

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 770 342**

51 Int. Cl.:

**G16B 10/00** (2009.01)

**C12Q 1/68** (2008.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.12.2011 PCT/US2011/066938**

87 Fecha y número de publicación internacional: **28.06.2012 WO12088456**

96 Fecha de presentación y número de la solicitud europea: **22.12.2011 E 11851621 (0)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019 EP 2656263**

54 Título: **Procedimientos para pruebas prenatales no invasivas de paternidad**

30 Prioridad:

**22.12.2010 US 201061426208 P**  
**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**  
**18.11.2011 US 201113300235**  
**22.12.2011 US 201113335043**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**01.07.2020**

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 770 342 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Procedimientos para pruebas prenatales no invasivas de paternidad

- 5 **[0001]** La presente descripción se refiere generalmente a métodos para pruebas prenatales no invasivas de paternidad.

## ANTECEDENTES

- 10 **[0002]** El parentesco poco claro es un problema significativo, y las estimaciones oscilan entre 4% y 10% de los niños que creen que su padre biológico es un hombre que no es su padre biológico real. En los casos en que una mujer está embarazada, pero las personas relevantes no están seguras de quién es el padre biológico, existen varias opciones para determinar el padre biológico correcto del feto. Un método es esperar hasta el nacimiento y realizar huellas genéticas en el niño y comparar la huella genética del genoma del niño con la de los padres sospechosos. Sin embargo, la madre a menudo desea conocer la identidad del padre biológico de su feto prenatalmente. Otro método es realizar un muestreo de vellosidades coriónicas en el primer trimestre o una amniocentesis en el segundo trimestre, y utilizar el material genético recuperado para realizar la toma de huellas genéticas prenatalmente. Sin embargo, estos métodos son invasivos y conllevan un riesgo significativo de aborto espontáneo.

- 20 **[0003]** Recientemente se ha descubierto que el ADN fetal libre de células (cfADN) y las células fetales intactas pueden entrar en la circulación de sangre materna. En consecuencia, el análisis de este material genético fetal puede permitir el diagnóstico genético prenatal no invasivo temprano (NIPGD o NPD). Un desafío clave en la realización de NIPGD en células fetales es la tarea de identificar y extraer células fetales o ácidos nucleicos de la sangre de la madre. La concentración de células fetales en la sangre materna depende de la etapa del embarazo y la condición del feto, pero las estimaciones varían de una a cuarenta células fetales en cada mililitro de sangre materna, o menos de una célula fetal por cada 100.000 células nucleadas maternas. Las técnicas actuales pueden aislar pequeñas cantidades de células fetales de la sangre de la madre, aunque es difícil enriquecer las células fetales hasta la pureza en cualquier cantidad. La técnica más efectiva en este contexto implica el uso de anticuerpos monoclonales, pero otras técnicas utilizadas para aislar células fetales incluyen la centrifugación de densidad, la lisis selectiva de eritrocitos adultos y FACS. Un desafío clave es realizar NIPGD en el ADNc fetal es que generalmente se mezcla con el ADNc materno y, por lo tanto, el análisis del ADNc se ve obstaculizado por la necesidad de tener en cuenta la señal genotípica materna. El análisis de ADN fetal se ha demostrado usando amplificación por PCR usando cebadores que están diseñados para hibridarse con secuencias que son específicas de los genes heredados por vía paterna. Estas fuentes de material genético fetal abren la puerta a técnicas de diagnóstico prenatal no invasivas.

- 35 **[0004]** Una vez se ha aislado el ADN fetal, ya sea puro o en una mezcla, que se puede amplificar. Existen varios métodos disponibles para la amplificación del genoma completo (WGA): PCR mediada por ligadura (LM-PCR), PCR de cebador oligonucleotídico degenerado (DOP-PCR) y amplificación de desplazamiento múltiple (MDA). Existen varios métodos disponibles para la amplificación dirigida, incluida la PCR y las sondas de circularización, como las SONDA DE INVERSIÓN MOLECULAR (MIP) y las sondas PADLOCK. Existen otros métodos que pueden usarse para enriquecer preferentemente el ADN fetal, como la separación por tamaño y las sondas de captