélicos.

En concreto, la proyección de la información de dos canales para el ratio r donde f(x,y) is r = x/y, no resulta recomendable para los modelos precisos de ruido del canal y sesgo. El ruido de un SNP concreto no es una función del ratio, es decir ruido(x,y) ≠ f(x,y) sino que se trata de hecho de una función conjunta de ambos canales. Por ejemplo, m en el modelo binomial, el ruido del ratio medido tiene una varianza de r(1-r)/(x+y) que no es una función puramente de r. En este modelo, cuando se incluye cualquier ruido o sesgo del canal, se supone que en SNP i, el valor X del canal observado es x=aiX+bi, donde X es el valor verdadero del canal, bi es el sesgo del canal y el ruido aleatorio adicionales. De forma similar, se supone que y=ciY+di. El ratio observado r=x/y no puede predecir de forma precisa el ratio verdadero X/Y ni elaborar un modelo del ruido restante, dado que (aiX+bi)/(ciY+di) no es una función de X/Y.

El método divulgado en el presente documento describe una forma efectiva de elaborar un modelo del ruido y el sesgo utilizando distribuciones binomiales conjuntas de todos los canales de medición individualmente. Las ecuaciones relevantes se pueden encontrar en otro apartado del presente documento en las secciones que tratan sobre el sesgo constante de SNP, P(bueno) y P(ref]malo), P (mut|mala) que se ajustan efectivamente al comportamiento de SNP. En una realización, un método de la presente divulgación utiliza una distribución Beta-binomial, que evita la práctica restrictiva de confiar exclusivamente en los ratios de los alelos, sino que elabora modelos del comportamiento basándose en los recuentos de ambos canales.

En una realización, un método divulgado en el presente documento puede determinar el estado de ploidía de un feto en gestación a partir de los datos genéticos que se encuentran en el plasma materno utilizando todas las mediciones disponibles. En una realización, un método divulgado en el presente documento puede determinar el estado de ploidía de un feto en gestación a partir de los datos genéticos que se encuentran en el plasma materno utilizando las mediciones de solo un subconjunto de contextos parentales. Algunos métodos conocidos en la técnica solamente utilizan datos genéticos medidos en los que el contexto parental procede del contexto AA|BB, es decir, cuando ambos progenitores son homocigotos en un determinado locus, pero para un alelo diferente. Un problema de este método es que una pequeña proporción de loci polimórficos proceden del contexto AA|BB, típicamente menos del 10%. En una realización de un método divulgado en el presente documento, el método no utiliza mediciones genéticas del plasma materno realizadas en loci en los que el contexto parental es AA|BB. En una realización, el método instantáneo utiliza mediciones de plasma solo para aquellos loci polimórficos con el contexto parental AA|AB, AB|AA, y AB|AB.

Algunos métodos conocidos en la técnica implican la determinación de la media de los ratios de los alelos de los SNP en el contexto AA|BB, donde los genotipos de ambos progenitores se encuentran presentes, y reivindican la determinación del estado de ploidía a partir del ratio medio de los alelos en estos SNP. Este método presenta una imprecisión significativa debido al comportamiento diferencial de los SNP. Cabe señalar que este método asume que se conocen los genotipos de ambos progenitores. Por el contrario, en algunas realizaciones, el método instantáneo utiliza un modelo de distribución de canales conjunto que no asume la presencia de ninguno de los progenitores ni asume el comportamiento uniforme de los SNP. En algunas realizaciones, el método instantáneo tiene en cuenta la diferente ponderación/comportamiento de los SNP. En algunas realizaciones, el método instantáneo no requiere el conocimiento de uno o ambos genotipos parentales. Un ejemplo de cómo puede conseguir esto el método instantáneo es el siguiente:

En algunas realizaciones, la probabilidad logarítmica de una hipótesis se puede determinar para cada SNP. En un SNP i concreto, asumiendo la hipótesis de ploidía fetal H y un porcentaje de ADN fetal cf, la probabilidad logarítmica de los datos observados D se define como:

$$\mathrm{LIK(D|H,i)} = \log \mathrm{P(D|H,cf,i)} = \log\left(\sum_{\mathrm{m,f,c}} \mathrm{P(D|m,f,c,H,cf,i)P(c|m,f,H)P(m|i)P(f|i)}\right)$$

donde m son posibles genotipos verdaderos de la madre, f son posibles genotipos verdaderos del padre, donde m,f c {AA,AB,BB}, y donde c son posibles genotipos del niño dada la hipótesis H. En particular, para la monosomía c {A, B}, para la disomía c ϵ {AA,AB, BB}, para la trisomía c ϵ {AAA, AAB, ABB, BBB}. Cabe señalar que la inclusión de los datos genotípicos parentales típicamente proporciona determinaciones de la ploidía más precisas; sin embargo, los datos genotípicos parentales no son necesarios para que el método instantáneo funcione bien.

Algunos métodos conocidos en la técnica implican la determinación de la media de los ratios de alelos de los SNP en los que la madre es homocigota, pero un alelo diferente se mide en el plasma (contextos AA|AB o AA|BB) y reivindican determinar los estados del ploidía a partir del ratio de alelos medio de estos SNP. Este método está pensado para los casos en los que el genotipo parental no está disponible. Cabe señalar que es cuestionable la precisión con la que se puede reivindicar que el plasma es heterocigoto en un SNP concreto sin la presencia de un padre homocigoto y opuesto BB: para los casos con una escasa fracción del niño, lo que parece la presencia de un alelo B podría ser simplemente la presencia de ruido; por otra parte, lo que parece la ausencia de B podría ser una simple pérdida de alelos de las mediciones fetales. Incluso en el caso de que se pueda determinar realmente la heterocigosidad del plasma, este método no será capaz de distinguir trisomías paternas. En concreto, para los SNP en los que la madre es AA, y donde algún B se mide en el plasma, si el padre es GG, el genotipo del niño resultante es AGG, lo que produce un ratio medio del 33% de A (para una fracción del niño=100%). Sin embargo, en el caso en el que el padre es AG, el genotipo del niño resultante podría ser AGG para la trisomía emparejada, contribuyendo al ratio del 33% de A, o AAG para la trisomía no emparejada, arrastrando el ratio medio más hacia un 66% de A. Dado que muchas trisomías se encuentran en los cromosomas con cruces, el cromosoma total puede encontrarse en cualquier estado entre la ausencia de trisomía no emparejada y todas las trisomías no emparejadas, y este ratio puede variar entre el 33 y el 66%. Para una disomía plana, el ratio se debería encontrar en torno al 50%. Sin el uso de un modelo de enlace o un modelo de error preciso de la media, este método no detectaría muchos casos de trisomía parental. Por el contrario, el método divulgado en el presente documento asigna probabilidades del genotipo parental para cada candidato genotípico parental, basándose en la información genotípica disponible y la frecuencia de la población, y no requiere explícitamente genotipos parentales.

Por otra parte, el método divulgado en el presente documento es capaz de detectar la trisomía incluso en ausencia o presencia de datos genotípicos parentales y puede compensar identificando los puntos de posibles cruces de trisomía emparejada a no emparejada utilizando un modelo de enlace.

Algunos métodos conocidos en la técnica reivindican un método para determinar la media de los ratios alélicos de los SNP cuando no se conoce el genotipo materno ni paterno, y para determinar el estado de ploidía a partir del ratio medio en estos SNP. Sin embargo, no se divulga un método para lograr este fin. El método divulgado en el presente documento es capaz de realizar determinaciones precisas del estado de ploidía en esta situación, y la reducción a la práctica se divulga en otro apartado del documento, utilizando un método de la probabilidad máxima conjunto y opcionalmente utiliza modelos de sesgo y ruido de los SNP, así como un modelo de enlace.

Algunos métodos conocidos en la técnica implican la estimación de la media de los ratios alélicos y reivindican la determinación de los estados de ploidía a partir del ratio alélico medio en uno o unos cuantos SNP. Sin embargo, estos métodos no utilizan el concepto del enlace. Los métodos divulgados en el presente documento no presentan estas desventajas.

*Utilización de la longitud de la secuencia como dato previo para determinar el origen del ADN*

Se ha determinado que la distribución de la longitud de las secuencias difiere entre el ADN materno y fetal, siendo por lo general más cortas en el fetal. En una realización de la presente divulgación, se puede utilizar el conocimiento previo en forma de datos empíricos y construir una distribución previa para la longitud prevista del ADN tanto materno (P(X| materno)) como fetal (P(X| fetal)). Dada una nueva secuencia de ADN no identificada de longitud x, se puede asignar una probabilidad de que una determinada secuencia de ADN sea ADN materno o fetal, basándose en la probabilidad previa de que x sea materna o fetal. En concreto, si P(x|materna) > P(x|fetal), entonces la secuencia de ADN se puede clasificar como materna, con P(x|materna) = P(x|materna)/[(P(x|materna) + P(x| fetal)], y si p(x|materna) < p(x|fetal), entonces la secuencia de DNA se puede clasificar como fetal, P(x| fetal) = P(x| fetal)/[(P(x|materna) + P(x| fetal)]. En una realización de la presente divulgación, una distribución de las longitudes de secuencia materna y fetal se puede determinar que es específica para esa muestra considerando las secuencias que se pueden clasificar como maternas o fetales con elevada probabilidad y, después, esa distribución específica de la muestra se puede utilizar como la distribución del tamaño prevista para esa muestra.

*Profundidad de lectura variable para minimizar el coste de la secuenciación*

En múltiples ensayos clínicos relativos al diagnóstico, por ejemplo, en Chiu et al. BMJ 2011;342:c7401, se establece un protocolo con un número de parámetros y, a continuación, el mismo protocolo se ejecuta con los mismos parámetros para cada uno de los pacientes del ensayo. En el caso de la determinación del estado de ploidía de un feto en gestación en una madre utilizando la secuenciación como método para medir el material genético un parámetro pertinente es el número de lecturas. El número de lecturas se puede referir al número de lecturas reales,

el número de lecturas previstas, líneas parciales, líneas completas o células del flujo completas en un secuenciador. En estos estudios, el número de lecturas se fija típicamente a un nivel que garantizará que todas o prácticamente todas las muestras alcancen el nivel de precisión deseado. En estos momentos la secuenciación es una tecnología cara, con un coste de unos 200 dólares por cinco millones de lecturas correlacionables, y a pesar de que el precio está bajando, cualquier método que permita un diagnóstico basado en la secuenciación para operar con un nivel de precisión similar pero menos lecturas supondrá necesariamente un ahorro considerable de dinero.

La precisión de la determinación del estado de ploidía depende típicamente de una serie de factores, incluyendo el número de lecturas y la fracción del ADN fetal en la mezcla. La precisión es típicamente superior cuando la fracción de ADN fetal de la mezcla es superior. Al mismo tiempo, la precisión es típicamente superior si el número de lecturas es mayor. Es posible tener una situación con dos casos en la que el estado de ploidía se determine con precisiones comparables, donde el primer caso tiene una fracción inferior de ADN fetal en la mezcla que el segundo, y se secuenciaron más lecturas en el primer caso que en el segundo. Se puede utilizar la fracción estimada de ADN fetal de la mezcla como guía para determinar el número de lecturas necesarias para conseguir un determinado nivel de precisión.

En una realización de la presente divulgación, se puede utilizar un conjunto de muestras donde las diferentes muestras del conjunto se secuencian a diferentes profundidades de lectura, donde el número de lecturas ejecutado en cada una de las muestras se selecciona para conseguir un determinado nivel de precisión teniendo en cuenta la fracción calculada de ADN fetal en cada mezcla. En una realización de la presente divulgación, esto puede implicar realizar una medición de la muestra combinada para determinar la fracción de ADN fetal en la mezcla; esta estimación de la fracción fetal se puede realizar con secuenciación, se puede realizar con TaqMan, se puede realizar con qPCR, se puede realizar con arrays de SNP, se puede realizar con cualquier método que pueda distinguir diferentes alelos en un determinado locus. La necesidad de la estimación de la fracción fetal se puede eliminar incluyendo hipótesis que cubran todas las fracciones fetales o un conjunto de fracciones fetales del conjunto de hipótesis consideradas cuando se comparan con los datos medidos reales. Una vez que se ha determinado la fracción de ADN fetal de la muestra, se puede determinar el número de secuencias a leer para cada muestra.

En una realización de la presente divulgación, 100 mujeres embarazadas visitan a sus respectivos obstetras; se les extrae sangre que se pasa a continuación a tubos de análisis con un anti-lisante y/o algún elemento para inactivar la ADNasa. Se llevan a casa un kit para el padre de su feto en gestación que proporciona una muestra de saliva. Ambos conjuntos de materiales genéticos de las 100 parejas se envían al laboratorio, donde la sangre materna es agitada y la capa leucocitaria se aísla, así como el plasma. El plasma comprende una mezcla de ADN materno y ADN de origen placentario. Se determina el genotipo de la capa leucocitaria materna y la sangre paterna utilizando un array de SNP, y el ADN de las muestras de plasma materno es focalizado con sondas de hibridación SURESELECT. El ADN extraído con las sondas se utiliza para generar 100 bibliotecas etiquetadas, una para cada una de las muestras maternas, donde cada muestra es etiquetada con una etiqueta diferente. Se extrae una fracción de cada biblioteca, cada una de esas fracciones se mezcla y añade a dos líneas de un secuenciador de ADN ILLUMINA HISEQ de forma multiplexada, donde cada línea da como resultado unos 50 millones de lecturas correlacionables, lo que proporciona unos 100 millones de lecturas correlacionables sobre las 100 mezclas multiplexadas o aproximadamente un millón de lecturas por muestra. Las lecturas de secuencia se utilizaron para determinar la fracción de ADN fetal de cada mezcla. 50 de las muestras tenían más de un 15% de ADN fetal en la mezcla y un millón de lecturas fue suficiente para determinar el estado de ploidía de los fetos con una certeza del 99,9%.

De las mezclas restantes, 25 tenían entre un 10 y un 15% de ADN fetal; una fracción de cada una de las bibliotecas relevantes preparada a partir de estas mezclas se multiplexó y se procesó en la línea del HISEQ generando otros dos millones de lecturas para cada muestra. Los dos conjuntos de datos de secuencias para cada una de las mezclas con un 10-15% de ADN fetal se juntaron y los tres millones de lecturas por muestra resultantes fueron suficientes para determinar el estado de ploidía de esos fetos con una certeza del 99,9%.

De las mezclas restantes, 13 tenían entre un 6 y un 10% de ADN fetal; una fracción de cada una de las bibliotecas relevantes preparada a partir de estas mezclas se multiplexó y se procesó en la línea del HISEQ generando otros cuatro millones de lecturas para cada muestra. Los dos conjuntos de datos de secuencias para cada una de las mezclas con un 6-10% de ADN fetal se juntaron y los cinco millones de lecturas totales por mezcla resultantes fueron suficientes para determinar el estado de ploidía de esos fetos con una certeza del 99,9%.

De las mezclas restantes, 8 tenían entre un 4 y un 6% de ADN fetal; una fracción de cada una de las bibliotecas relevantes preparada a partir de estas mezclas se multiplexó y se procesó en la línea del HISEQ generando otros seis millones de lecturas para cada muestra. Los dos conjuntos de datos de secuencias para cada una de las mezclas con un 4-6% de ADN fetal se juntaron y los siete millones de lecturas totales por mezcla resultantes fueron suficientes para determinar el estado de ploidía de esos fetos con una certeza del 99,9%.

De las cuatro mezclas restantes, todas ellas tenían entre un 2 y un 4% de ADN fetal; una fracción de cada una de las bibliotecas relevantes preparada a partir de estas mezclas se multiplexó y se procesó en la línea del HISEQ generando otros doce millones de lecturas para cada muestra. Los dos conjuntos de datos de secuencias para cada una de las mezclas con un 2-4% de ADN fetal se juntaron y los 13 millones de lecturas totales por mezcla resultantes fueron suficientes para determinar el estado de ploidía de esos fetos con una certeza del 99,9%.

Este método precisó seis líneas de secuenciación en una máquina HISEQ para conseguir una precisión del 99,9% con las 100 muestras. Si se hubiese requerido el mismo número de procesamientos para cada muestra, para garantizar una determinación del estado de ploidía con una precisión del 99,9% se habrían necesitado 25 líneas de secuenciación, y si se hubiese tolerado una tasa de no determinación o una tasa de error del 4%, se habrían podido obtener 14 líneas de secuenciación.

*Uso de datos del genotipo en bruto*

Existen varios métodos que pueden realizar un NPD utilizando información genética fetal medida en ADN fetal que se encuentra en la sangre materna. Algunos de estos métodos implican la realización de mediciones del ADN fetal utilizando arrays de SNP, algunos métodos implican la secuenciación no focalizada y algunos métodos implican la secuenciación focalizada. La secuenciación focalizada puede centrarse en los SNP, puede centrarse en STR, puede centrarse en otros loci polimórficos, puede centrarse en loci no polimórficos o en alguna combinación de estos. Algunos de estos métodos pueden implicar el uso de un instrumento de determinación alélica propio o comercializado que determine la identidad de los alelos a partir de los datos de intensidad procedentes de los sensores de la máquina que realiza la medición. Por ejemplo, el sistema ILLUMTNA INFINIUM o el sistema de microarrays AFFYMETRIX GENECHIP utiliza perlas o microchips con secuencias de ADN que se pueden hibridar con segmentos complementarios de ADN; tras la hibridación, se produce un cambio en las propiedades fluorescentes de la molécula sensora que puede ser detectada. También existen métodos de secuenciación, por ejemplo, ILLUMINA SOLEXA GENOME SEQUENCER o ABI SOLID GENOME SEQUENCER, donde la secuencia genética de fragmentos de ADN es secuenciada; tras la ampliación de la cadena de ADN complementaria a la cadena que se va a secuenciar, la identidad del nucleótido ampliado se detecta típicamente a través de una etiqueta fluorescente o radioetiqueta unida al nucleótido complementario. En todos estos métodos los datos genotípicos o de secuenciación se determinan típicamente sobre la base de señales fluorescentes o de otro tipo, o de la ausencia de estas. Estos sistemas se combinan típicamente con paquetes de software de bajo nivel que realizan determinaciones alélicas específicas (datos genéticos secundarios) a partir de los resultados analógicos del dispositivo de detección de fluorescencia o de otro tipo (datos genéticos primarios). Por ejemplo, en el caso de un alelo determinado en un array de SNP, el software determinará, por ejemplo, que un determinado SNP se encuentra presente o no presente si la intensidad de la fluorescencia está por encima o por debajo de determinado umbral. De forma similar, el producto de un secuenciador es un cromatograma que indica el nivel de fluorescencia detectado para cada uno de los colorantes, y el software determinará que un determinado par de bases es A o T o C o G.

Los secuenciadores de alto rendimiento típicamente realizan una serie de mediciones de este tipo, denominadas lecturas, que representan la estructura más probable de la secuencia de ADN que se ha secuenciado. El producto análogo directo del cromatograma se define aquí como datos genéticos primarios, y la determinación del par de bases/SNP realizada por el software se considera en el presente documento los datos genéticos secundarios. En una realización, por datos primarios se entiende los datos de intensidad bruta que son el producto no procesado de una plataforma de determinación del genotipo, donde la plataforma de determinación del genotipo se puede referir a un array de SNP o a una plataforma de secuenciación. Los datos genéticos secundarios se refieren a los datos genéticos procesados, donde se ha realizado una determinación alélica, o se han asignado pares de bases a los datos de la secuencia, y/o las lecturas de secuencia se han correlacionado con el genoma.

Muchas aplicaciones de nivel superior aprovechan estas determinaciones alélicas, determinaciones de los SNP y lecturas de secuencia, es decir, los datos genéticos secundarios que produce el software de determinación del genotipo. Por ejemplo, DNA NEXUS, ELAND o MAQ realizarán las lecturas de secuenciación y las correlacionarán con el genoma. Por ejemplo, en el contexto del diagnóstico prenatal no invasivo, un sistema informático complejo, como PARENTAL SUPPORT™, puede aprovechar un gran número de determinaciones de SNP para determinar el genotipo de un individuo. Por otra parte, en el contexto del diagnóstico genético previo al implante, se puede tomar un conjunto de lecturas de secuencia correlacionadas con el genoma y realizando un recuento normalizado de las lecturas que están correlacionadas con cada cromosoma, o con una sección de un cromosoma, se puede determinar el estado de ploidía de un individuo. En el contexto del diagnóstico prenatal no invasivo, se puede tomar un conjunto de lecturas de secuencia que se han medido en el ADN presente en el plasma materno y correlacionarlas con el genoma. A continuación, se puede realizar un recuento normalizado de las lecturas correlacionadas con cada cromosoma, o sección de un cromosoma, y utilizar esos datos para determinar el estado de ploidía de un individuo. Por ejemplo, se puede concluir que los cromosomas que tienen un número desproporcionadamente grande de lecturas son trisómicos en el feto que se está gestando en la madre a la que se le ha extraído la sangre.

Sin embargo, en realidad el producto inicial de los instrumentos de medición es una señal analógica. Cuando se determina un determinado par de bases a través del software asociado con el software de secuenciación, por ejemplo, el software puede determinar el par de bases T, en realidad, la determinación es la que el software considera más probable. Sin embargo, en algunos casos la determinación puede ser de baja certeza, por ejemplo, la señal analógica puede indicar que el par de bases concreto tiene solo un 90% de probabilidad de ser T y un 10% de probabilidad de ser A. En otro ejemplo, el software de determinación del genotipo que está asociado con un lector del array SNP puede determinar que un cierto alelo es G. Sin embargo, en realidad la señal analógica subyacente puede indicar que tiene solo un 70% de probabilidad de ser G y un 30% de probabilidad de ser T. En estos casos, cuando las aplicaciones de nivel superior utilizan las determinaciones del genotipo y las determinaciones de la secuencia realizadas por el software de nivel inferior, pierden cierta información. Es decir, los datos genéticos primarios, medidos directamente por la plataforma de determinación del genotipo, pueden ser menos fiables que los

datos genéticos secundarios determinados por los paquetes de software adjuntos, aunque contienen más información. Para correlacionar las secuencias de datos genéticos secundarios con el genoma, muchas lecturas se descartan porque algunas bases no se leen con suficiente claridad o porque la correlación no está clara. Cuando se utilizan lecturas de secuencia de datos genéticos primarios, todas o muchas de esas lecturas que se han descartado cuando se convirtieron por primera vez en lecturas de secuencia de datos genéticos secundarios se pueden utilizar tratando las lecturas de manera probabilística.

En una realización de la presente divulgación, el software de nivel superior no se basa en las determinaciones alélicas, las determinaciones del SNP o las lecturas de secuencia determinadas por el software de nivel inferior. En vez de esto, el software de nivel superior basa sus cálculos en las señales analógicas medidas directamente por la plataforma de determinación del genotipo. En una realización de la presente divulgación, un método basado en la informática como PARENTAL SUPPORT™ es modificado para que su capacidad para reconstruir datos genéticos del embrión/feto/niño sea organizada para utilizar directamente los datos genéticos primarios medidos por la plataforma de determinación del genotipo. En una realización de la presente divulgación, un método basado en la informática como PARENTAL SUPPORT™ puede realizar determinaciones alélicas y/o determinar el número de copias del cromosoma utilizando datos genéticos primarios y sin usar datos genéticos secundarios. En una realización de la presente divulgación, todas las determinaciones genéticas, determinaciones de SNP, lecturas de secuencia, correlaciones de secuencia se tratadas de manera probabilística utilizando los datos de intensidad brutos medidos directamente por la plataforma de determinación del genotipo, en lugar de convertir los datos genéticos primarios en determinaciones genéticas secundarias. En una realización, las mediciones de ADN de la muestra preparada utilizada para calcular las probabilidades del recuento de alelos y determinar la probabilidad relativa de cada hipótesis comprenden datos genéticos primarios.

En algunas realizaciones, el método puede aumentar la precisión de los datos genéticos de un individuo diana que incorporan los datos genéticos de al menos un individuo relacionado, donde el método comprende la obtención de datos genéticos primarios específicos del genoma de un individuo diana y datos genéticos específicos del genoma o genomas del individuo o individuos relacionados, la creación de un conjunto de una o más hipótesis sobre qué segmentos de qué cromosomas del individuo o individuos relacionados corresponden posiblemente a esos segmentos en el genoma del individuo diana, la determinación de la probabilidad de cada una de las hipótesis teniendo en cuenta los datos genéticos primarios del individuo diana y los datos genéticos del individuo o individuos relacionados, y la utilización de las probabilidades asociadas con cada hipótesis para determinar el estado más probable del material genético real del individuo diana. En algunas realizaciones, el método puede determinar el número de copias de un segmento de un cromosoma del genoma de un individuo diana, donde el método comprende la creación de un conjunto de hipótesis del número de copias sobre cuántas copias del segmento del cromosoma se encuentran presentes en el genoma de un individuo diana, la incorporación de datos genéticos primarios del individuo diana e información genética de uno o más individuos relacionados al conjunto de datos, la estimación de las características de la respuesta de la plataforma asociada con el conjunto de datos, donde la respuesta de la plataforma puede variar de un experimento a otro, la computación de las probabilidades condicionales de cada hipótesis del número de copias, teniendo en cuenta el conjunto de datos y las características de la respuesta de la plataforma, y la determinación del número de copias del segmento del cromosoma basándose en la hipótesis del número de copias más probable. En una realización, un método de la presente divulgación puede determinar un estado de ploidía de al menos un cromosoma en un individuo diana, donde el método comprende la obtención de datos genéticos primarios del individuo diana y de uno o más individuos relacionados, la creación de un conjunto de al menos una hipótesis del estado de ploidía para cada uno de los cromosomas del individuo diana, la utilización de una o más técnicas especializadas para determinar una probabilidad estadística para cada hipótesis del estado de ploidía del conjunto, para cada técnica especializada utilizada, teniendo en cuenta los datos genéticos obtenidos, la combinación para cada hipótesis del estado de ploidía de las probabilidades estadísticas determinadas mediante la técnica o técnicas especializadas, y la determinación del estado de ploidía para cada uno de los cromosomas del individuo diana basándose en las probabilidades estadísticas combinadas de cada una de las hipótesis del estado de ploidía. En una realización, un método de la presente divulgación puede determinar un estado alélico de un conjunto de alelos en un individuo diana y a partir de uno o los dos progenitores del individuo diana, y opcionalmente de uno o más individuos relacionados, donde el método comprende la obtención de datos genéticos primarios del individuo diana, y de uno o los dos progenitores, y de cualquier individuo relacionado, la creación de un conjunto de al menos una hipótesis alélica para el individuo diana, y para uno o los dos progenitores, y opcionalmente para uno o más individuos relacionados, donde las hipótesis describen los posibles estados alélicos del conjunto de alelos, determinando una probabilidad estadística para cada hipótesis alélica del conjunto de hipótesis, teniendo en cuenta los datos genéticos obtenidos, y la determinación del estado alélico para cada uno de los alelos del conjunto de alelos para el individuo diana, y para uno o los dos progenitores, y opcionalmente para uno o más individuos relacionados, basándose en las probabilidades estadísticas de cada una de las hipótesis alélicas.

En algunas realizaciones de la divulgación, los datos genéticos de la muestra combinada pueden comprender datos de secuencia donde los datos de secuencia pueden no corresponder de forma exclusiva al genoma humano. En algunas realizaciones, los datos genéticos de la muestra combinada pueden comprender datos de secuencia donde los datos de secuencia corresponden a una pluralidad de ubicaciones en el genoma, donde cada posible correlación está asociada con una probabilidad de que la correlación en cuestión es correcta. En algunas realizaciones, no se asume que las lecturas de secuencia están asociadas con una posición concreta en el genoma. En algunas

realizaciones, las lecturas de secuencia están asociadas con una pluralidad de posiciones en el genoma y una probabilidad asociada perteneciente a esa posición.

*Combinación de métodos para el diagnóstico prenatal*

Hay muchos métodos que se pueden utilizar para el diagnóstico prenatal y la detección prenatal de la aneuploidía u otros defectos genéticos. En otra sección de este documento y en la Solicitud de Patente USA con el número de serie 11/603 406, presentada el 28 de noviembre de 2006; la Solicitud de Patente USA con el número de serie 12/076 348, presentada el 17 de marzo de 2008, y la Solicitud de Patente PCT con el número de serie PCT/S09/52730 se describe uno de estos métodos que utiliza los datos genéticos de individuos relacionados para aumentar la precisión del conocimiento o la estimación de los datos genéticos de un individuo diana, como un feto. Otros métodos utilizados para el diagnóstico prenatal implican la medición de los niveles de determinadas hormonas en la sangre materna, donde estas hormonas están correlacionadas con diversas anomalías genéticas. Un ejemplo de esto se denomina la triple prueba, una triple prueba en la que se miden los niveles de varias (normalmente dos, tres, cuatro o cinco) hormonas diferentes en la sangre materna. En un caso en el que se utilizan múltiples métodos para determinar la probabilidad de un determinado resultado, donde ninguno de los métodos es definitivo por sí solo, se puede combinar la información proporcionada por estos métodos para realizar una predicción que es más precisa que cualquiera de los métodos individuales. En la triple prueba, la combinación de la información proporcionada por las tres hormonas diferentes puede proporcionar una predicción de anomalías genéticas que es más precisa que la efectuada por los niveles de hormonas individuales.

En el presente documento se divulga un método para realizar predicciones más precisas sobre el estado genético de un feto, concretamente la posibilidad de anomalías genéticas en un feto, que comprende la combinación de predicciones de anomalías genéticas en un feto, donde estas predicciones se realizaron utilizando diversos métodos. Un método "más preciso" se puede referir a un método para el diagnóstico de una anomalía que tiene una tasa de falsos negativos inferior para una determinada tasa de falsos positivos. En una realización preferible de la presente divulgación, una o más de las predicciones se realizan basándose en los datos genéticos conocidos sobre el feto, donde el conocimiento genético se determinó utilizando el método PARENTAL SUPPORT™, es decir, utilizando datos genéticos del individuo relacionado con el feto para determinar los datos genéticos del feto con una mayor precisión. En algunas realizaciones, los datos genéticos pueden incluir estados de ploidía del feto. En algunas realizaciones, los datos genéticos se pueden referir a un conjunto de determinaciones alélicas sobre el genoma del feto. En algunas realizaciones, algunas de las predicciones se pueden haber realizado utilizando la triple prueba. En algunas de las realizaciones, algunas de las predicciones se pueden haber realizado utilizando mediciones de otros niveles de hormonas en la sangre materna. En algunas realizaciones, las predicciones realizadas mediante métodos considerados diagnósticos se pueden combinar con predicciones realizadas mediante métodos considerados de detección. En algunas realizaciones, el método implica la medición de los niveles en sangre materna de alfa-fetoproteína (AFP). En algunas realizaciones, el método implica la medición de los niveles en sangre materna de estriol no conjugado (UE3). En algunas realizaciones, el método implica la medición de los niveles en sangre materna de gonadotropina coriónica humana beta (beta-hCG). En algunas realizaciones, el método implica la medición de los niveles en sangre materna de antígeno trofoblástico invasivo (ITA). En algunas realizaciones, el método implica la medición de los niveles en sangre materna de inhibina. En algunas realizaciones, el método implica la medición de los niveles en sangre materna de proteína A plasmática asociada el embarazo (PAPP-A). En algunas realizaciones, el método implica la medición de los niveles en sangre materna de otras hormonas o marcadores del suero materno. En algunas de las realizaciones, algunas de las predicciones se pueden haber realizado utilizando otros métodos. En algunas realizaciones, algunas de las predicciones se pueden haber realizado utilizando una prueba plenamente integrada, como una que combina los ultrasonidos y el análisis de sangre a las 12 semanas aproximadas de embarazo y un segundo análisis de sangre sobre las 16 semanas de embarazo. En algunas realizaciones, el método implica la medición de la translucencia nucal (NT) fetal. En algunas realizaciones, el método implica el uso de los niveles medidos de las mencionadas hormonas para realizar predicciones. En algunas realizaciones, el método implica una combinación de los métodos mencionados.

Existen muchas formas de combinar las predicciones, por ejemplo, se podrían convertir las mediciones de las hormonas en un múltiplo de la mediana (MoM) y a continuación en ratios de probabilidad (LR). De forma similar, otras mediciones se podrían transformar en LR utilizando el modelo de la mezcla de las distribuciones de NT. Los LR para NET y los marcadores bioquímicos se podrían multiplicar por la edad y el riesgo relacionado con la gestación para obtener el riesgo para las diversas condiciones, como la trisomía 21. Las tasas de detección (DR) y las tasas de falsos positivos (FPR) se podrían calcular tomando las proporciones con los riesgos anteriores que superan un determinado umbral de riesgo.

En una realización de la divulgación, un método para determinar el estado de ploidía implica la combinación de las probabilidades relativas de cada una de las hipótesis de ploidía determinadas utilizando el modelo de distribución conjunto y las probabilidades del recuento alélico con las probabilidades relativas de cada una de las hipótesis de ploidía que se calculan utilizando técnicas estadísticas tomadas de otros métodos que determinan una puntuación de riesgo de que un feto sea trisómico, incluyendo, entre otros: un análisis del recuento de lecturas, la comparación de las tasas de heterocigosidad, una estadística que solo está disponible cuando se utiliza la información genética parental, la probabilidad de señales normalizadas del genotipo para determinados contextos parentales, una estadística que se calcula utilizando la fracción fetal estimada de la primera muestra o de la muestra preparada y combinaciones de estos.

Otro método podría implicar una situación con cuatro niveles de hormonas medidos, donde la distribución de la probabilidad con respecto a esas hormonas es conocida: p(x1, x2, x3, x4|e) para el caso euploide y p(x1, x2, x3, x4|a) para el caso aneuploide. A continuación, se podría medir la distribución de la probabilidad para las mediciones del ADN, g(y|e) y g(y|a) para el caso euploide y el caso aneuploide, respectivamente. Suponiendo que son independientes teniendo en cuenta el supuesto de euploide/aneuploide, se podrían combinar como p(x1, x2, x3, x4|a)g(y|a) y p(x1, x2, x3, x4|e)g(y|e) y, a continuación, multiplicar cada uno por el dato previo p(a) y p(e) teniendo en cuenta la edad materna. Se podría elegir el más elevado.

En una realización de la divulgación, se puede recurrir al teorema del límite central para asumir que la distribución en g(y|a o e) es gaussiana, y medir la desviación media y estándar analizando múltiples muestras. En otra realización, se podría asumir que no son independientes teniendo en cuenta el resultado y recopilar suficientes muestras para estimar la distribución conjunta p(x1, x2, x3, x4|a o e).

En una realización de la divulgación, el estado de ploidía para el individuo diana se determina que es el estado de ploidía asociado con la hipótesis con la probabilidad más elevada. En algunos casos, una hipótesis tendrá una probabilidad normalizada combinada superior al 90%. Cada hipótesis está asociada con uno o varios estados de ploidía, y el estado de ploidía asociado con la hipótesis cuya probabilidad normalizada combinada sea superior al 90%, o algún otro valor umbral, como el 50%, 80%, 95%, 98%, 99% o 99,9% podrá ser seleccionado como el umbral requerido para establecer que una hipótesis es el estado de ploidía determinado.

*ADN de hijos de embarazos anteriores en la sangre materna*

Una dificultad del diagnóstico prenatal no invasivo consiste en diferenciar las células fetales del embarazo actual de las células fetales de embarazos anteriores. Algunos piensan que el material genético de los embarazos anteriores desaparecerá tras algún tiempo, pero no se han obtenido pruebas concluyentes. En una realización de la presente divulgación, se puede determinar el ADN fetal presente en la sangre materna de origen paterno (es decir, ADN del feto heredado del padre) utilizando el método PARENTAL SUPPORT™ (PS) y el conocimiento del genoma paterno. Este método puede utilizar información genética parental por fases. Se puede determinar el genotipo parental por fases a partir de la información genotípica no determinada por fases, utilizando datos genéticos de los abuelos (como los datos genéticos medidos del esperma del abuelo) o los datos genéticos de otro niño nacido o la muestra de un aborto. También se podría obtener información genética por fases a partir de la información genética no determinada por fases mediante HapMap o determinando el haplotipo de las células paternas. La determinación con éxito del haplotipo se ha demostrado deteniendo las células en la fase de mitosis cuando los cromosomas son agrupaciones apretadas y utilizando microfluídicos para poner los cromosomas separados en pocillos separados. En otra realización se pueden utilizar los datos del haplotipo parental por fases para detectar la presencia de más de un homólogo del padre, lo que implica que el material genético de más de un niño se encuentra presente en la sangre. Al centrarse en los cromosomas que se espera que sean euploides en un feto, se podría descartar la posibilidad de que el feto se vea afectado por una trisomía. También se puede determinar si el ADN fetal no procede del padre actual, en cuyo caso se podrían utilizar otros métodos como la triple prueba para predecir anomalías genéticas.

Puede haber otras fuentes de material genético fetal disponible a través de métodos distintos de la extracción de sangre. En el caso del material genético fetal disponible en la sangre materna, existen dos categorías principales: 1) células fetales completas, por ejemplo, glóbulos rojos fetales nucleados o eritroblastos, y 2) ADN fetal flotante libre. En el caso de las células fetales completas, existen algunas evidencias de que las células fetales pueden persistir en la sangre materna durante un periodo de tiempo prolongado de forma que se puede aislar una célula de una mujer embarazada que contenga ADN de un niño o un feto de un embarazo anterior. También existen evidencias de que el ADN fetal flotante se elimina del sistema en el plazo unas semanas. Un problema consiste en determinar la identidad del individuo cuyo material genético se encuentra en la célula, concretamente para garantizar que el material genético medido no corresponda a un feto de un embarazo anterior. En una realización de la presente divulgación, el conocimiento del material genético materno se puede utilizar para garantizar que el material genético en cuestión no es material genético materno. Existen diversos métodos para ello, incluyendo el método basado en la informática PARENTAL SUPPORT™, descrito en este documento o en cualquiera de las patentes mencionadas en el mismo.

En una realización de la presente divulgación, la sangre extraída de la madre embarazada puede ser separada en una fracción que comprende ADN fetal flotante libre y una fracción que comprende glóbulos rojos nucleados. El ADN flotante libre se puede opcionalmente enriquecer, y la información genotípica del ADN se puede medir. A partir de la información genotípica medida del ADN flotante libre, el conocimiento del genotipo materno se puede utilizar para determinar aspectos del genotipo fetal. Estos aspectos se pueden referir al estado de ploidía y/o a las identidades de un conjunto de alelos. A continuación, se puede determinar el genotipo de los glóbulos rojos nucleados utilizando métodos descritos en otro apartado de este documento y en otras patentes mencionadas, especialmente las mencionadas en la primera parte de este documento. El conocimiento del genoma materno permitiría determinar si un glóbulo rojo dado es genéticamente materno o no. Y los aspectos del genotipo fetal determinados como se ha descrito anteriormente permitirían determinar si la célula sanguínea individual procede genéticamente del feto que se está gestando en estos momentos. Básicamente, este aspecto de la presente divulgación permite utilizar el conocimiento genético de la madre y posiblemente la información genética de otros individuos relacionados, como el padre, junto con la información genética medida del ADN flotante libre que se encuentra en la sangre materna para determinar si una célula nucleada aislada que se encuentra en la sangre materna es a) genéticamente materna, b)

genéticamente del feto que se está gestando en estos momentos, o c) genéticamente de un feto de un embarazo anterior.

*Determinación de aneuploidía del cromosoma sexual prenatal*

En métodos conocidos en la técnica, las personas que desean determinar el sexo de un feto en gestación a partir de la sangre materna han utilizado el hecho de que el ADN flotante libre fetal (fffDNA) se encuentra presente en el plasma de la madre. si se pueden detectar los loci específicos de Y en el plasma materno, esto implica que el feto en gestación es un varón. Sin embargo, el hecho de que no se detecten los loci específicos de Y en el plasma no siempre garantiza que el feto en gestación sea una mujer cuando se utilizan métodos conocidos en la técnica anterior, dado que en algunos casos la cantidad de fffDNA es demasiado escasa para garantizar que los loci específicos de Y se vayan a detectar en caso de que el feto sea un varón.

En el presente documento se proporciona un método novedoso que no requiere la medición de ácidos nucleicos específicos de Y, es decir, ADN procedente de loci de origen exclusivamente paterno. El método Parental Support, anteriormente divulgado, utiliza datos de frecuencia de cruces, datos genotípicos parentales y técnicas informáticas para determinar el estado de ploidía de un feto en gestación. El sexo del feto es simplemente el estado de ploidía del feto en los cromosomas sexuales. Un feto que es XX es mujer y XY es varón. El método descrito en el presente documento también se puede utilizar para determinar el estado de ploidía del feto. Cabe señalar que la determinación del sexo es efectivamente sinónimo de la determinación del estado de ploidía de los cromosomas sexuales; en el caso de la determinación del sexo, normalmente se presupone que el niño es euploide y, por tanto, hay menos hipótesis posibles.

El método divulgado en el presente documento implica analizar los loci que son comunes a los cromosomas X e Y para crear una línea de partida en términos de cantidad prevista de ADN fetal presente para un feto. A continuación, las regiones que son específicas solo para el cromosoma X se pueden analizar para determinar si el feto es hembra o varón. En el caso de un varón, esperamos observar menos ADN fetal de loci que son específicos para el cromosoma X que de loci que son específicos tanto para el cromosoma X como para el cromosoma Y. Por el contrario, en los fetos hembra, esperamos que el ADN para cada uno de estos grupos sea el mismo. El ADN en cuestión se puede medir a través de cualquier técnica que pueda cuantificar la cantidad de ADN presente en una muestra, por ejemplo, qPCR, arrays de SNP, arrays para la determinación del genotipo o secuenciación. Para el ADN que es exclusivamente de un individuo esperaríamos observar lo siguiente:

| | ADN específico de X | ADN específico de X e Y | ADN específico de Y |
|---|---|---|---|
| Varón (XY) | A | 2A | A |
| Hembra (XX) | 2A | 2A | 0 |

En el caso del ADN de un feto mezclado con ADN de la madre y donde la fracción de ADN fetal de la mezcla es F y donde la fracción de ADN materno de la mezcla es M, de forma que F+M = 100%, esperaríamos observar lo siguiente:

| | ADN específico de X | ADN específico de X e Y | ADN específico de Y |
|---|---|---|---|
| Feto varón (XY) | m + ½ f | M + F | ½ F |
| Feto hembra (XX) | M + F | M + F | 0 |

En el caso en el que F y M son conocidos, los ratios previstos se pueden computar y los datos observados se pueden comparar con los datos esperados. En el caso en el que F y M no son conocidos, se puede seleccionar un umbral basado en datos históricos. En ambos casos, la cantidad de ADN medida en loci específicos tanto para X como para Y se pueden utilizar como línea de partida, y la prueba para determinar el sexo del feto se puede basar en la cantidad de ADN observada en loci específicos exclusivamente para el cromosoma X. Si esa cantidad es inferior que la línea de partida en una cantidad aproximadamente igual a V2 F o en una cantidad que hace que caiga por debajo de un umbral predefinido, se determina que el feto es varón, y si esa cantidad es aproximadamente igual a la línea de partida o si no es inferior en una cantidad que hace que caiga por debajo de un umbral predefinido, se determina que el feto es mujer.

En otra realización de la divulgación, se pueden analizar solo los loci que son comunes a los cromosomas X e Y, lo que a menudo se denomina el cromosoma Z. Típicamente un subconjunto de los loci del cromosoma Z son siempre A en el cromosoma X y B en el cromosoma Y. Si se descubre que los SNP del cromosoma Z tienen el genotipo B, entonces se dirá que el feto es un varón; si se descubre que los SNP del cromosoma Z solo tienen el genotipo A, se dirá que el feto es una mujer. En otra realización, se pueden analizar los loci que se encuentran únicamente en el

cromosoma X. Contextos como AA|B resultan particularmente informativos, dado que la presencia de una B indica que el feto tiene un cromosoma X del padre. Contextos como AB|B también son informativos, dado que esperamos ver B presente únicamente la mitad de las veces en el caso de un feto hembra en comparación con el feto de un varón. En otra realización, se pueden analizar los SNP del cromosoma Z donde tanto los alelos A como B se encuentran presentes en el cromosoma X y el cromosoma Y, y donde se sabe qué SNP proceden del cromosoma Y paterno y cuáles proceden del cromosoma X paterno.

En una realización de la divulgación, se pueden amplificar posiciones de un único nucleótido que se sabe que varían entre la región homóloga no recombinante (HNR) que comparten el cromosoma Y y el cromosoma X. La secuencia de la región HNR es en gran medida idéntica en los cromosomas X e Y. Dentro de esta región idéntica se encuentran posiciones de un único nucleótido que, a pesar de que no varían entre los cromosomas X ni entre los cromosomas Y de la población, son diferentes en los cromosomas X e Y. Cada ensayo por PCR podrían amplificar una secuencia de los loci que se encuentran presentes tanto en los cromosomas X como Y. Dentro de cada secuencia amplificada se encontraría una única base que se puede detectar utilizando la secuenciación o cualquier otro método.

En una realización de la divulgación, el sexo del feto se podría determinar a partir del ADN flotante libre fetal que se encuentra en el plasma materno, donde el método comprende algunos de los pasos siguientes o todos ellos: 1) Diseñar cebadores PCR (ordinaria o mini-PCR, más multiplexado si se desea) para amplificar las posiciones de un único nucleótido variable X/Y dentro de la región HNR, 2) obtener plasma materno, 3) amplificar por PCR diana del plasma materno utilizando ensayos PCR HRN X/Y, 4) secuenciar los amplicones, 5) examinar los datos de secuencia para determinar la presencia del alelo Y en una o más de las secuencias amplificadas. La presencia de uno o más indicaría un feto varón. La ausencia de todos los alelos Y de los amplicones indica un feto hembra.

En una realización de la divulgación, se podría utilizar la secuenciación focalizada para medir el ADN en el plasma materno y/o los genotipos parentales. En una realización se podrían ignorar todas las secuencias que proceden claramente de ADN de origen paterno. Por ejemplo, en el contexto AA|AB, se podría recontar el número de secuencias A e ignorar todas las secuencias B. Para determinar la tasa de heterocigosidad del anterior algoritmo, se podría comparar el número de secuencias A observadas con el número previsto de secuencias totales para la sonda en cuestión. Hay muchas formas de calcular un número previsto de secuencias para cada sonda y para cada muestra. En una realización, se pueden utilizar datos históricos para determinar qué fracción de todas las lecturas de secuencia pertenece a cada sonda específica y, a continuación, utilizar esta fracción empírica, combinada con el número total de lecturas de secuencia, para estimar el número de secuencias de cada sonda. Otro método sería focalizar algunos alelos homocigotos conocidos y, a continuación, utilizar los datos históricos relativos al número de lecturas en cada sonda con el número de lecturas en los alelos homocigotos conocidos. A continuación, para cada muestra se podría medir el número de lecturas en los alelos homocigotos y después utilizar esta medición, junto con las relaciones obtenidas empíricamente, para estimar el número de lecturas de secuencia en cada sonda.

En algunas realizaciones de la divulgación, se puede determinar el sexo del feto combinando las predicciones realizadas por una pluralidad de métodos. En algunas realizaciones, la pluralidad de métodos se selecciona de los métodos descritos en la presente divulgación. En algunas realizaciones, al menos uno de la pluralidad de métodos se selecciona de los métodos descritos en la presente divulgación.

En algunas realizaciones, el método descrito en el presente documento también se puede utilizar para determinar el estado de ploidía del feto en gestación. En una realización, el método de determinación del estado de ploidía utiliza loci que son específicos para el cromosoma X, o comunes para el cromosoma X y el cromosoma Y, pero no emplea los loci específicos de Y. En una realización, el método de determinación del estado de ploidía utiliza uno o más de los siguientes loci: loci que son específicos del cromosoma X, loci que son comunes para el cromosoma X y el cromosoma Y, y loci que son específicos para el cromosoma Y. En una realización, donde los ratios de cromosomas sexuales son similares, por ejemplo, 45,X (síndrome de Turner), 46,XX (hembra normal) y 47,XXX (trisomía X), la diferenciación se puede realizar comparando las distribuciones alélicas con las distribuciones alélicas previstas en función de las diversas hipótesis. En otra realización, esto se puede conseguir comparando el número relativo de lecturas de secuencia para los cromosomas sexuales con uno o una pluralidad de cromosomas de referencia que se presuponen euploides. Cabe señalar también que estos métodos se pueden ampliar para incluir los casos aneuploides.

*Detección de enfermedades en un solo gen*

En una realización de la divulgación, un método para determinar el estado de ploidía del feto se puede ampliar para permitir la detección simultánea de trastornos en un único gen. El diagnóstico de enfermedades en un único gen utiliza el mismo método focalizado que se emplea para las pruebas de la aneuploidía y requiere dianas específicas adicionales. En una realización, el diagnóstico de NPD en un único gen se realiza a través del análisis del enlace. En muchos casos, las pruebas directas de la muestra de cfDNA no resulta fiable, dado que la presencia de ADN materno hace que resulte prácticamente imposible determinar si el feto ha heredado la mutación de la madre. La detección de un alelo único de origen paterno resulta menos problemática, pero solamente resulta plenamente informativa si la enfermedad es dominante y portada por el padre, lo que limita la utilidad del método. En una realización el método implica la PCR o métodos de amplificación relacionados.

En algunas realizaciones de la divulgación, el método implica la determinación por fases del alelo anómalo con SNP alrededor que presentan enlaces muy firmes en los progenitores, utilizando información de parientes de primer grado. Después se puede ejecutar el método Parental Support en los datos de secuencia focalizados obtenidos de estos SNP para determinar que homólogos, normales o anómalos, han sido heredados por el feto de ambos progenitores. Siempre que los SNP están suficientemente unidos, la herencia del genotipo del feto se puede determinar de manera muy fiable. En algunas realizaciones, el método comprende a) la adición de un conjunto de loci de SNP para flanquear de forma densa un conjunto especificado de enfermedades comunes a nuestro grupo multiplexado para la prueba de la aneuploidía; b) determinación por fases de los alelos de estos SNP añadidos con los alelos normales y anómalos basados en los datos genéticos de diversos parientes; y c) reconstruir el diplotipo fetal, o el conjunto de alelos de SNP por fases en los homólogos maternos y paternos heredados de la región que rodea al locus de la enfermedad para determinar el genotipo fetal. En algunas realizaciones, las sondas adicionales que están estrechamente unidas a un locus vinculado a una enfermedad se añaden al conjunto de loci polimórficos que se van a utilizar para la prueba de la aneuploidía.

La reconstrucción del diplotipo fetal resulta difícil porque la muestra es una mezcla de ADN materno y fetal. En algunas realizaciones, el método incorpora información relativa para determinar las fases de los SNP y los alelos de la enfermedad, a continuación, tiene en cuenta la distancia física de los SNP y los datos de recombinación de las probabilidades de recombinación específicas de una ubicación y los datos observados de las mediciones genéticas del plasma materno para obtener el genotipo más probable del feto.

En una realización de la divulgación, se incluye un número de sondas adicionales por locus vinculadas a enfermedad en el conjunto de loci polimórficos focalizados; el número de sondas adicionales por locus vinculadas a enfermedad puede ser de entre 4 y 10, entre 11 y 20, entre 21 y 40, entre 41 y 60, entre 61 y 80, o combinaciones de estos.

*Determinación del número de moléculas en una muestra*

En el presente documento se describe un método para determinar el número de moléculas de ADN en una muestra que genera una molécula exclusivamente identificada para cada molécula de ADN original de la muestra durante la primera ronda de amplificación del ADN. En el presente documento se describe un procedimiento para conseguir este fin seguido de un método de secuenciación molecular o clonal.

Este método implica la focalización de uno o más loci específicos y la generación de una copia etiquetada de las moléculas originales de forma que la mayoría o todas las moléculas etiquetadas de cada locus focalizado tendrán una etiqueta única que se podrán distinguir entre sí tras la secuenciación de este código de barras utilizando la secuenciación de una única molécula o clonal. Cada código de barras único secuenciado representa una única molécula en la muestra original. Simultáneamente, los datos de secuencia se utilizan para establecer el locus del que procede la molécula. Utilizando esta información se puede determinar el número de moléculas únicas de la muestra original para cada locus.

Este método se puede utilizar para cualquier aplicación en la que se requiere la evaluación cuantitativa del número de moléculas de una muestra original. Por otra parte, el número de moléculas únicas de una o más dianas se puede relacionar con el número de moléculas únicas de una o más dianas diferentes para determinar el número de copias, la distribución de los alelos o el ratio alélico relativos. Alternativamente, se puede establecer un modelo del número de copias detectadas de las diversas dianas mediante una distribución, a fin de identificar el número de copias más probable de las dianas originales. Las aplicaciones incluyen, entre otras, la detección de inserciones y deleciones como las que se encuentran en los portadores de la distrofia muscular de Duchenne; la cuantificación de deleciones o duplicaciones de segmentos de cromosomas como las observadas en variantes del número de copias; el número de copias del cromosoma de las muestras de individuos nacidos; el número de copias del cromosoma de las muestras de individuos no nacidos como embriones o fetos.

El método se puede combinar con la evaluación simultánea de variaciones contenidas en el elemento focalizado por secuencia. Esto se puede utilizar para determinar el número de moléculas que representan cada alelo en la muestra original. El método del número de copias se puede combinar con la evaluación de los SNP u otras variaciones de secuencia para determinar el número de copias del cromosoma de los individuos no nacidos; la discriminación y cuantificación de las copias de los loci que tienen variaciones de secuencias cortas, pero en las que la PCR puede amplificar a partir de múltiples regiones diana como en la detección de un portador de atrofia muscular espinal; determinación del número de copias de diferentes fuentes de moléculas de muestras que consisten en mezclas de diferentes individuos como en la detección de aneuploidía fetal a partir del ADN flotante libre obtenido del plasma materno.

En una realización de la divulgación, el método se refiere a un locus diana único que puede comprender uno o más de los pasos siguientes: 1) Diseño de un par estándar de oligómeros para amplificación PCR de un locus específico. 2) Adición, durante la síntesis, de una secuencia de bases especificadas sin complementariedad o con una complementariedad mínima con el locus diana o el genoma para el extremo 5' del oligómero específico diana. Esta secuencia, denominada cola, es una secuencia conocida, que se utilizará para la amplificación posterior, seguida por una secuencia de nucleótidos aleatorios. Estos nucleótidos aleatorios comprenden la región aleatoria. La región aleatoria comprende una secuencia generada aleatoriamente de ácidos nucleicos que probabilísticamente difieren entre cada molécula de la sonda. Por consiguiente, tras la síntesis, el grupo de oligómeros con cola se compondrá de una recopilación de oligómeros que comienzan con una secuencia conocida seguida de una secuencia

desconocida que difiere entre moléculas, seguida de una secuencia específica diana. 3) Realización de una ronda de amplificación (desnaturalización, hibridación, ampliación) utilizando solo el oligómero con cola. 4) Adición de exonucleasa a la reacción, detención efectiva de la reacción por PCR e incubación de la reacción a la temperatura apropiada para eliminar los oligómeros de cadena simple directos que no se han hibridado y ampliar para formar un producto de doble cadena. 5) Incubación de la reacción a temperatura elevada para desnaturalizar la exonucleasa y eliminar su actividad. 6) Adición a la reacción de un nuevo oligonucleótido que es complementario a la cola del oligómero utilizado en la primera reacción junto con el otro oligómero específico diana para permitir la amplificación PCR del producto generado en la primera ronda de PCR. 7) Continuación de la amplificación para generar suficiente producto para la secuenciación clonal en sentido descendente. 8) Medición del producto de la PCR amplificado a través de múltiples métodos, por ejemplo, secuenciación clonal, para disponer de un número suficiente de bases para prolongar la secuencia.

En una realización, un método de la presente divulgación implica la focalización de múltiples loci en paralelo o de otro modo. Se pueden generar cebadores para diferentes loci diana independientemente y mezclados para crear grupo para la PCR multiplexada. En una realización, las muestras originales se pueden dividir en subconjuntos y se pueden focalizar diferentes loci en cada subconjunto antes de su recombinación y secuenciación. En una realización, el paso de etiquetado y un número de ciclos de amplificación se pueden realizar antes de subdividir el conjunto, a fin de garantizar un etiquetado eficiente de todas las dianas antes de la división y mejorar la amplificación posterior mediante la amplificación continua utilizando conjuntos más pequeños de cebadores en grupos subdivididos.

Un ejemplo de una aplicación en la que esta tecnología resultaría particularmente útil es el diagnóstico no invasivo de la aneuploidía prenatal, donde el ratio de alelos en un determinado locus o una distribución de los alelos en un número de loci se puede utilizar para ayudar a determinar el número de copias de un cromosoma presente en el feto. En este contexto, resulta recomendable amplificar el ADN presente en la muestra inicial al tiempo que se mantienen las cantidades relativas de los diversos alelos. En algunas circunstancias, especialmente en los casos en los que hay una cantidad muy pequeña de ADN, por ejemplo, menos de 5000 copias del genoma, menos de 1000 copias del genoma, menos de 500 copias del genoma y menos de 100 copias del genoma, se puede dar un fenómeno denominado cuello de botella. Esto ocurre cuando hay un pequeño número de copias de cualquier alelo dado en la muestra inicial y los sesgos de amplificación pueden resultar en el conjunto amplificado de ADN que tiene unos ratios significativamente diferentes de los alelos que se encuentran en la mezcla inicial de ADN. Aplicando un conjunto de códigos de barras únicos o prácticamente únicos a cada cadena de ADN antes de la amplificación por PCR estándar, se pueden excluir n-1 copias de ADN de un conjunto de n moléculas idénticas del ADN secuenciado que procedía de la misma molécula original.

Por ejemplo, supongamos un SNP heterocigoto en el genoma de un individuo y una mezcla de ADN del individuo donde 10 moléculas de cada alelo se encuentran presentes en la muestra original de ADN.

Tras la amplificación podrá haber 100 000 moléculas de ADN correspondientes a ese locus. Debido a los procesos estocásticos, el ratio de ADN podría ser cualquiera entre 1:2 y 2:1; sin embargo, dado que cada una de las moléculas originales se ha etiquetado con una etiqueta única, se podría determinar que el ADN del grupo amplificado procedía exactamente de 10 moléculas de ADN de cada alelo. Por tanto, este método proporcionaría una medición más precisa de las cantidades relativas de cada alelo que un método que no utilice este método. Para los métodos en los que resulta deseable que la cantidad relativa de sesgo alélico sea minimizada, este método proporcionará datos más precisos.

La asociación del fragmento secuenciado con el locus diana se puede conseguir de múltiples maneras. En una realización, se obtiene una secuencia de longitud suficiente a partir de un fragmento focalizado para ampliar el código de barras de la molécula, así como un número suficiente de bases únicas correspondientes a la secuencia diana para permitir la identificación sin ambigüedades del locus diana. En otra realización, el cebador del código de barras molecular que contiene el código de barras molecular generado aleatoriamente también puede contener un código de barras específico de locus (código de barras de locus) que identifique la diana a la que se va a asociar. Este código de barras de locus sería idéntico entre todos los cebadores del código de barras molecular para cada diana individual y, por tanto, todos los amplicones resultantes, aunque diferente de todas las demás dianas. En una realización, el método de etiquetado que se describe en el presente documento se puede combinar con un protocolo de anidado unilateral.

En una realización de la divulgación, el diseño y la generación de cebadores del código de barras molecular se puede reducir en la práctica a lo siguiente: los cebadores de códigos de barras moleculares se pueden componer de una secuencia que no es complementaria de la secuencia diana, seguida de una región de código de barras molecular aleatoria, seguida de una secuencia específica diana. La secuencia 5' del código de barras molecular se puede utilizar para la amplificación por PCR de la subsecuencia y se puede componer de secuencias útiles para la conversión del amplicón en una biblioteca para la secuenciación. La secuencia del código de barras molecular aleatoria se podría generar de múltiples maneras. El método preferible sintetiza el cebador de etiquetado de la molécula de tal forma que incluye las cuatro bases para la reacción durante la síntesis de la región del código de barras. Todas o diversas combinaciones de bases pueden ser especificadas utilizando los códigos de ambigüedad de ADN IUPAC. De esta manera la recopilación sintetizada de moléculas contendrá una mezcla aleatoria de secuencias en la región del código de barras molecular. La longitud del código de barras determinará cuántos cebadores contendrán códigos de barras únicos. El número de secuencias únicas está relacionado con la longitud

de la región del código de barras como $N^1$, donde N es el número de bases, típicamente 4, y L es la longitud del código de barras. Un código de barras de cinco bases puede producir hasta 1024 secuencias únicas; un código de barras de ocho bases puede producir 65536 secuencias únicas. En una realización, el ADN se puede medir mediante un método de secuenciación, donde los datos de secuencia representan la secuencia de una única molécula. Esto puede incluir métodos en los que las moléculas únicas son secuenciadas directamente o métodos en los que las moléculas únicas son amplificadas para formar clones detectables mediante el instrumento de la secuencia, pero que aun así representan moléculas únicas, en el presente documento denominados secuenciación clonal.

*Algunas realizaciones*

En algunas realizaciones, se divulga un método en el presente documento para generar un informe en el que se divulga el estado determinado de ploidía de un cromosoma de un feto en gestación, donde el método comprende:

la obtención de una primera muestra que contiene ADN de la madre del feto y ADN del feto; la obtención de datos genotípicos de uno o los dos progenitores del feto; la preparación de la primera muestra aislando el ADN para obtener una muestra preparada; la medición del ADN de la muestra preparada en diversos loci polimórficos; el cálculo, por ordenador, de los recuentos de alelos o las probabilidades de los recuentos de alelos en la pluralidad de loci polimórficos de las mediciones de ADN realizadas en la muestra preparada; la creación, por ordenador, de una pluralidad de hipótesis de ploidía relativas a las probabilidades de recuentos de alelos previstas en la pluralidad de loci polimórficos del cromosoma para los diferentes estados de ploidía posibles del cromosoma; la creación, por ordenador, de un modelo de distribución conjunto para la probabilidad del recuento de alelos de cada locus polimórfico del cromosoma para cada hipótesis del estado de ploidía utilizando datos genotípicos de uno o los dos progenitores del feto; la determinación, por ordenador, de una probabilidad relativa de cada una de las hipótesis de ploidía utilizando el modelo de distribución conjunto y las probabilidades del recuento de alelos calculadas para la muestra preparada; la determinación del estado de ploidía del feto, seleccionando el estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada; y la generación de un informe en el que se divulga el estado de ploidía determinado.

En algunas realizaciones de la divulgación, el método se utiliza para determinar el estado de ploidía de una pluralidad de fetos en gestación en diversas madres, donde el método también comprende:

la determinación del porcentaje de ADN que es de origen fetal en cada una de las muestras preparadas; y donde el paso de medición del ADN en la muestra preparada se realiza mediante secuenciación de un número de moléculas de ADN en cada una de las muestras preparadas, donde se secuencian más moléculas de ADN de esas muestras preparadas que tienen una fracción menor de ADN fetal que de las muestras preparadas que tienen una fracción mayor de ADN fetal.

En algunas realizaciones de la divulgación, el método se utiliza para determinar el estado de ploidía de una pluralidad de fetos en gestación en diversas madres, y donde la medición del ADN en la muestra preparada se realiza, para cada uno de los fetos, mediante la secuenciación de una primera fracción de la muestra preparada de ADN para obtener un primer conjunto de mediciones. El método comprende además: la realización de una primera determinación de la probabilidad relativa para cada una de las hipótesis de ploidía para cada uno de los fetos, teniendo en cuenta el primer conjunto de mediciones de ADN; la resecuenciación de una segunda fracción de la muestra preparada de esos fetos donde la primera determinación de la probabilidad relativa para cada una de las hipótesis de ploidía indica que una hipótesis de ploidía correspondiente a un feto aneuploide tiene una probabilidad significativa aunque no concluyente, para obtener un segundo conjunto de mediciones; la realización de una segunda determinación de la probabilidad relativa para las hipótesis de ploidía para los fetos utilizando el segundo conjunto de mediciones y opcionalmente también el primer conjunto de mediciones; y la determinación de los estados de ploidía de los fetos cuya segunda muestra fue resecuenciada mediante la selección del estado de ploidía correspondiente a la hipótesis con la probabilidad más elevada determinada por la segunda determinación de probabilidad relativa.

En algunas realizaciones, se divulga una composición de interés, donde la composición de interés comprende: una muestra de ADN enriquecido preferentemente, donde la muestra de ADN enriquecido preferentemente ha sido enriquecida preferentemente en diversos loci polimórficos de una primera muestra de ADN, donde la primera muestra de ADN consistía en una mezcla de ADN materno y ADN fetal obtenida de plasma materno, donde el grado de enriquecimiento es al menos de un factor 2, y donde el sesgo alélico entre la primera muestra y la muestra enriquecida preferentemente se selecciona, de media, del grupo compuesto por menos de un 2%, menos de un 1%, menos de un 0,5%, menos de un 0,2%, menos de un 0,1%, menos de un 0,05%, menos de un 0,02% y menos de un 0,01%. En algunas realizaciones, se divulga un método para crear una muestra de este ADN enriquecido preferentemente.

En algunas realizaciones se divulga un método para determinar la presencia o ausencia de una aneuploidía fetal en una muestra de tejido materno que comprende ADN genómico fetal y materno, donde el método consiste en lo siguiente: a) obtención de una mezcla de ADN genómico fetal y materno de dicha muestra de tejido materno; b) enriquecimiento selectivo de la mezcla de ADN fetal y materno en diversos alelos polimórficos; c) distribución de los fragmentos selectivamente enriquecidos de la mezcla de ADN genómico fetal y materno del paso a) para proporcionar muestras de reacción que comprenden una única molécula de ADN genómico o productos de

amplificación de una única molécula de ADN genómico; d) la realización de una secuenciación de ADN masivamente paralela de los fragmentos selectivamente enriquecidos en las muestras de reacción del paso c) para determinar la secuencia de dichos fragmentos selectivamente enriquecidos; e) la identificación de los cromosomas a los que pertenecen las secuencias obtenidas en el paso d); 1) análisis de los datos del paso d) para determinar i) el número de fragmentos de ADN genómico del paso d) que pertenecen al menos a un primer cromosoma diana que se presume que es diploide tanto en la madre como en el feto, y ii) el número de fragmentos de ADN genómico del paso d) que pertenecen a un segundo cromosoma diana, donde dicho segundo cromosoma diana se sospecha que es aneuploide en el feto; g) el cálculo de la distribución prevista del número de fragmentos de ADN genómico del paso d) del segundo cromosoma diana si el segundo cromosoma diana es euploide, utilizando el número determinado en el paso f) parte i); h) el cálculo de la distribución prevista del número de fragmentos de ADN genómico del paso d) para el segundo cromosoma diana si el segundo cromosoma diana es aneuploide, utilizando el primer número del paso f) parte i) y una fracción estimada de ADN fetal que se encuentra en la mezcla del paso b); e i) la utilización de una probabilidad máxima o un método a posteriori máximo para determinar si el número de fragmentos de ADN genómico determinado en el paso f) parte ii) es más probable que forme parte de la distribución calculada en el paso g) o la distribución calculada en el paso h); indicando de este modo la presencia o ausencia de aneuploidía fetal.

Sección experimental

Las realizaciones divulgadas en el presente documento se describen en los siguientes Ejemplos, que se incluyen para ayudar a entender la divulgación y no se deberá interpretar en ningún caso que limitan el alcance de la divulgación definida en las reivindicaciones que se recoge a continuación. Los siguientes ejemplos se incluyen para proporcionar a los expertos en la técnica una divulgación y descripción completas de cómo utilizar las realizaciones descritas y no pretenden limitar el alcance de la divulgación ni pretenden declarar que los siguientes experimentos son todos los experimentos o los únicos realizados. Se ha procurado en lo posible garantizar la precisión con respecto a las cifras utilizadas (por ejemplo, cantidades, temperaturas, etc.) pero deben tenerse en cuenta algunos errores y desviaciones experimentales. Salvo que se indique lo contrario, las partes son partes por volumen y la temperatura se expresa en grados centígrados. Se entenderá que las variaciones en los métodos descritos se pueden introducir sin modificar los aspectos fundamentales que pretenden ilustrar los experimentos.

*Experimento 1*

El objetivo consistía en demostrar que un algoritmo de la estimación de la probabilidad máxima (MLE) bayesiana que utiliza genotipos paternos para calcular la fracción fetal mejora la precisión del diagnóstico no invasivo de trisomía prenatal en comparación con los métodos publicados.

Se crearon datos de secuencia simulados para el cfDNA materno mediante lecturas de muestreo obtenidas en la trisomía -21 y las respectivas líneas celulares maternas. La tasa de determinaciones correctas de disomía y trisomía se determinó a partir de 500 simulaciones en diversas fracciones fetales para un método publicado (Chiu et al. BMJ 2011;342:c7401) y de nuestro algoritmo basado en MLE. Validamos las simulaciones mediante la obtención de cinco millones de lecturas de secuenciación forzada de cuatro madres embarazadas y los respectivos padres recopiladas de conformidad con un protocolo aprobado por el IRB. Los genotipos parentales se obtuvieron en un array de 290 000 SNP. (Ver Figura 14)

En las simulaciones, el método basado en la MLE consiguió una precisión del 99,0% para las fracciones fetales de tan solo el 9% y certezas notificadas que correspondían perfectamente a la precisión general. Validamos estos resultados utilizando cuatro muestras reales donde obtuvimos todas las determinaciones correctas con una certeza computada superior al 99%. Por el contrario, nuestra implementación del algoritmo publicado para Chiu et al. precisó una fracción fetal del 18% para conseguir una precisión de 99,0%, y consiguió tan solo una precisión del 87,8% al 9% de ADN fetal.

La determinación de la fracción fetal de los genotipos parentales conjuntamente con un método basado en MLE consigue una mayor precisión que los algoritmos publicados a las fracciones fetales previstas durante el primer y el segundo trimestre. Por otra parte, el método divulgado en el presente documento produce una métrica de certeza que resulta fundamental para determinar la fiabilidad del resultado, especialmente a fracciones fetales bajas donde la detección de la ploidía resulta más difícil. Los métodos publicados utilizan un método de umbral menos preciso para la determinación del estado de ploidía basado en grandes conjuntos de datos de formación sobre disomía, y un método que predefine una tasa de falsos positivos. Además, sin una métrica de la certeza, los métodos publicados corren el riesgo de dar resultados falsos negativos cuando se dispone de un cfDNA fetal insuficiente para determinar el estado de ploidía. En algunas realizaciones, se calcula una estimación de certeza para el estado de ploidía determinado.

*Experimento 2*

El objetivo consistía en mejorar la detección no invasiva de la trisomía fetal 18, 21 y X, particularmente en muestras compuestas de baja fracción fetal, utilizando un método de secuenciación focalizado combinado con los datos de los genotipos parentales y de Hapmap en un algoritmo de estimación de la probabilidad máxima (MLE) bayesiana.

Las muestras maternas de cuatro embarazos euploides y dos positivos en trisomía y las respectivas muestras paternas se obtuvieron según un protocolo aprobado por el IRB de pacientes en los que se conocía el cariotipo fetal.

El cfDNA materno se extrajo del plasma y se obtuvieron unos 10 millones de lecturas de secuencia tras el enriquecimiento preferente que se focalizó en SNP específicos. Las muestras parentales fueron secuenciadas similarmente para obtener los genotipos.

El algoritmo descrito determinó correctamente la disomía del cromosoma 18 y 21 para todas las muestras euploides y los cromosomas normales de las muestras aneuploides. Las determinaciones de la trisomía 18 y 21 fueron correctas, al igual que los números de copias del cromosoma X en los fetos varones y hembras. La certeza producida por el algoritmo superó el 98% en todos los casos.

El método descrito documentó con precisión el estado de ploidía de todos los cromosomas testados de seis muestras, incluyendo muestras compuestas por menos de un 12% de ADN fetal, que representan aproximadamente el 30% de las muestras del primer trimestre y el segundo trimestre. La diferencia crucial entre el algoritmo de la MLE instantánea y los métodos publicados es que aprovecha los genotipos parentales y los datos de Hapmap para mejorar la precisión y generar una métrica de la certeza. Con bajas fracciones fetales, todos los métodos resultan menos precisos; es importante identificar correctamente las muestras con una cantidad insuficiente de cfDNA fetal para realizar una determinación fiable. Otros han utilizado sondas específicas del cromosoma Y para estimar la fracción fetal de los fetos varones, pero la determinación del genotipo parental concurrente permite la estimación de la fracción fetal para ambos sexos. Otra limitación inherente de los métodos publicados utilizando secuenciación forzada no focalizada es que la precisión del estado de ploidía varía entre cromosomas debido a diferencias en factores como la riqueza de GC. El método de la secuenciación focalizada instantánea es independiente en gran medida de estas variaciones a escala cromosómica y ofrece un rendimiento más uniforme entre cromosomas.

*Experimento 3*

El objetivo consistía en determinar si la trisomía se puede detectar con un elevado grado de certeza en un feto triploide, utilizando novedosos métodos informáticos para analizar loci de SNP de ADN fetal flotante libre del plasma materno.

Se extrajeron 20 ml de sangre de una paciente embarazada tras una prueba ultrasónica anormal. Tras la centrifugación, el ADN materno se extrajo de la capa leucocitaria (DNEASY, QIAGEN); el cfADN se extrajo del plasma (QIAAMP QIAGEN). La secuenciación focalizada se aplicó a loci de SNP de los cromosomas 2, 21 y X, en ambas muestras de ADN. La estimación de la probabilidad máxima bayesiana seleccionó las hipótesis más probables del conjunto de todos los estados de ploidía posibles. El método determina la fracción de ADN fetal, el estado de ploidía y las certezas explícitas en la determinación del estado del ploidía. No se realizan suposiciones sobre la ploidía de un cromosoma de referencia. El diagnóstico utiliza una estadística de prueba que es independiente de los recuentos de lecturas de secuencias, que es el estado de la técnica reciente.

El método instantáneo diagnosticó con precisión la trisomía de los cromosomas 2 y 21. La fracción del niño se estimó al 11,9% [CI 11.7-12,1]. Se descubrió que el feto tenía una copia materna y dos copias paternas de los cromosomas 2 y 21 con una certeza de efectivamente 1 (probabilidad de error < 10-30). Esto se consiguió con 92 600 y 258 100 lecturas en los cromosomas 2 y 21, respectivamente.

Esta es la primera demostración de diagnóstico no invasivo prenatal de cromosomas trisómicos de la sangre materna donde el feto era triploide, tal y como se confirmó por el cariotipo de metafase. Los métodos existentes de diagnóstico no invasivo no detectarían la aneuploidía en esta muestra. Los métodos actuales confían en el exceso de lecturas de secuencia sobre un cromosoma trisómico respecto de los cromosomas de referencia disómicos; pero un feto triploide no tiene referencia disómica.

Por otra parte, los métodos existentes no alcanzarían una determinación del estado de ploidía con una certeza similarmente elevada con esta fracción de ADN fetal y el número de lecturas de secuencia. Es sencillo ampliar el método a los 24 cromosomas.

*Experimento 4*

El siguiente protocolo se utilizó para la amplificación de 800-plex de ADN aislado de plasma materno de un embarazo euploide y también ADN genómico de una línea de células de triploidía 21 utilizando una PCR estándar (lo que significa que no se utilizó el anidado). La preparación de la biblioteca y la amplificación implicaron la generación de extremos romos en un único tubo seguida de la adición de una cola A. La unión del adaptador se realizó utilizando el kit de enlace presente en el kit de AGILENT SURESELECT, y la PCR se ejecutó durante 7 ciclos. A continuación, 15 ciclos de STA (95°C durante 30 s; 72°C durante 1 min; 60°C durante 4 min; 65°C durante 1 min; 72°C durante 30 s) utilizando 800 pares de cebadores diferentes focalizados en los SNP de los cromosomas 2, 21 y X. La reacción se ejecutó con una concentración de cebador de 12,5 nM. Después, se secuenció el ADN con un secuenciador ILLUMINAIIGAX. El secuenciador produjo 1,9 millones de lecturas, de las cuales, el 92% correspondían al genoma; de estas lecturas que correspondían al genoma, más del 99% correspondían a una de las regiones focalizadas por los cebadores focalizados. Estas cifras fueron básicamente las mismas para el ADN del plasma y para el ADN genómico. La Figura 15 muestra el ratio de los dos alelos para los aproximadamente 780 SNP que fueron detectados por el secuenciador en el ADN genómico que se tomó de una línea celular con trisomía conocida en el cromosoma 21. Cabe señalar que los ratios alélicos se recogen aquí para facilitar la visualización, porque las distribuciones alélicas no son fáciles de leer a simple vista. Los círculos representan SNP en cromosomas disómicos, mientras que las estrellas representan SNP en cromosomas trisómicos. La Figura 16 es otra

representación de los mismos datos que en la Figura X, donde el eje Y es el número relativo de A y B medido para cada SNP, y donde el eje X es el número de SNP donde los SNP están separados por cromosomas. En la Figura 16, los SNP 1 a 312 están en el cromosoma 2, los SNP 313 a 605 están en el cromosoma 21, que es trisómico, y los SNP 606 a 800 están en el cromosoma X. Los datos de los cromosomas 2 y X muestran un cromosoma disómico, dado que los recuentos de secuencia relativos se encuentran en tres grupos: AA en la parte superior del gráfico, BB en la parte inferior del gráfico y AB en el centro del gráfico. Los datos del cromosoma 21, que es trisómico, muestran cuatro grupos: AAA en la parte superior del gráfico, AAB alrededor de la línea 0,65 (2/3), ABB alrededor de la línea 35 (1/3) y BBB en la parte inferior del gráfico.

La Figura 17 muestra los datos para el mismo protocolo de 800-plex, pero medidos en ADN amplificado a partir de cuatro muestras de plasma de mujeres embarazadas. Para estas cuatro muestras, esperamos observar siete grupos de puntos: 1) a lo largo de la parte superior del gráfico están los loci donde tanto la madre como el feto son AA; 2) ligeramente por debajo de la parte superior del gráfico están los loci en los que la madre es AA y el feto es AB; 3) ligeramente por encima de la línea 0,5 están los loci en los que la madre es AB y el feto es AA; 4) a lo largo de la línea 0,5 están los loci en los que la madre y el feto son AB; 5) ligeramente por debajo de la línea 0,5 están los loci en los que la madre es AB y el feto es BB; 6) ligeramente por encima de la parte inferior del gráfico están los loci en los que la madre es BB y el feto es AB; 7) a lo largo de la parte inferior del gráfico están los loci en los que la madre y el feto son BB. Cuanto menor sea la fracción fetal, menor será la separación entre los grupos 1) y 2), entre los grupos 3), 4) y 5), y entre los grupos 6) y 7). Se espera que la separación sea la mitad de la fracción de ADN que es de origen fetal. Por ejemplo, si el ADN es un 20% fetal y un 80% materno, esperamos que 1) a 7) se centren en 1,0, 0,9, 0,6, 0,5, 0,4, 0,1 y 0,0 respectivamente; ver por ejemplo, la Figura 17, POOL1_BC5_ref rate. Si, por el contrario, el ADN es un 8% fetal y un 92% materno, esperamos que 1) a 7) se centren en 1,00, 0,96, 0,54, 0,50, 0,46, 0,04 y 0,00 respectivamente; ver por ejemplo, la Figura 17, POOL1_BC2_ref_rate. Si no se detecta ADN fetal, no esperamos observar 2), 3), 5) ni 6); alternativamente podríamos decir que la separación es cero y, por tanto, 1) y 2) se encuentran uno encima del otro, al igual que 3), 4) y 5), y también 6) y 7); ver por ejemplo, la Figura 17, POOL1_BC7_ref_rate. Cabe señalar que la fracción fetal para la Figura 17, POOL1_BC1 ref_rate es aproximadamente del 25%.

*Experimento 5*

La mayoría de los métodos de amplificación y medición de ADN producirán cierto sesgo alélico, donde los dos alelos que se encuentran típicamente en un locus se detectan con intensidades o recuentos que no son representativos de las cantidades reales de alelos en la muestra de ADN. Por ejemplo, para un único individuo, en un locus heterocigoto esperamos observar un ratio 1:1 de los dos alelos, que es el ratio teórico previsto para un locus heterocigoto; sin embargo, debido al sesgo alélico, podemos observar 55:45 o incluso 60:40. Cabe señalar también que en el contexto de la secuenciación, si la profundidad de lectura es baja, entonces el simple ruido estocástico podría provocar un sesgo alélico significativo. En una realización, se puede elaborar un modelo del comportamiento de cada SNP de forma que si se observa un sesgo constante para determinados alelos, este sesgo se pueda corregir. La Figura 18 muestra la fracción de datos que se puede explicar por la varianza binomial antes y después de la corrección del sesgo. En la Figura 18, las estrellas representan el sesgo alélico observado en los datos de secuencia en bruto para el experimento de 800.plex; los círculos representan el sesgo alélico tras la corrección. Cabe señalar que si no hubiese sesgo alélico en absoluto, cabría esperar que los datos cayesen por debajo de la línea x=y. Un conjunto similar de datos que se produjo amplificando ADN utilizando una amplificación focalizada de 150-plex generó datos muy cercanos a la línea 1:1 tras la corrección del sesgo.

*Experimento 6*

La amplificación universal de ADN utilizando adaptadores ligados con cebadores específicos para las etiquetas de los adaptadores, donde la hibridación del cebador y los tiempos de extensión se limitan a unos pocos minutos, tiene el efecto de enriquecer la proporción de cadenas de ADN más cortas. La mayoría de los protocolos de bibliotecas diseñadas para crear bibliotecas de ADN adecuadas para la secuenciación contienen este paso y algunos ejemplos de protocolos están publicados y son conocidos por los expertos en la técnica. En algunas realizaciones de la invención, los adaptadores con una etiqueta universal están ligados al ADN plasmático y son amplificados utilizando cebadores específicos para la etiqueta del adaptador. En algunas realizaciones, la etiqueta universal puede ser la misma etiqueta que la utilizada para la secuenciación, puede ser una etiqueta universal exclusiva para la amplificación PCR o puede ser un conjunto de etiquetas. Dado que típicamente el ADN fetal es de naturaleza corta, mientras que el ADN materno puede ser de naturaleza corta y larga, este método tiene el efecto de enriquecer la proporción de ADN fetal en la mezcla. El ADN flotante libre, que se cree que es ADN de células apoptóticas, y que contiene tanto ADN fetal como materno, es corto —en su mayoría menos de 200 pb—. El ADN celular liberado por la lisis celular, un fenómeno común tras la flebotomía, suele ser materno en su práctica totalidad y también es bastante largo —en su mayoría más de 500 pb—. Por tanto, las muestras de sangre que se han asentado durante más de unos pocos minutos contendrán una mezcla de ADN corto (fetal + materno) y más largo (materno). Al realizar una amplificación universal con unos tiempos de extensión relativamente cortos con plasma materno, seguida de una amplificación focalizada, tenderá a aumentar la proporción relativa de ADN fetal en comparación con el plasma que se ha amplificado utilizando solamente la amplificación focalizada. Esto se puede observar en la Figura 19, que muestra el porcentaje fetal medido cuando se introduce ADN plasmático (eje vertical) frente al porcentaje fetal medido cuando se introduce ADN plasmático para el que se ha preparado una biblioteca utilizando el protocolo de preparación de la biblioteca de ILLUMINA GAIIx. Todos los puntos se encuentran por debajo de la línea, lo que

indica que el paso de preparación de la biblioteca enriquece la fracción de ADN de origen fetal. Dos muestras de plasma que eran rojas, indicando hemolisis y por tanto que se produciría una cantidad aumentada de ADN materno largo presente del lisado celular, muestran un enriquecimiento particularmente significativo de la fracción fetal cuando la preparación de la biblioteca se realiza antes de la amplificación focalizada. El método divulgado en el presente documento resulta particularmente útil en los casos en los que existe hemolisis o se ha producido alguna otra situación en la que las células que comprenden cadenas relativamente largas de ADN contaminante se han lisado, contaminando la muestra combinada de ADN corto con ADN largo. Típicamente, los tiempos de hibridación y extensión relativamente cortos son de entre 30 segundos y dos minutos, aunque podrían ser de tan solo 5 o 10 segundos o menos o de hasta 5 o 10 minutos.

*Experimento 7*

El siguiente protocolo se utilizó para la amplificación de 1.200-plex de ADN aislado de plasma materno de un embarazo euploide y también ADN genómico de una línea de células de triploidía 21 utilizando un protocolo de PCR directa y también un método de semi-anidado. La preparación de la biblioteca y la amplificación implicaron la generación de extremos romos en un único tubo seguida de la adición de una cola A. La unión del adaptador se realizó utilizando una modificación del kit de unión que se encuentra en el kit de AGILENT SURESELECT, y la PCR se ejecutó durante 7 ciclos. En el conjunto de cebadores focalizados, había 550 ensayos para SNP del cromosoma 21 y 325 ensayos para los SNP de cada uno de los cromosomas 1 y X. Ambos protocolos implicaban 15 ciclos de STA (95°C durante 30 s; 72°C durante 1 min; 60°C durante 4 min; 65°C durante 30 s; 72°C durante 30 s) utilizando una concentración de cebador de 16 nM. El protocolo de PCR semi-anidada implicaba una segunda amplificación de 15 ciclos de STA (95°C durante 30 s; 72°C durante 1 min; 60°C durante 4 min; 65°C durante 30 s; 72°C durante 30 s) utilizando una concentración de etiqueta directa interna de 29 nM, y una concentración de etiqueta inversa de 1 uM o 0,1 uM. A continuación, se secuenció el ADN con un secuenciador ILLUMINA IIGAX. Para el protocolo de PCR directa, el 73% de las lecturas corresponden al genoma; para el protocolo semi-anidado, el 97,2% de las lecturas de secuencia corresponden al genoma. Por tanto, el protocolo semi-anidado proporciona aproximadamente un 30% más de información, presumiblemente debido en gran medida a la eliminación de cebadores que es más probable que causen dímeros del cebador.

La variabilidad de la profundidad de lectura tiende a ser mayor cuando se utiliza el protocolo semi-anidado que cuando se utiliza el protocolo de PCR directa (ver la Figura 20), donde los rombos se refieren a la profundidad de lectura para los loci ejecutados con el protocolo semi-anidado y los cuadrados se refieren a la profundidad de lectura para los loci ejecutados sin anidado. Los SNP están dispuestos por profundidad de lectura para los rombos, de forma que todos los rombos se encuentran en una línea curvada, mientras que los cuadrados parecen estar estrechamente correlacionados; la disposición de los SNP es arbitraria y es la altura del punto la que denota la profundidad de lectura y no su ubicación de izquierda a derecha.

En algunas realizaciones, los métodos descritos en el presente documento pueden conseguir unas varianzas excelentes de la profundidad de lectura (DOR). Por ejemplo, en una versión de este experimento (Figura 21) que utiliza una amplificación por PCR directa de 1200-plex de ADN genómico, de los 1200 ensayos: 1186 ensayos tenían una DOR superior a 10; la profundidad de lectura media era de 400; 1063 ensayos (88,6%) tenían una profundidad de lectura de entre 200 y 800, y una ventana ideal donde el número de lecturas para cada alelo es lo suficientemente elevada como para proporcionar datos significativos, mientras que el número de lecturas para cada alelo no es tan elevado por lo que el uso marginal de estas lecturas fue particularmente limitado. Solo 12 alelos tenían una profundidad de lectura superior con la máxima en 1035 lecturas. La desviación estándar de la DOR fue 290, la DOR media fue 453, el coeficiente de varianza de la DOR fue 64%; hubo 950 000 lecturas totales y el 63,1% de las lecturas correspondían al genoma. En otro experimento (Figura 22) que utilizó un protocolo semi-anidado de 1200-plex, la DOR fue superior. La desviación estándar de la DOR fue 583, la DOR media fue 630, el coeficiente de varianza de la DOR fue 93%; hubo 870 000 lecturas totales y el 96,3% de las lecturas correspondían al genoma. Cabe señalar que en ambos casos, los SNP están dispuestos por la profundidad de lectura para la madre, por lo que la línea curvada representa la profundidad de lectura materna. La diferenciación entre el niño y el padre no es significativa; solo la tendencia es significativa a efectos de la presente explicación.

*Experimentos*

En un experimento, se utilizó el protocolo de PCR semi-anidada de 1200-plex para amplificar ADN de una célula y de tres células. Este experimento es relevante para las pruebas de aneuploidía prenatal que utilizan células fetales aisladas de sangre materna o para el diagnóstico genético previo al implante utilizando blastómeros biopsiados o muestras de trofectoderma. Hubo tres replicados de 1 y 3 células de dos individuos (46 XY y 47 XX+21) por condición. Los ensayos estaban dirigidos a los cromosomas 1, 21 y X. Se utilizaron tres métodos de lisis diferentes: ARCTURUS, MPERv2 y lisis alcalina. La secuenciación se ejecutó multiplexando 48 muestras en una línea de secuenciación. El algoritmo devolvió determinaciones correctas del estado de ploidía para cada uno de los tres cromosomas y para cada uno de los replicados.

*Experimento 9*

En un experimento se prepararon cuatro muestras de plasma materno y se amplificaron utilizando un protocolo hemi-anidado de 9600-plex. Las muestras se prepararon como sigue: Hasta 40 ml de sangre materna se centrifugaron para aislar la capa leucocitaria y el plasma. El ADN genómico de la muestra materna se preparó a

partir de la capa leucocitaria y el ADN paterno se preparó a partir de una muestra de sangre o una muestra de saliva. El cfADN del plasma materno se aisló utilizando el kit QIAGEN CIRCULATING NUCLEIC ACID y se eluyó en 45 uL de tampón TE conforme a las instrucciones del fabricante. Los adaptadores de unión universal se unieron al extremo de cada molécula de 35 uL de ADN de plasma purificado y se amplificaron las bibliotecas durante siete ciclos utilizando cebadores específicos de adaptador. Las bibliotecas se purificaron con perlas AGENCOURT AMPURE y se eluyeron en 50 ul de agua.

3 ul de DNA se amplificaron con 15 ciclos de STA (95 °C durante 10 min para la activación de la polimerasa inicial, a continuación 15 ciclos de 95°C durante 30 s; 72°C durante 10 s; 65°C durante 1 min; 60°C durante 8 min; 65°C durante 3 min y 72°C durante 30 s; y una extensión final a 72°C durante 2 min) utilizando una concentración de cebador de 14,5 nM de 9600 cebadores inversos etiquetados específicos diana y un cebador inverso específico de adaptador de la biblioteca a 500 nM.

El protocolo de PCR hemi-anidada implicó una segunda amplificación de una dilución del producto de la primera STA durante 15 ciclos de STA (95°C durante 10 min para la activación de la polimerasa inicial, a continuación 15 ciclos de 95°C durante 30 s; 65°C durante 1 min; 60°C durante 5 min; 65°C durante 5 min y 72°C durante 30 s; y una extensión final a 72°C durante 2 min) utilizando una concentración de etiqueta inversa de 1000 nM, y una concentración de 16,6 u nM para cada uno de los 9600 cebadores directos específicos diana.

Una parte alícuota de los productos de la STA se amplificó posteriormente mediante PCR estándar durante 10 ciclos con 1 uM de cebadores directos específicos de etiqueta y cebadores inversos con código de barras para generar bibliotecas de secuenciación con código de barras. Una parte alícuota de cada biblioteca se mezcló con bibliotecas de diferentes códigos de barras y se purificó utilizando una columna de centrifugación.

De este modo, se utilizaron 9600 cebadores en las reacciones de un único pocillo; los cebadores estaban diseñados para focalizar los SNP que se encuentran en los cromosomas 1, 2, 13, 18, 21, X e Y. A continuación, los amplicones fueron secuenciados utilizando un secuenciador ILLUMINA GAIIX. Por cada muestra, se generaron aproximadamente 3,9 millones de lecturas mediante el secuenciador, con 3,7 millones de lecturas correspondientes al genoma (94%), y de ellas, 2,9 millones de lecturas (74%) correspondientes a los SPN etiquetados con una profundidad de lectura promedio de 344 y una mediana de profundidad de lectura media de 255. Se determinó que la fracción fetal de las cuatro muestras era de 9,9%, 18,9%, 16,3%, y 21,2%.

Las muestras del ADN genómico materno y paterno relevantes se amplificaron utilizando un protocolo semi-anidado de 9600-plex y se secuenciaron. El protocolo semi-anidado es diferente en el sentido de que aplica 9600 cebadores directos externos y cebadores inversos etiquetados a 7,3 nM en la primera STA. Las condiciones de termociclado y la composición de la segunda STA, y la adición de códigos de barras a la PCR fueron las mismas que para el protocolo hemi-anidado.

Los datos de secuencia se analizaron utilizando métodos informáticos divulgados en el presente documento y el estado de ploidía se determinó en seis cromosomas para los fetos cuyo ADN estaba presente en las cuatro muestras de plasma materno. Las determinaciones del estado de ploidía para los 28 cromosomas del conjunto se realizaron correctamente con niveles de certeza superiores al 99,2%, salvo en el caso de un cromosoma, en el que se determinó correctamente pero con una certeza del 83%.

La Figura 23 muestra la profundidad de lectura del método de hemi-anidado de 9600-plex junto con la profundidad de lectura del método semi-anidado de 1200-plex descrito en el Experimento 7, aunque el número de SNP con una profundidad de lectura superior a 100, superior a 100 y superior a 400 fue significativamente superior que en el protocolo de 1200-plex. El número de lecturas en el percentil 90 se puede dividir por el número de lecturas en el percentil 10 para proporcionar una métrica sin dimensiones que es indicativa de la uniformidad de la profundidad de lectura; cuando menor es el número, más uniforme (estrecha) es la profundidad de lectura. El ratio medio del percentil 90/percentil 10 es de 11,5 para el método ejecutado en el Experimento 9, aunque es de 5,6 para el método ejecutado en el Experimento 5. Una profundidad de lectura más estrecha para el multiplexado dado de un protocolo es mejor para la eficiencia de secuenciación, dado que se necesitan menos lecturas de secuencia para garantizar que un determinado porcentaje de lecturas se encuentra por encima de un umbral del número de lecturas.

*Experimento 10*

En un experimento se prepararon cuatro muestras de plasma materno y se amplificaron utilizando un protocolo semi-anidado de 9600-plex. Los detalles del Experimento 10 fueron muy similares a los del Experimento 9, salvo por el protocolo de anidado, e incluyendo la identidad de las cuatro muestras. Los estados de ploidía de los 28 cromosomas del conjunto se determinaron correctamente con certezas superiores al 99,7%. 7,6 millones (97%) de las lecturas correspondían al genoma y 6,3 millones (80%) de las lecturas correspondían a los SNP focalizados. La profundidad de lectura promedio fue de 751 y la profundidad media de lectura fue de 396.

*Experimento 11*

En un experimento, se dividieron tres muestras de plasma materno en cinco porciones iguales y cada porción se amplificó utilizando 2400 cebadores multiplexados (cuatro porciones) o 1200 cebadores multiplexados (una porción) y se amplificaron utilizando un protocolo semi-anidado, para un total de 10 800 cebadores. Tras la amplificación, las porciones se agruparon para la secuenciación. Los detalles del Experimento 11 fueron muy similares a los del

Experimento 9, salvo por el protocolo de anidado y el método de división y reagrupación. Las determinaciones del estado de ploidía para los 21 cromosomas del conjunto se realizaron correctamente con niveles de certeza superiores al 99,7%, salvo en el caso de una determinación errónea, donde la certeza era del 83%. 3,4 millones de lecturas correspondían a los SNP focalizados, la profundidad de lectura promedio fue de 404 y la profundidad media de lectura fue de 258.

*Experimento 12*

En un experimento, se dividieron cuatro muestras de plasma materno en cuatro porciones iguales y cada porción se amplificó utilizando 2400 cebadores multiplexados y se amplificó utilizando un protocolo semi-anidado, para un total de 9600 cebadores. Tras la amplificación, las porciones se agruparon para la secuenciación. Los detalles del Experimento 12 fueron muy similares a los del Experimento 9, salvo por el protocolo de anidado y el método de división y reagrupación. Las determinaciones del estado de ploidía para los 28 cromosomas del conjunto se realizaron correctamente con niveles de certeza superiores al 97%, salvo en el caso de una determinación errónea, donde la certeza era del 78%. 4,5 millones de lecturas correspondían a los SNP focalizados, la profundidad de lectura promedio fue de 535 y la profundidad media de lectura fue de 412.

*Experimento 13*

En un experimento, se prepararon cuatro muestras de plasma materno y se amplificaron utilizando un protocolo triplemente hemi-anidado de 9600-plex, para un total de 9600 cebadores. Los detalles del Experimento 12 fueron muy similares a los del Experimento 9, salvo por el protocolo de anidado que implicó tres rondas de amplificación; las tres rondas implicaron 15, 10 15 ciclos de STA, respectivamente. Las determinaciones del estado de ploidía para 27 de los 28 cromosomas del conjunto se realizaron correctamente con niveles de certeza superiores al 99,9%, salvo en un caso en el que se determinó correctamente con una certeza del 94,6% y un caso en el que se determinó de forma incorrecta con una certeza del 80,8%. 3,5 promedio fue de 414 y la profundidad media de lectura fue de 249.

*Experimento 14*

En un experimento, se amplificaron 45 conjuntos de células utilizando un protocolo semi-anidado de 1200-plex, se secuenciaron y se determinaron los estados de ploidía en tres cromosomas. Cabe señalar que este experimento pretende simular las condiciones de realizar un diagnóstico genético previo a un trasplante con biopsias de una única célula de embriones de tres días o biopsias de trofectoderma de embriones de cinco días. 15 células únicas individuales y 30 conjuntos de tres células se colocaron en 45 tubos de reacción individuales para un total de 45 reacciones donde cada reacción contenía células de una única línea de células, pero las diferentes reacciones contenían células de diferentes líneas de células. Las células se prepararon en 5 ul de tampón de lavado, se sometieron a lisis añadiendo 5 ul de tampón de lisis ARCTURUS PICOPURE (APPLIED BIOSYSTEMS) y se incubaron a 56°C durante 20 min, 95ºC durante 10 min.

El ADN de las células únicas/tres células se amplificó con 25 ciclos de STA (95°C durante 10 min para la activación de la polimerasa inicial, a continuación 25 ciclos de 95°C durante 30 s; 72°C durante 10 s; 65°C durante 1 min; 60°C durante 8 min; 65°C durante 3 min y 72°C durante 30 s; y una extensión final a 72°C durante 2 min) utilizando una concentración de cebador de 50 nM de 1200 cebadores inversos etiquetados y directos específicos diana.

El protocolo de PCR semi-anidada implicó tres segundas amplificaciones paralelas de una dilución del producto de la primera STA durante 20 ciclos de STA (95°C durante 10 min para la activación de la polimerasa inicial, a continuación 15 ciclos de 95°C durante 30 s; 65°C durante 1 min; 60°C durante 5 min; 65°C durante 5 min y 72°C durante 30 s; y una extensión final a 72°C durante 2 min) utilizando una concentración de cebador inverso específico de etiqueta de 1000 nM, y una concentración de 60 nM para cada uno de los 400 cebadores directos anidados específicos diana. En las tres reacciones de 400-plex paralelas se amplificó el total de 1200 dianas amplificadas en la primera STA.

Una parte alícuota de los productos de la STA se amplificó posteriormente mediante PCR estándar durante 15 ciclos con 1 uM de cebadores directos específicos de etiqueta y cebadores inversos con código de barras para generar bibliotecas de secuenciación con código de barras. Una parte alícuota de cada biblioteca se mezcló con bibliotecas de diferentes códigos de barras y se purificó utilizando una columna de centrifugación.

De este modo, se utilizaron 1200 cebadores en las reacciones de células únicas; los cebadores estaban diseñados para focalizar los SNP que se encuentran en los cromosomas 1, 21 y X. A continuación, los amplicones fueron secuenciados utilizando un secuenciador ILLUMINA GAIIX. Por cada muestra se generaron aproximadamente 3,9 millones de lecturas a través del secuenciador, con 500 000 a 800 000 lecturas correspondientes al genoma (74% - 94% de todas las lecturas por muestra).

Las muestras de ADN genómico materno y paterno relevantes de las líneas de células se analizaron utilizando el mismo grupo de ensayo semi-anidado de 1200-plex con un protocolo similar con menos ciclos y una segunda STA de 1200-plex, y se secuenciaron.

Los datos de secuencia se analizaron utilizando métodos informáticos divulgados en el presente documento y el estado de ploidía se determinó en los tres cromosomas para las muestras.

La Figura 24 muestra unos ratios de profundidad de lectura normalizados (eje vertical) para seis muestras en tres cromosomas (1 = crom 1, 2 = crom 21; 3 = crom X). Los ratios se ajustaron para que fuesen iguales al número de lecturas correspondientes a ese cromosoma, se normalizaron y se dividieron por el número de lecturas correspondientes a ese cromosoma calculando la media sobre tres pocillos que comprendían cada uno tres células 46XY. Los tres conjuntos de puntos de datos correspondientes a las reacciones de 46XY se espera que presenten unos ratios de1:1. Los tres conjuntos de puntos de datos correspondientes a las células 47XX+21 se espera que presenten unos ratios de 1:1 para el cromosoma 1,1,5:1 para el cromosoma 21, y 2:1 para el cromosoma X.

La Figura 25 muestra los ratios alélicos representados para los tres cromosomas (1, 21, X) para tres reacciones. La reacción de la parte inferior izquierda muestra una reacción en las tres células 46XY. La región izquierda son los ratios alélicos del cromosoma 1, la región media son los ratios alélicos del cromosoma 21 y la región derecha son los ratios alélicos para el cromosoma X. Para las células 46XY, para el cromosoma esperamos observar ratios de 1, 0,5 y 0 correspondientes a los genotipos de SNP AA, AB y BB. Para las células 46XY, para el cromosoma 21 esperamos observar ratios de 1, 05, y 0 correspondientes a los genotipos de SNP AA, AB y BB. Para las células 46XY, para el cromosoma X esperamos observar ratios de 1 y 0 correspondientes a los genotipos de SNP A y B. La reacción de la parte inferior derecha muestra una reacción en las tres células 47XX+21. Los ratios alélicos están segregados por cromosoma como en el gráfico de la parte inferior izquierda. Para las células 47XX+21, para el cromosoma 1 esperamos observar ratios de 1, 0,5, y 0 correspondientes a los genotipos de SNP AA, AB y BB. Para las células 47XX+21, para el cromosoma 21 esperamos observar ratios de 1, 0,67, 0,33 y 0 correspondientes a los genotipos de SNP AAA, AAB, ABB y BBB. Para las células 47XX+21, para el cromosoma X esperamos observar ratios de 1, 05, y 0 correspondientes a los genotipos de SNP AA, AB y BB. La representación de la parte superior derecha se realizó sobre una reacción que comprende 1 ng de ADN genómico de la línea de células 47XX+21. La Figura 26 muestra los mismos gráficos que la Figura 25 pero para las reacciones realizadas sobre una única célula. El gráfico de la izquierda fue una reacción que contenía una célula 47XX+21, y el gráfico de la derecha fue para una reacción que contenía una célula 46XX.

A partir de los gráficos mostrados en la Figura 25 y la Figura 26, resulta visualmente evidente que hay dos grupos de puntos para los cromosomas en los que esperamos observar ratios de 1 y 0; tres grupos de puntos para los cromosomas en los que esperamos observar ratios de 1, 0,5 y 0, y cuatro grupos de puntos para los cromosomas en los que esperamos observar ratios de 1, 0,67, 0,33 y 0. El algoritmo de Parental Support fue capaz de realizar determinaciones correctas en los tres cromosomas para la totalidad de las 45 reacciones.

**REIVINDICACIONES**

1.Un método de amplificación de loci diana en una muestra de ácido nucleico, que consiste en lo siguiente:

a) realización de una reacción en cadena de la polimerasa (PCR) multiplexada en una muestra de ácido nucleico que comprende loci diana para amplificar simultáneamente al menos 1000 loci diana distintos utilizando i) al menos 1000 pares de cebadores diferentes o ii) al menos 1000 cebadores específicos diana y un cebador específico de etiqueta o universal, en un único volumen de reacción para producir productos amplificados que comprenden amplicones diana; y

b) secuenciación de los productos amplificados utilizando secuenciación de alto rendimiento, donde la concentración de cada cebador de los pares de cebadores o cada cebador específico diana tiene menos de 20 nM, y donde la longitud del paso de hibridación de la amplificación por PCR multiplexada supera los 10 minutos.

2.El método de la reivindicación 1, que comprende la obtención de cebadores para su uso en el paso a), mediante la identificación empíricamente o in silico de uno o más cebadores que forman un primer dímero con otro cebador de una biblioteca de potenciales cebadores con la frecuencia más elevada, eliminando el cebador o los cebadores identificados de la biblioteca de potenciales cebadores, y utilizando los cebadores que continúan en la biblioteca en el paso a).

3.El método de la reivindicación 1, que comprende la realización de la amplificación universal de ácidos nucleicos en la muestra antes del paso a), donde la amplificación universal comprende opcionalmente la PCR universal, la amplificación del genoma completo, la PCR mediada por unión, la PCR con cebadores de oligonucleótidos degenerados o la amplificación por desplazamiento múltiple.

4.El método de la reivindicación 1, donde la PCR multiplexada comprende la PCR anidada, semi-anidada o hemi-anidada; o donde cada par de cebadores comprende un cebador directo y un cebador inverso en el que el extremo 3' del cebador directo e inverso está diseñado para hibridarse a una región de ADN separada de un punto polimórfico de un locus diana por un pequeño número de bases, donde ese pequeño número es de entre 1 y 20 bases.

5.El método de la reivindicación 1, donde cada par de cebadores comprende un cebador directo y un cebador inverso en el que el extremo 3' del cebador directo e inverso está diseñado para hibridarse a una región de ADN separada de un punto polimórfico de un locus diana por un pequeño número de bases, donde ese pequeño número es de entre 2 y 20 bases.

6.El método de la reivindicación 1, que comprende simultáneamente la amplificación de al menos 5000 loci diana distintos.

7.El método de la reivindicación 1, donde al menos el 90% de los productos amplificados corresponden a los loci diana.

8.El método de la reivindicación 1, donde la amplificación por PCR multiplexada comprende 20 ciclos de PCR y el grado medio de sesgo alélico entre la muestra y los amplicones diana no es superior a un factor de 1,2.

9.El método de la reivindicación 1, donde la muestra comprende ADN materno de la madre embarazada de un feto y ADN fetal, y donde el método comprende la determinación de la presencia o ausencia de una anomalía de un cromosoma fetal a partir de los datos de la secuenciación.

10.El método de la reivindicación 1, donde los loci diana están presentes en el genoma humano.

11.El método de la reivindicación 1, donde los loci diana comprenden polimorfismos humanos de un único nucleótido.

12.El método de la reivindicación 1, donde la longitud de los amplicones diana es inferior a 100 nucleótidos.

13.El método de la reivindicación 1, donde la muestra de ácido nucleico comprende ADN de un tumor, trasplante o feto.

14.El método de la reivindicación 1, donde la muestra comprende ADN de una única célula.

15. El método de la reivindicación 1, que consiste en lo siguiente:

a) realización de una PCR multiplexada en una muestra de ácido nucleico que comprende moléculas de ADN con una longitud media de menos de 200 pares de bases que comprende loci diana para amplificar simultáneamente al menos 1000 loci diana distintos utilizando i) al menos 1000 pares de cebadores diferentes, o ii) al menos 1000 cebadores específicos diana y un cebador específico de etiqueta o universal, en un único volumen de reacción para producir productos amplificados que comprenden amplicones diana de menos de 100 nucleótidos de longitud; donde la concentración de cada cebador de los pares de cebadores o de cada cebador específico diana es inferior a 20 nM; y donde la longitud del paso de hibridación de la amplificación por PCR multiplexada es superior a 10 minutos; y

b) secuenciación de los productos amplificados utilizando una secuenciación de alto rendimiento.

El método de la reivindicación 4, donde cada par de cebadores comprende un cebador directo interno en el que el extremo 3' del cebador directo interno está diseñado para hibridarse en una región de ADN separada de un punto polimórfico de un locus diana por un pequeño número de bases, donde ese pequeño número es de entre 2 y 60 bases.

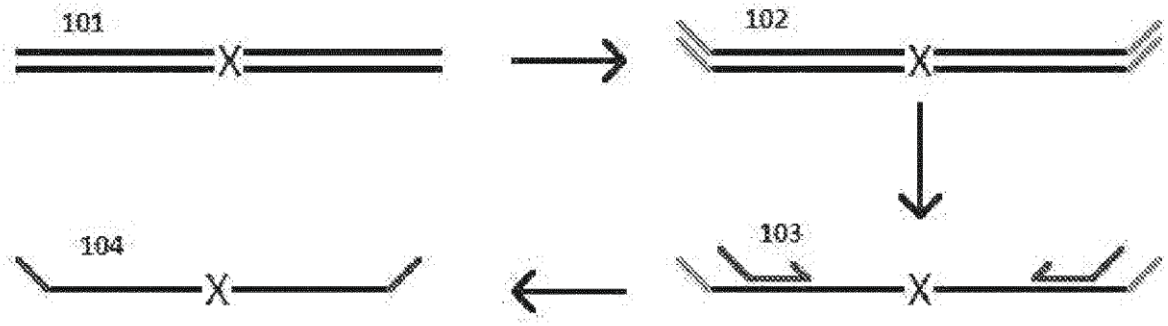



Fig 1

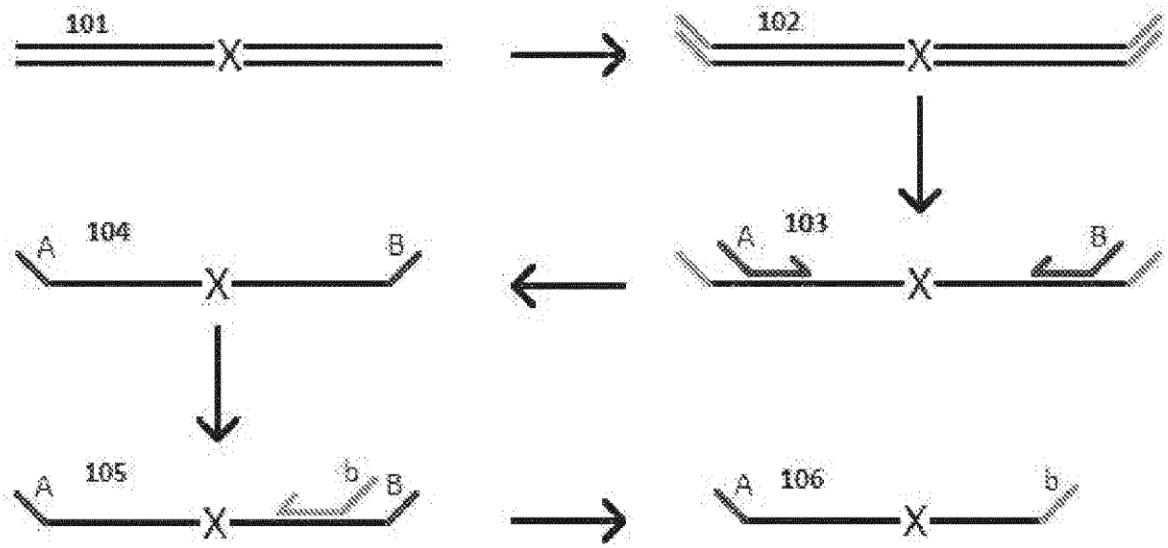



Fig 2

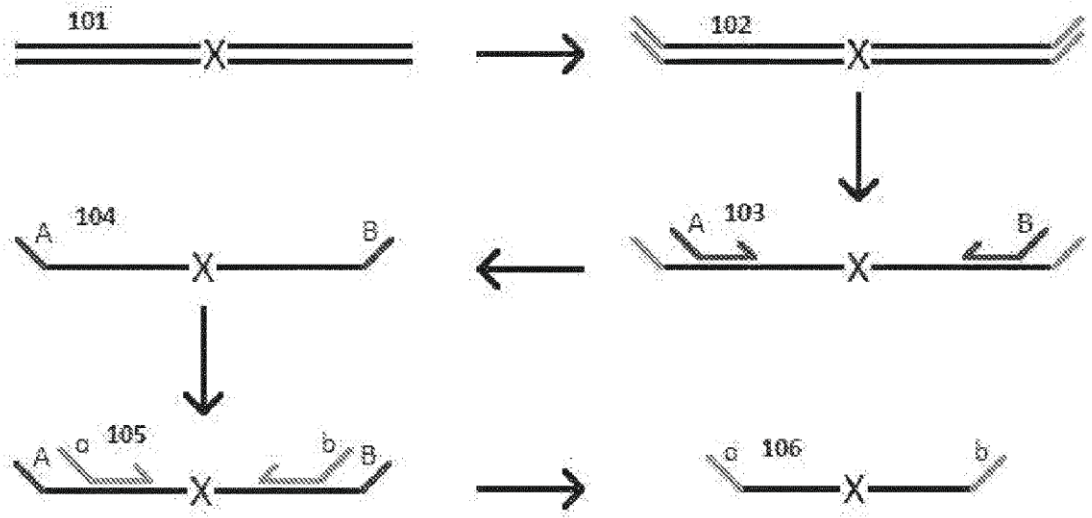

101
102
103
A
B
104
A
B
105
A
a
b
B
106
a
b

Fig 3

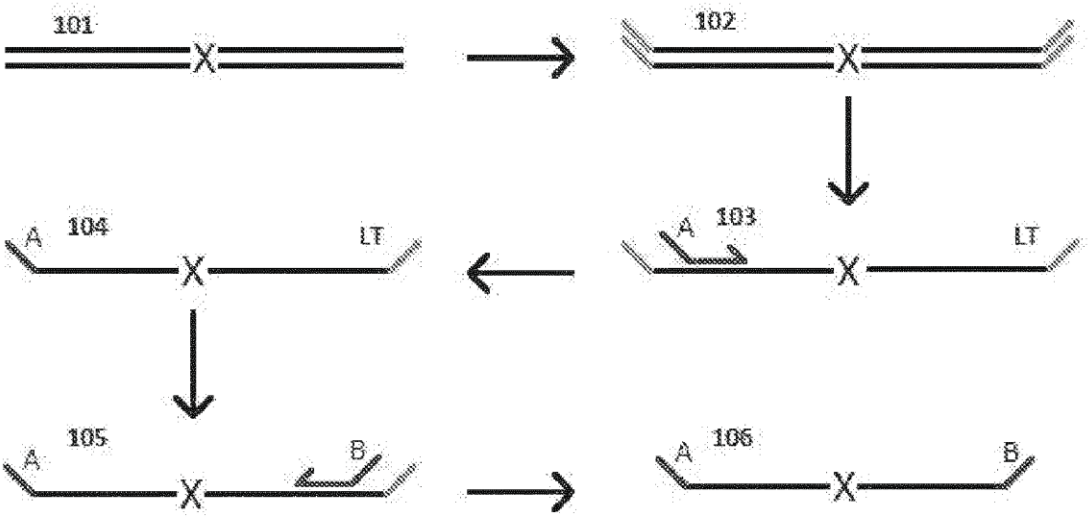

101
102
103
A
LT
104
A
LT
105
A
B
106
A
B

Fig 4

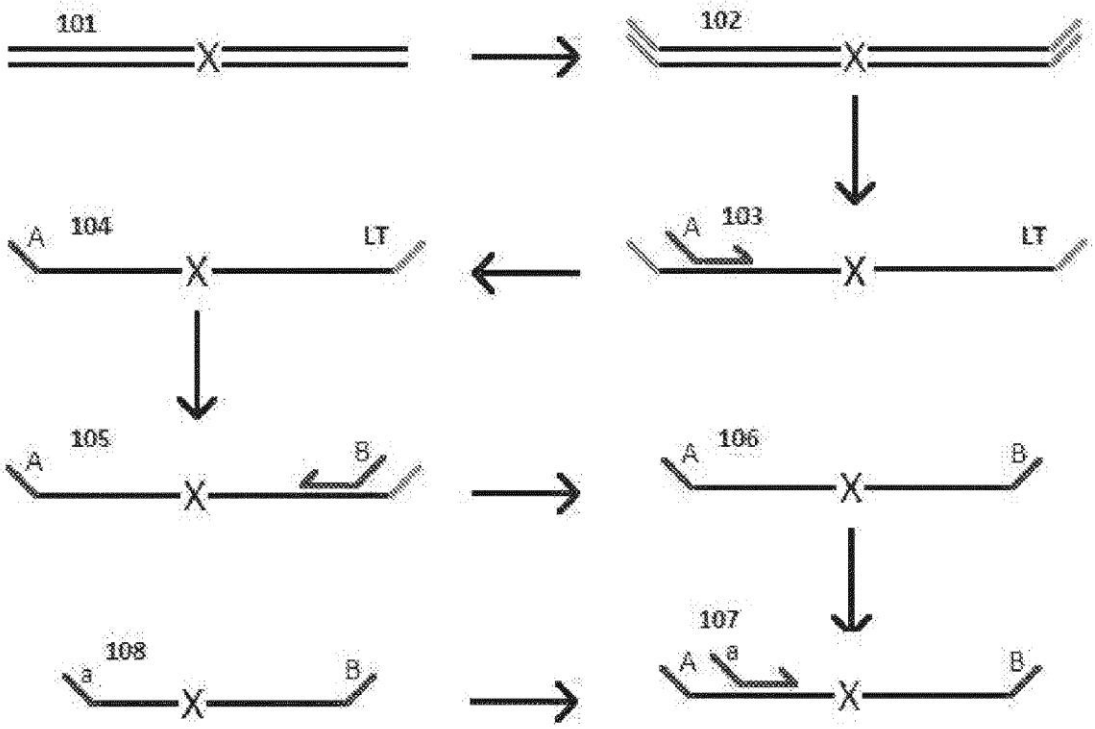

101
102
A
103
LT
A
104
LT
A
105
B
A
106
B
108
a
B
107
A
a
B

Fig 5

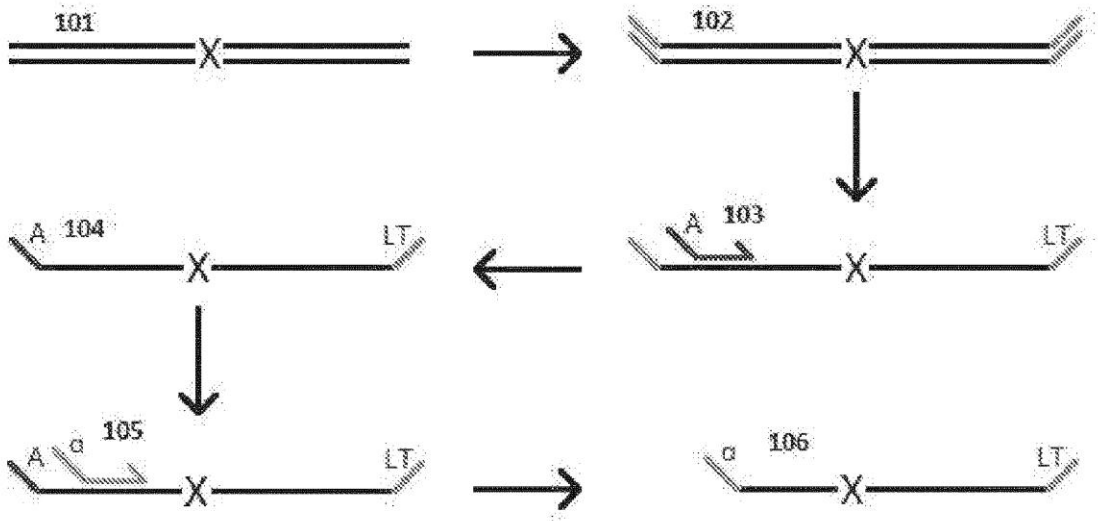

101
102
A
103
LT
A
104
LT
A
a
105
LT
a
106
LT

Fig 6

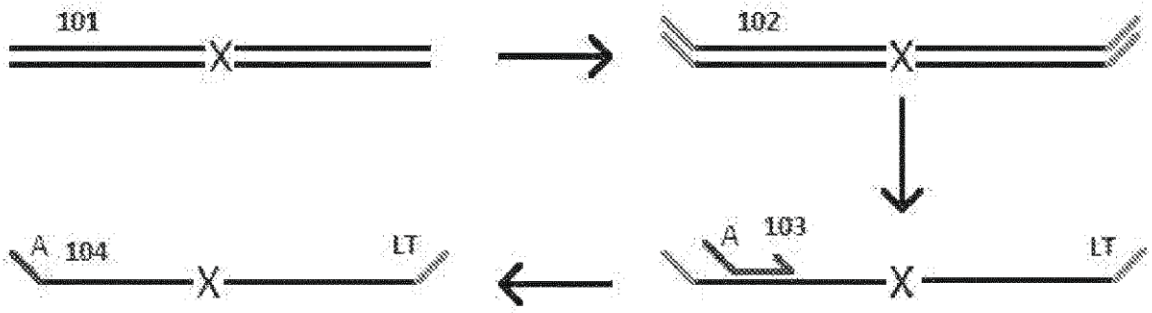



Fig 7

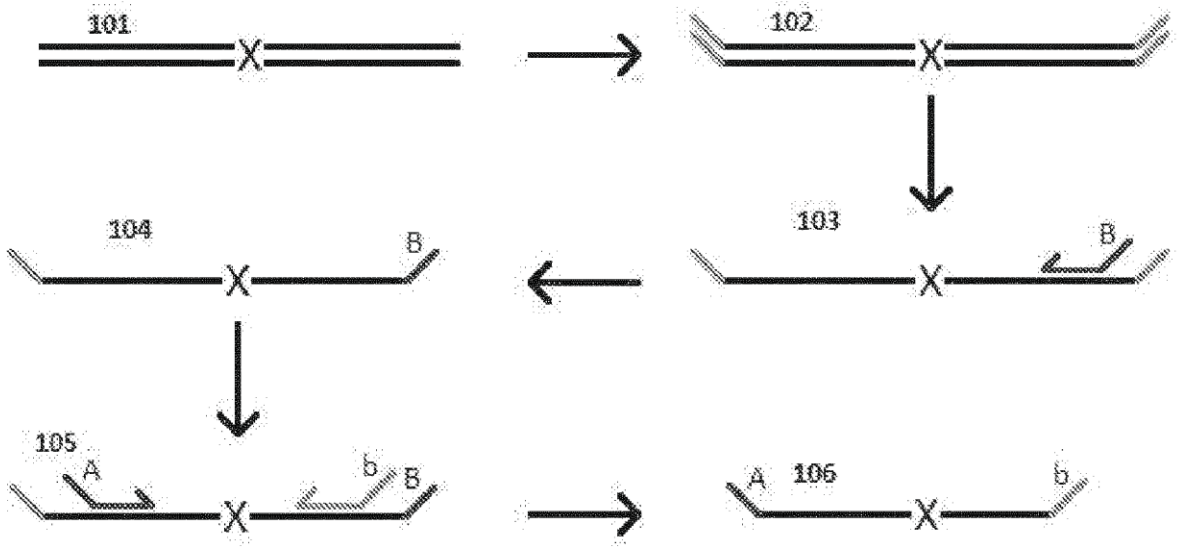



Fig 8

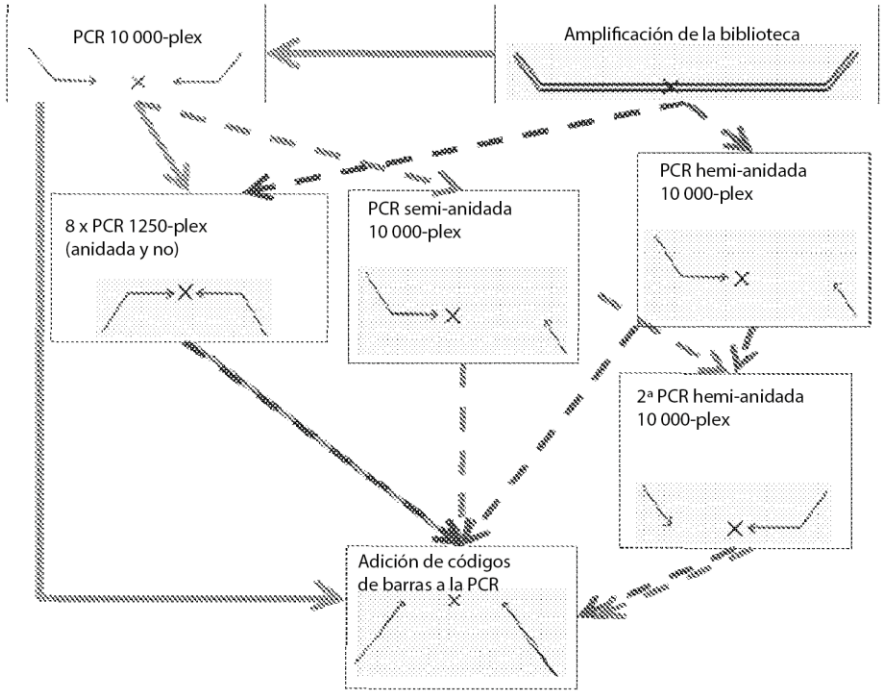

PCR 10 000-plex
Amplificación de la biblioteca
PCR hemi-anidada
10 000-plex
8 x PCR 1250-plex
(anidada y no)
PCR semi-anidada
10 000-plex
2ª PCR hemi-anidada
10 000-plex
Adición de códigos
de barras a la PCR

Fig 9

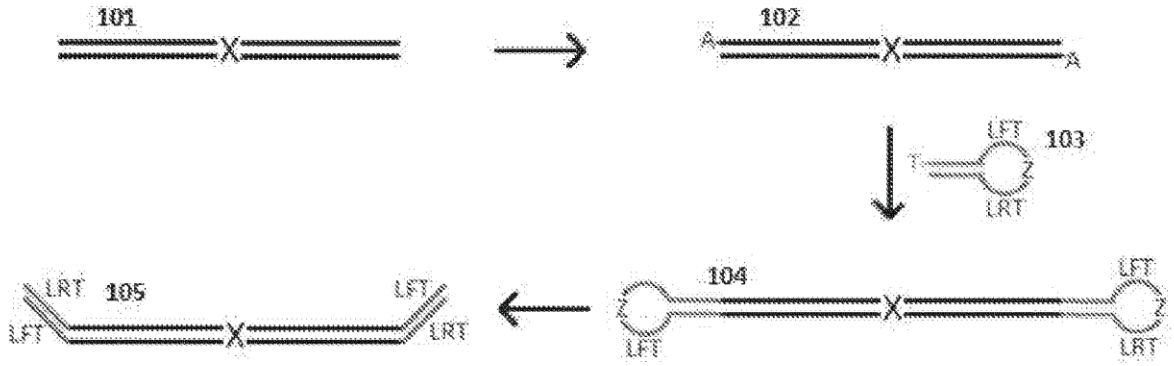

101
102
103
LFT
LRT
104
105

Fig 10

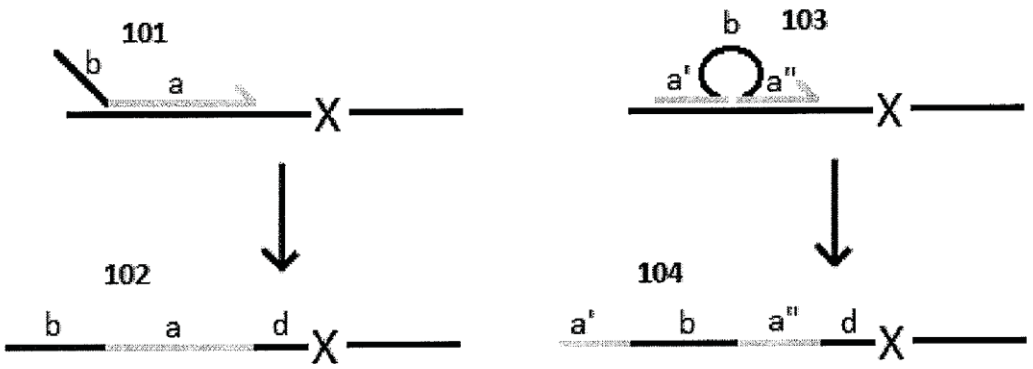



Fig 11

La secuencia del adaptador de secuenciación se encuentra dentro de la secuencia del cebador y flanqueada por una secuencia específica de diana a ambos lados. 10 bases son específicas de diana en el extremo 3' de cada cebador. Los cebadores se probaron con éxito en una PCR en tiempo real. Para la secuenciación, esto reduce el número de bases del cebador que es necesario secuenciar.

| | Etiqueta int. | | Id. de sec. |
|---|---|---|---|
| rs8130564 | 1.10 | AACTCACATAGC*ACACGACGCTCTTCCGATCT*TGCAAGCACA | 1 |
| rs2832093 | 2.10 | TCCTCTGTG*ACACGACGCTCTTCCGATCT* CCCTGCTCTT | 2 |
| rs12011281 | 3.10 | tcctctctctACACGACGCTCTTCCGATCTcGGGCTGTCA | 3 |
| rs6719561 | 4.10 | TACATCCTTGAGACACGACGCTCTTCCGATCTGCTGTGCAGT | 4 |
| rs10187018 | 5.10 | tttgcttgagctACACGACGCTCTTCCGATCTcgggagtttc | 5 |
| rs10460481 | 6.10 | gtcttatggtggACACGACGCTCTTCCGATCTcaaagccagt | 6 |

La secuencia del adaptador de secuenciación se encuentra dentro de la secuencia del cebador y flanqueada por una secuencia específica de diana a ambos lados. La etiqueta interna se forma en una estructura de horquilla con 10 bases complementarias en cualquiera de los extremos. Esto acerca mucho los extremos específicos de etiqueta del cebador y perjudica la unión no específica a la "etiqueta interna". 10 bases son específicas de diana en el extremo 3' de cada cebador. Los cebadores se testaron con éxito en una PCR en tiempo real.

| | Etiqueta int. loop | | Id. de sec. |
|---|---|---|---|
| rs8130564 | 1.10 | AACTCACATAGC*tgatcggtACACGACGCTCTTCCGATCT*TGCAAGCACA | 7 |
| rs2832093 | 2.10 | TCCTCTGTG*tgatcggtACACGACGCTCTTCCGATCT*CCCTGCTCTT | 8 |
| rs12011281 | 3.10 | tcctctctcttgatcggtACACGACGCTCTTCCGATCTcGGGCTGTCA | 9 |
| rs6719561 | 4.10 | TACATCCTTGAGtgatcggtACACGACGCTCTTCCGATCTGCTGTGCAGT | 10 |
| rs10187018 | 5.10 | tttgcttgagcttgatcggtACACGACGCTCTTCCGATCTcgggagtttc | 11 |
| rs10460481 | 6.10 | gtcttatggtggtgatcggtACACGACGCTCTTCCGATCTcaaagccagt | 12 |

Fig 12

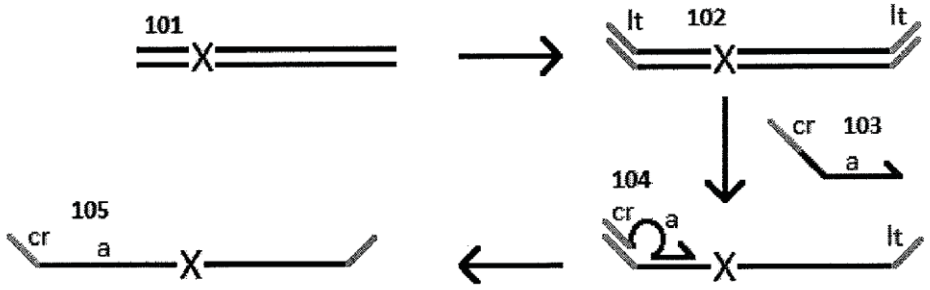



Fig 13

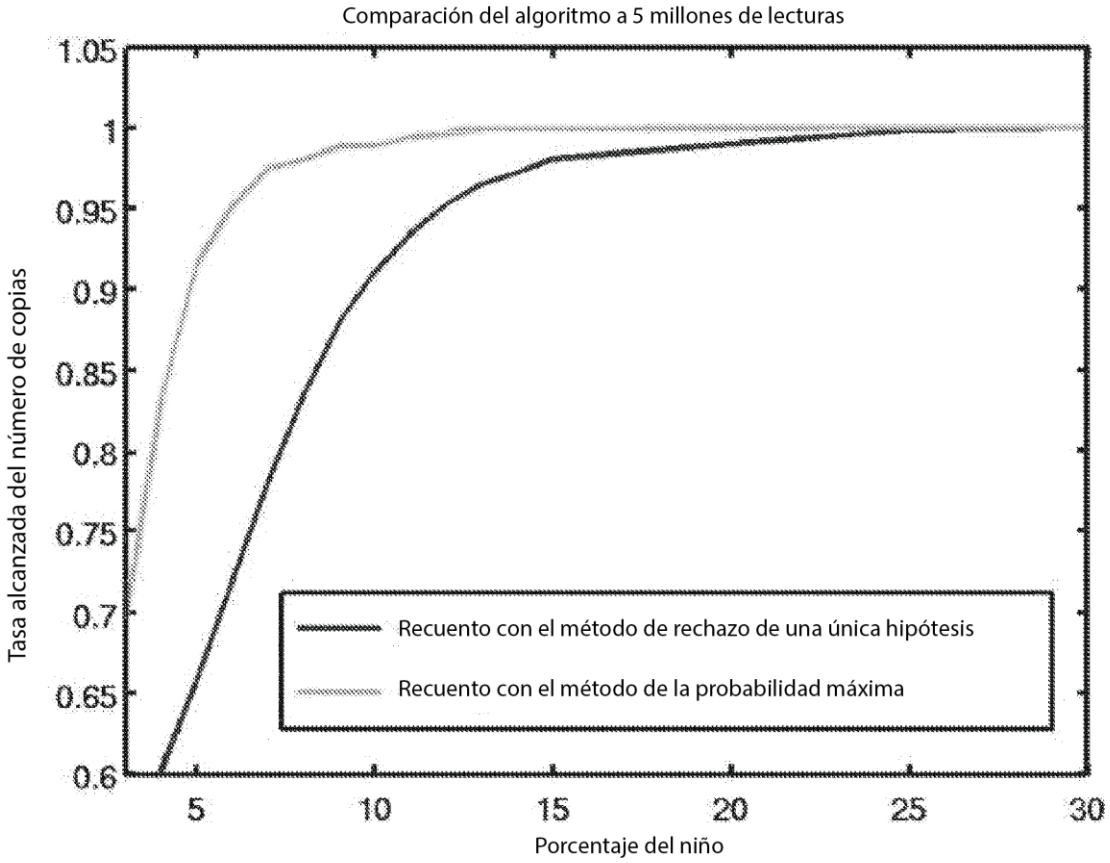

Comparación del algoritmo a 5 millones de lecturas
Tasa alcanzada del número de copias
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
5
10
15
20
25
30
Porcentaje del niño
Recuento con el método de rechazo de una única hipótesis
Recuento con el método de la probabilidad máxima

Fig 14

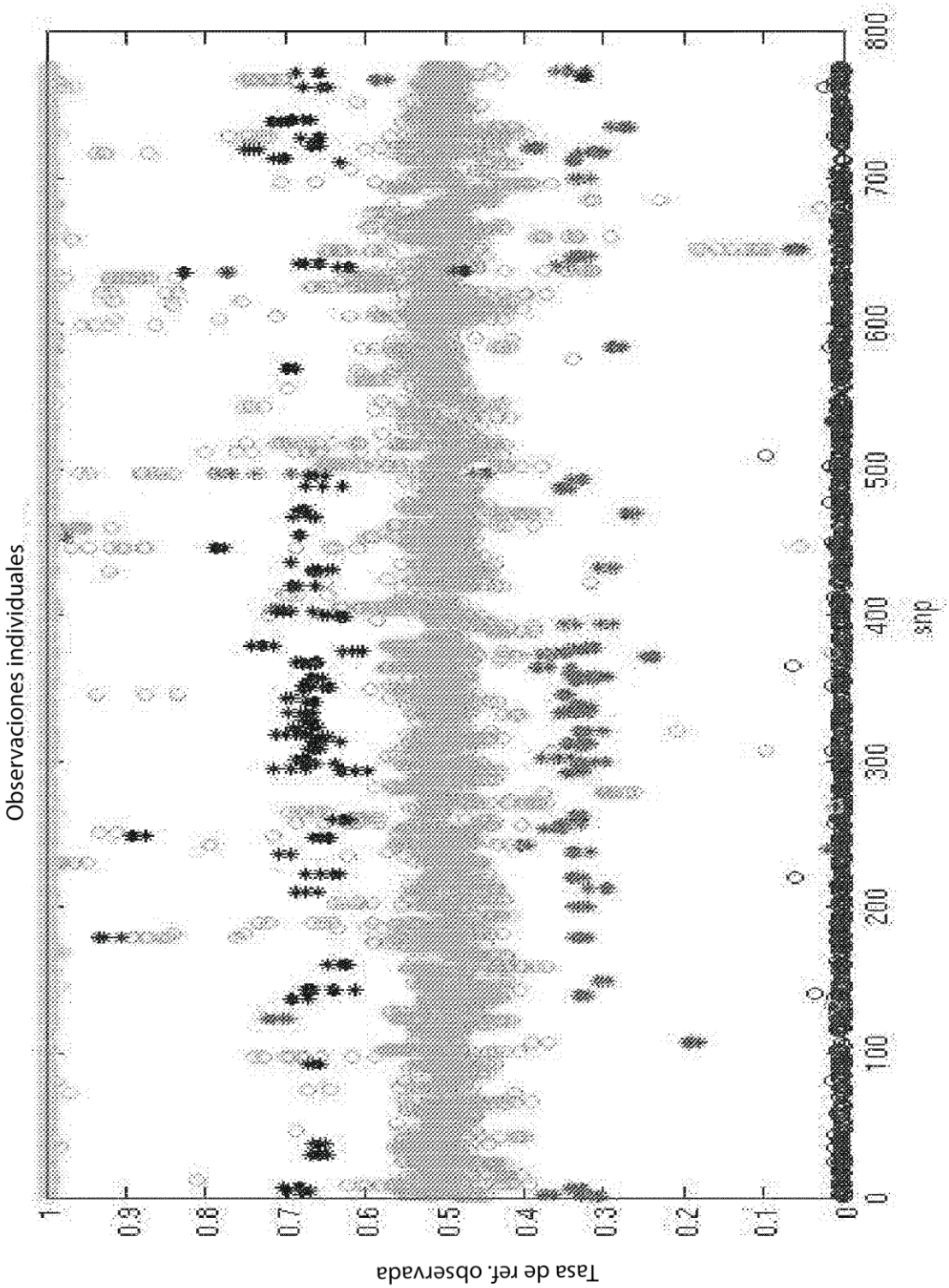

Observaciones individuales
Tasa de ref. observada
snp
0
100
200
300
400
500
600
700
800
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Fig 15

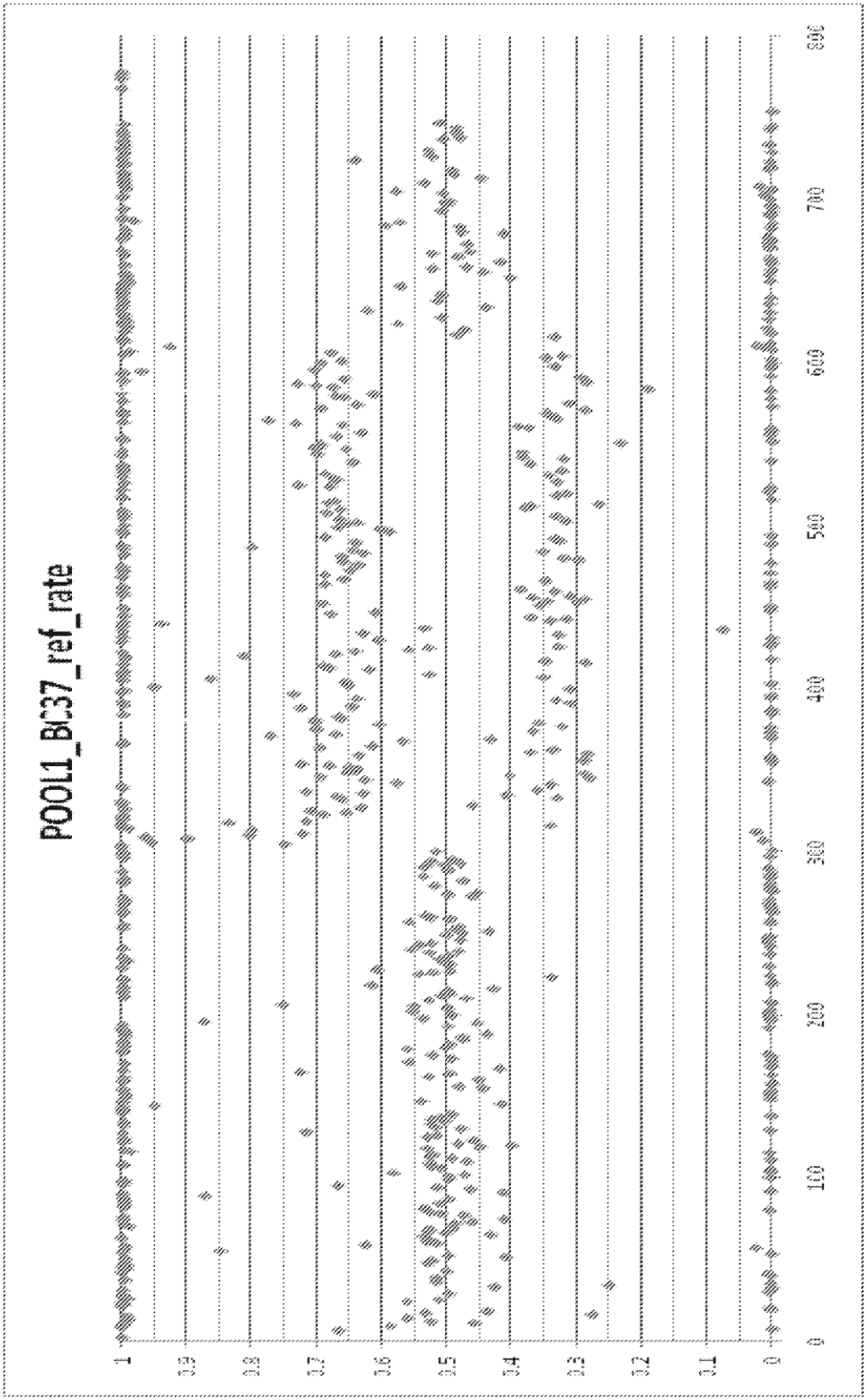

POOL1_BC37_ref_rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
600
700
800

Fig 16

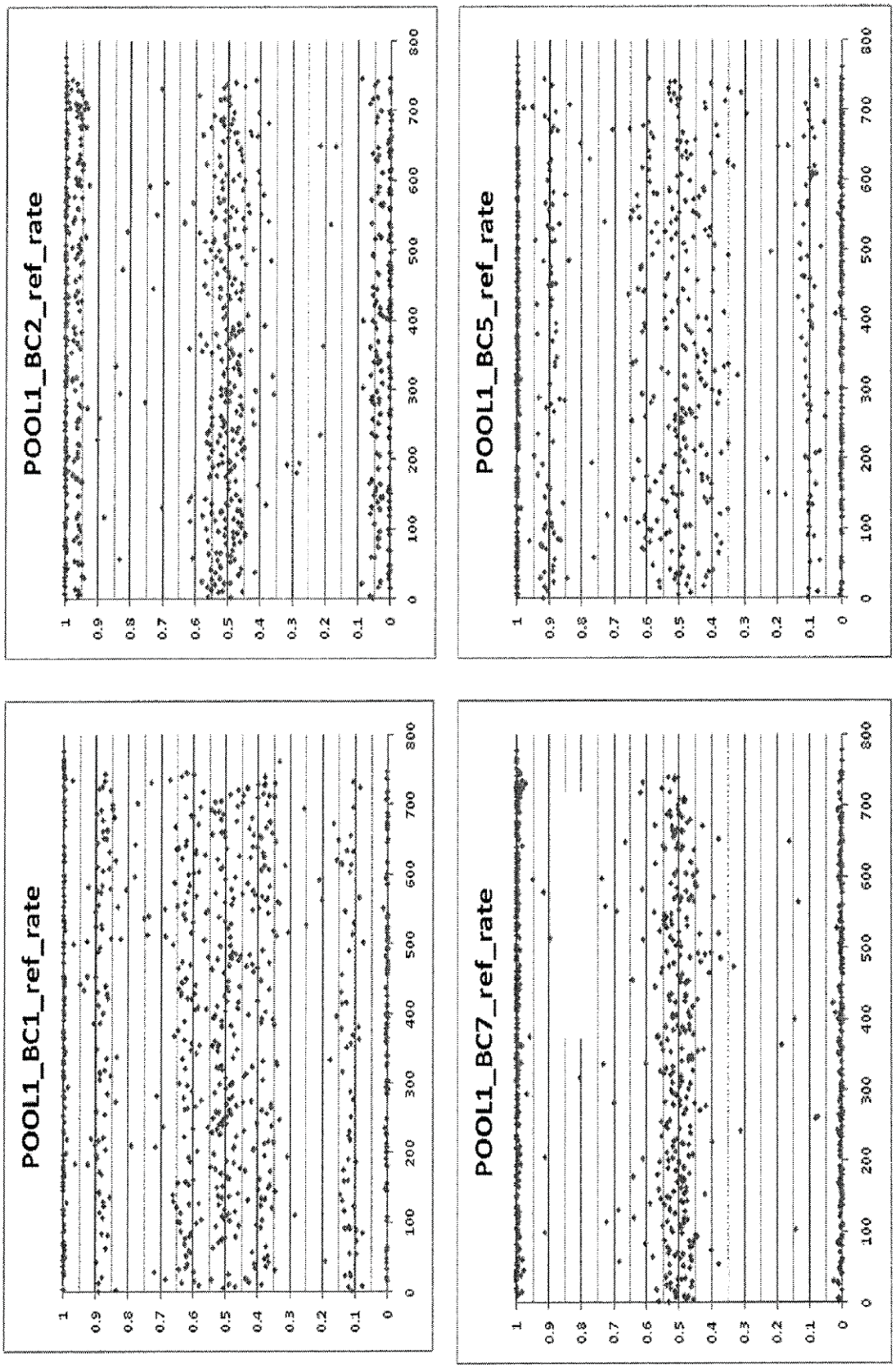

POOL1_BC1_ref_rate
POOL1_BC2_ref_rate
POOL1_BC7_ref_rate
POOL1_BC5_ref_rate

Fig 17

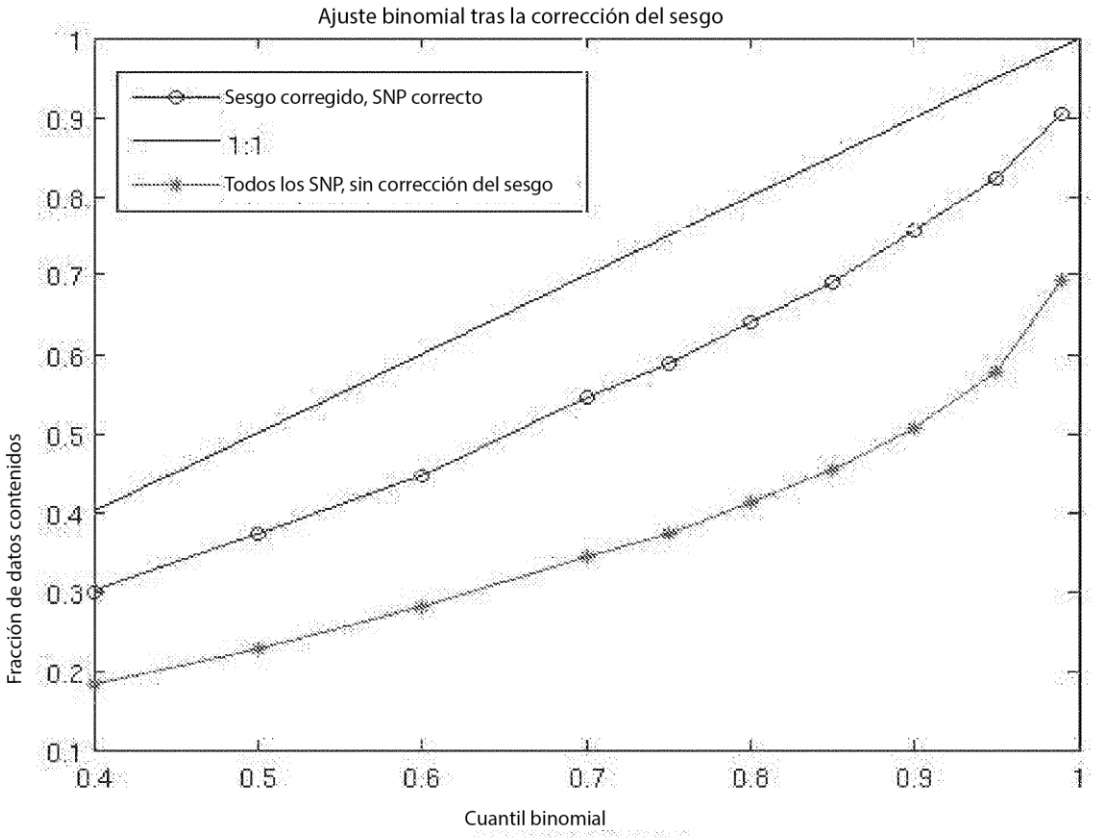

Ajuste binomial tras la corrección del sesgo
Sesgo corregido, SNP correcto
1:1
Todos los SNP, sin corrección del sesgo
Fracción de datos contenidos
Cuantil binomial
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.4
0.5
0.6
0.7
0.8
0.9
1

Fig 18

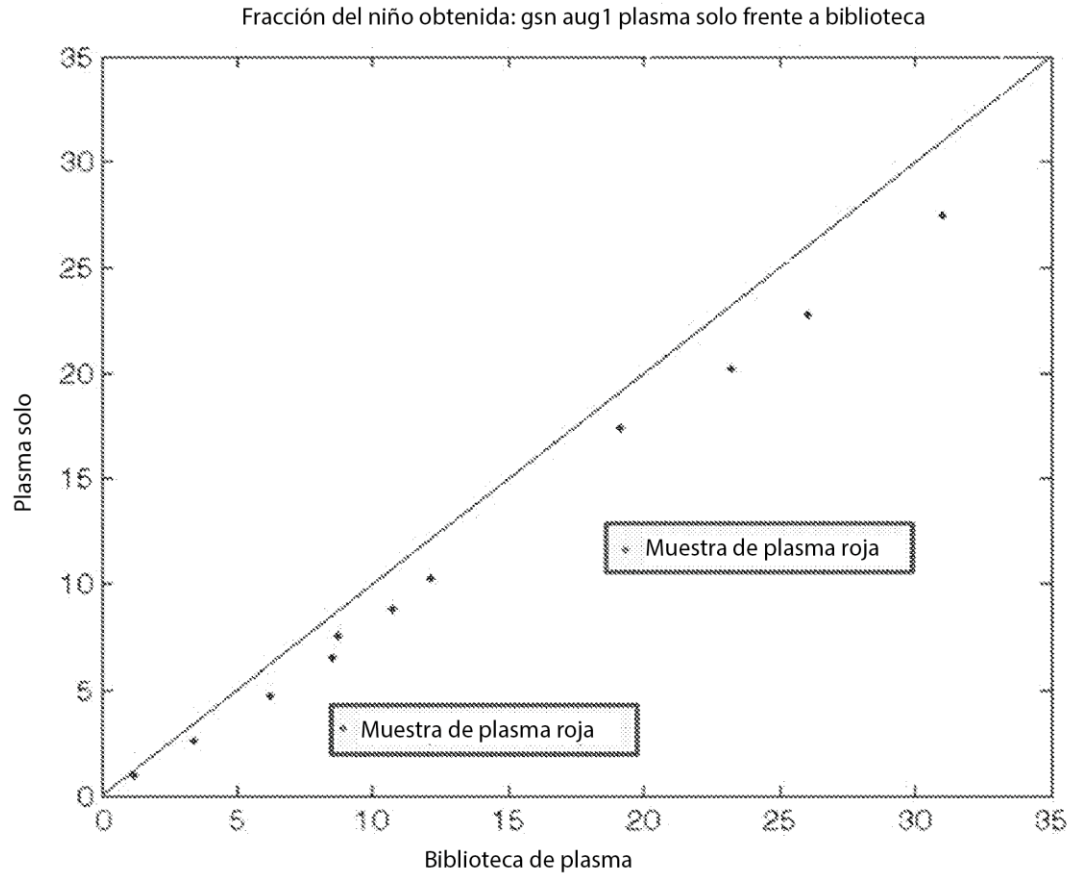

Fracción del niño obtenida: gsn aug1 plasma solo frente a biblioteca
Plasma solo
Biblioteca de plasma
Muestra de plasma roja
Muestra de plasma roja
0
5
10
15
20
25
30
35

Fig 19

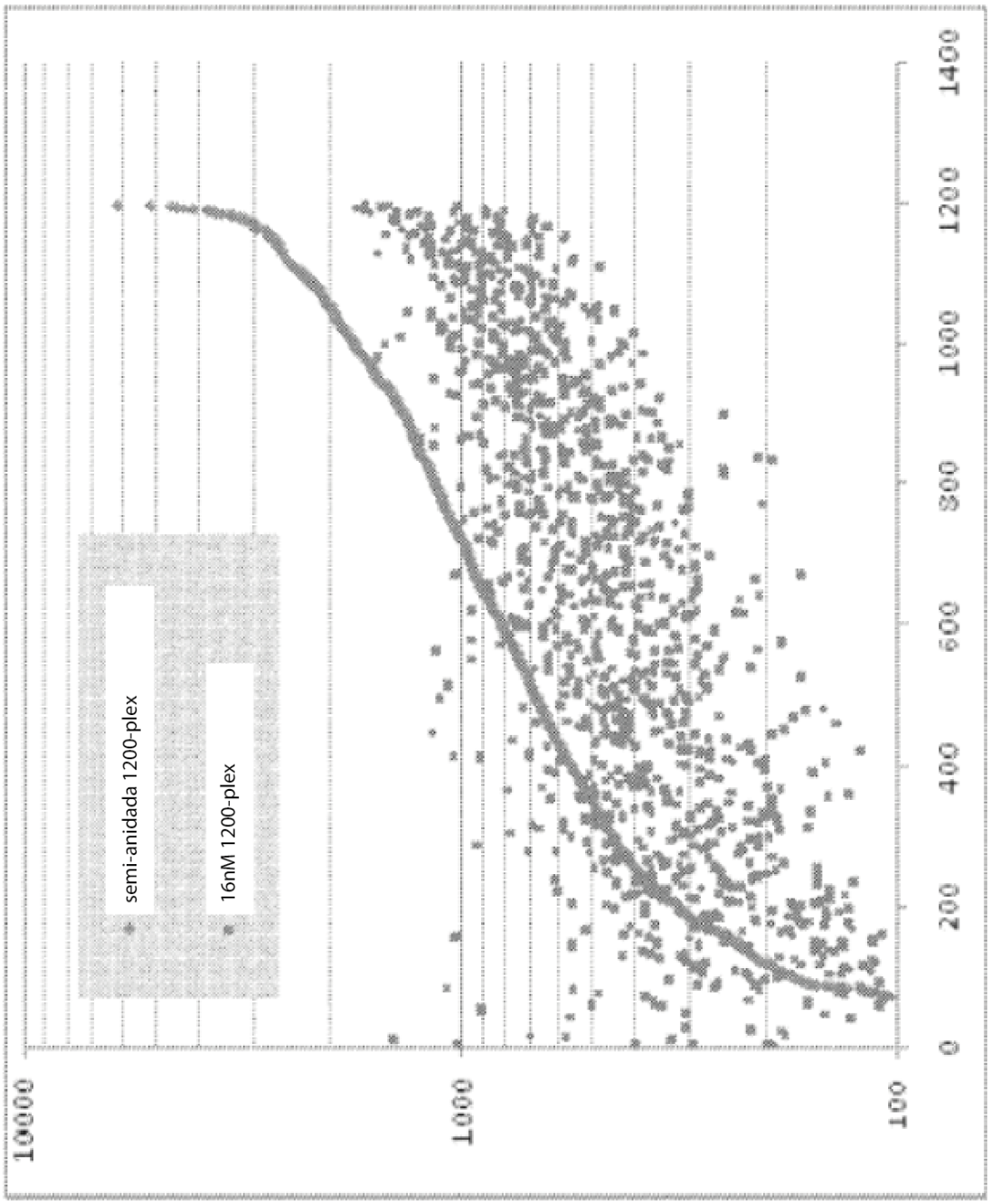

semi-anidada 1200-plex
16nM 1200-plex
10000
1000
100
0
200
400
600
800
1000
1200
1400

Fig 20

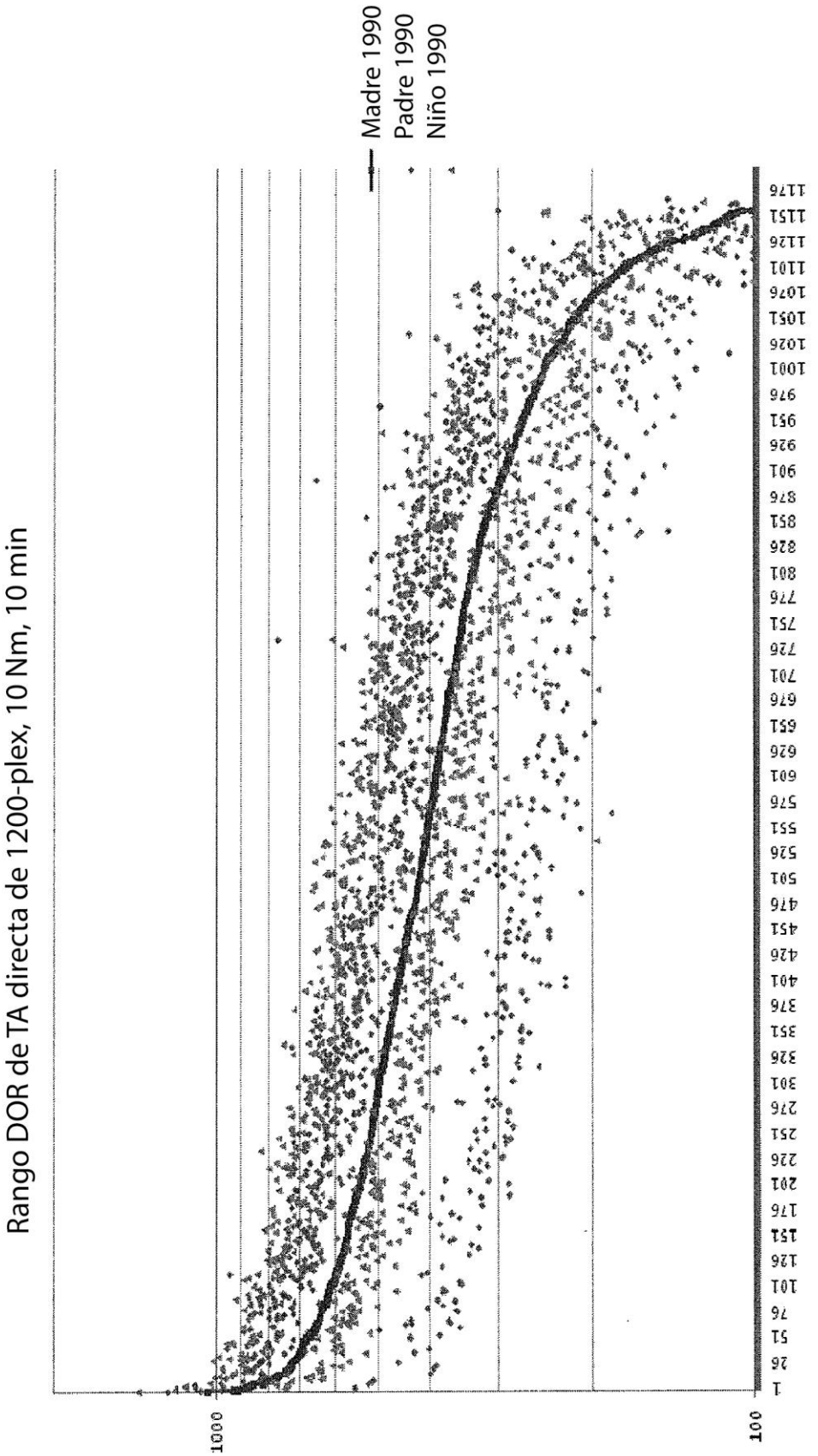

Rango DOR de TA directa de 1200-plex, 10 Nm, 10 min
Madre 1990
Padre 1990
Niño 1990
1000
100
1
26
51
76
101
126
151
176
201
226
251
276
301
326
351
376
401
426
451
476
501
526
551
576
601
626
651
676
701
726
751
776
801
826
851
876
901
926
951
976
1001
1026
1051
1076
1101
1126
1151
1176

Fig 21

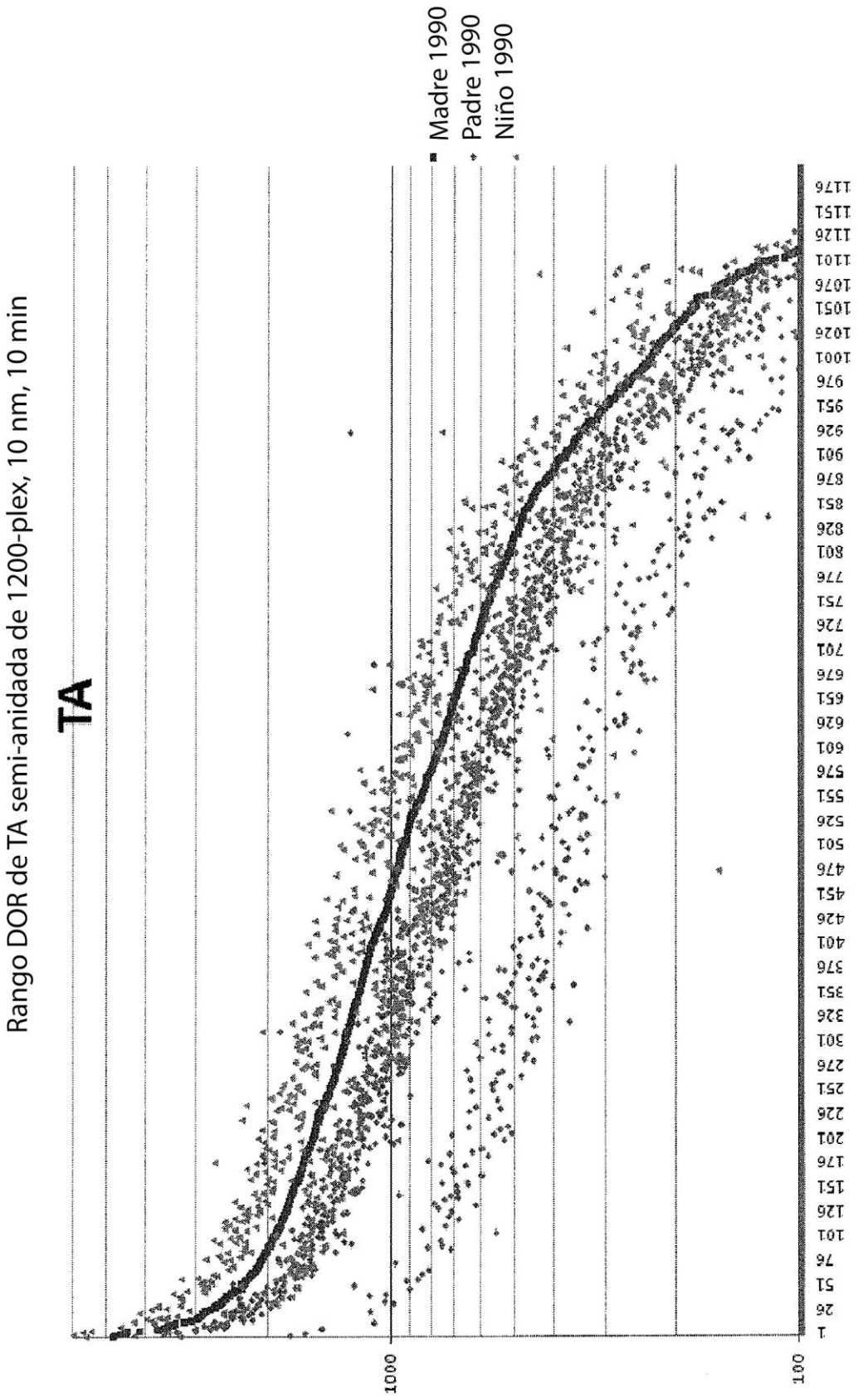

Rango DOR de TA semi-anidada de 1200-plex, 10 nm, 10 min
TA
Madre 1990
Padre 1990
Niño 1990
1000
100
1
26
51
76
101
126
151
176
201
226
251
276
301
326
351
376
401
426
451
476
501
526
551
576
601
626
651
676
701
726
751
776
801
826
851
876
901
926
951
976
1001
1026
1051
1076
1101
1126
1151
1176

Fig 22

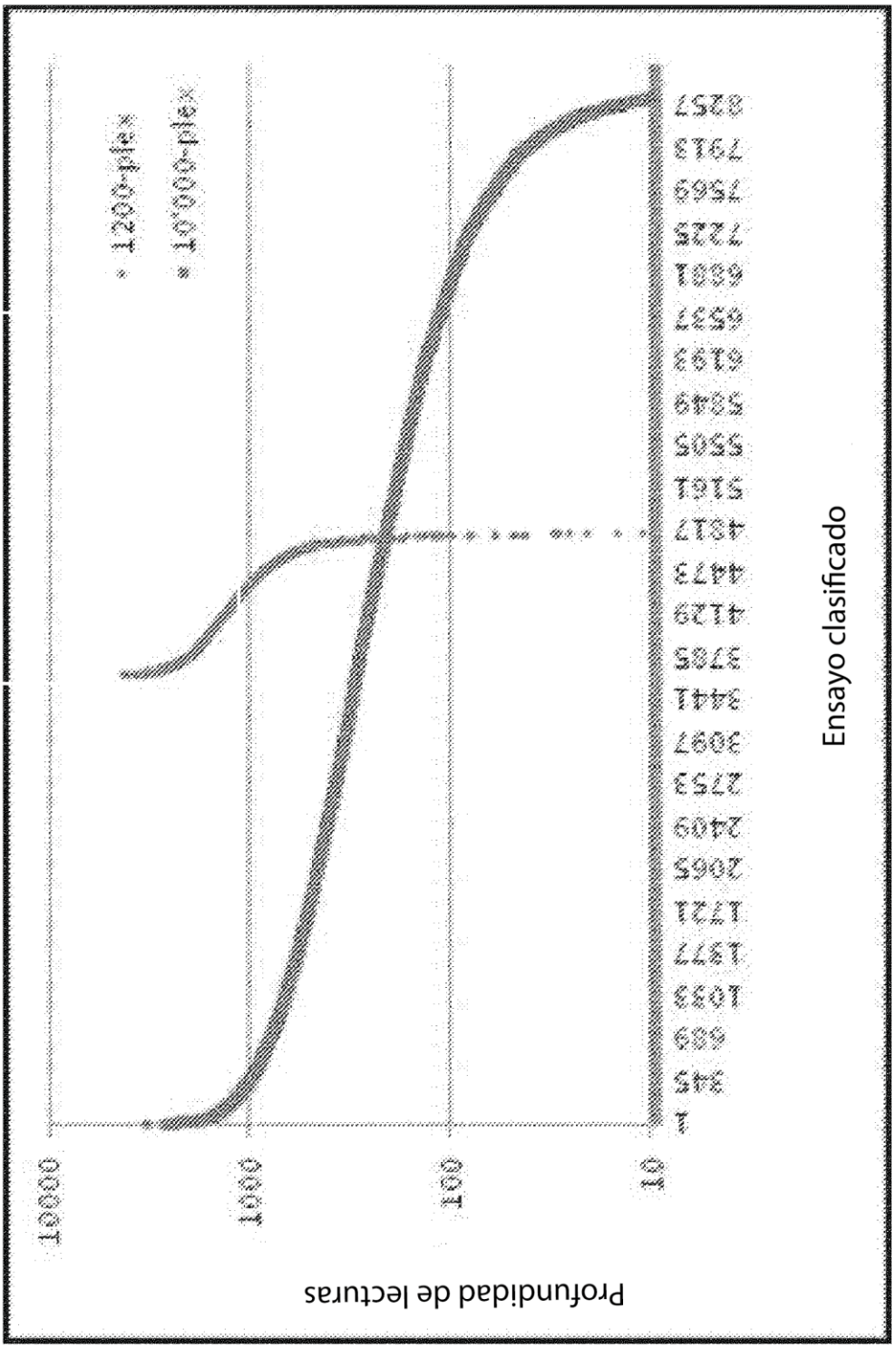

Profundidad de lecturas
10000
1000
100
10
Ensayo clasificado
1
345
689
1033
1377
1721
2065
2409
2753
3097
3441
3785
4129
4473
4817
5161
5505
5849
6193
6537
6881
7225
7569
7913
8257
1200-plex
10'000-plex

Fig. 23

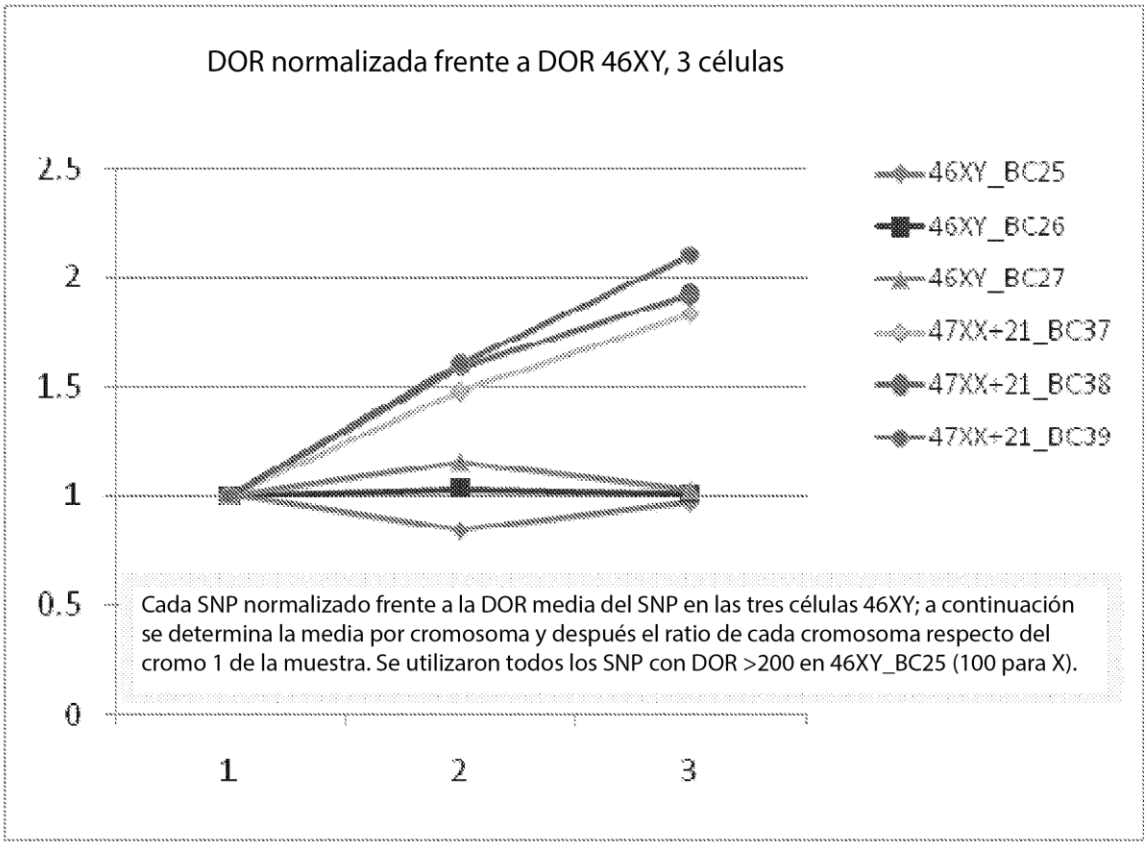
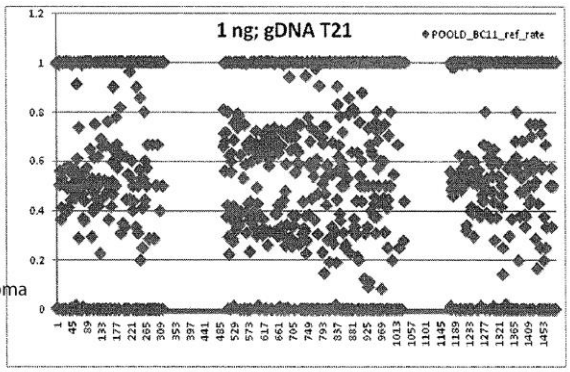
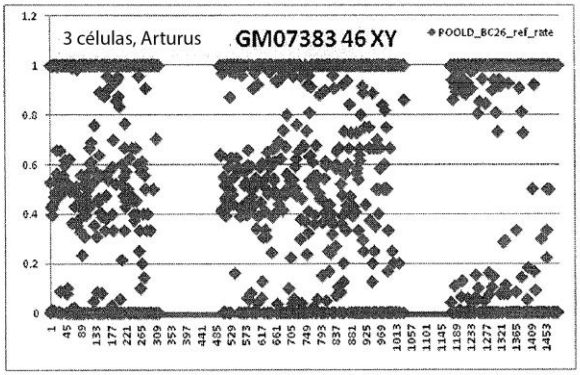



Fig 24

- Datos de 3 células con Arcturus: la mini-PCR puede detectar trisomía 21 de 3 células
- Izquierda: cromosoma 1; medio: cromosoma 21; y derecha: cromosoma X
- Clasificados de izquierda a derecha por la DOR para cada cromosoma



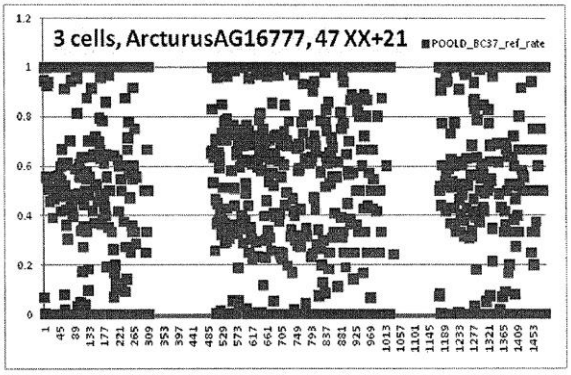
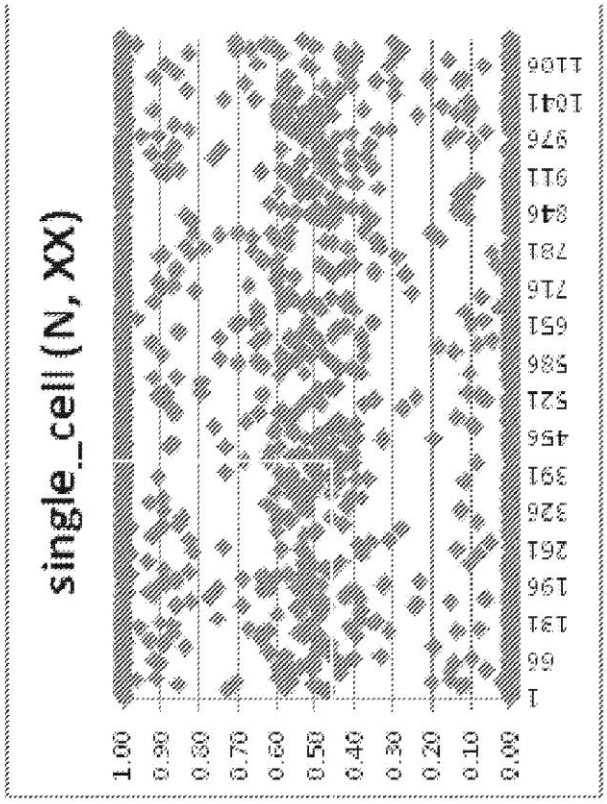




Fig 25


single_cell (N, XX)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
66
131
196
261
326
391
456
521
586
651
716
781
846
911
976
1041
1106

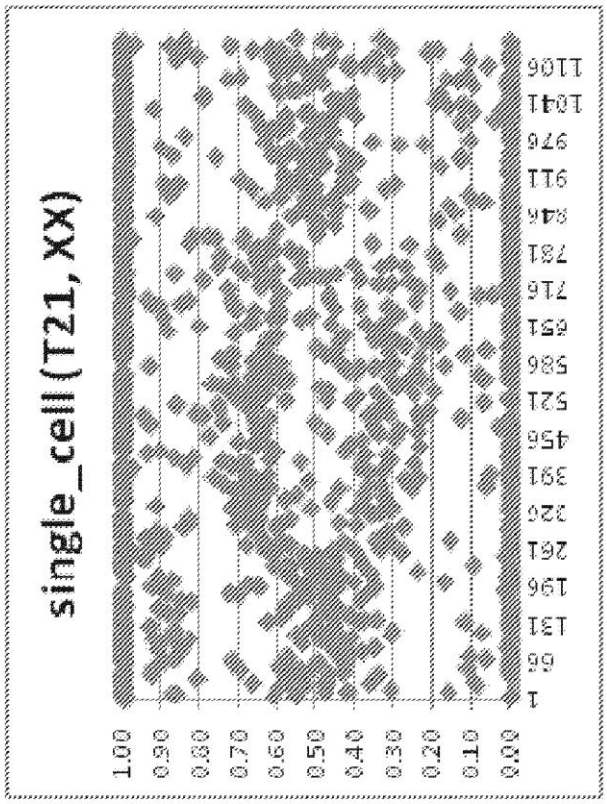

single_cell (T21, XX)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
66
131
196
261
326
391
456
521
586
651
716
781
846
911
976
1041
1106

Fig 26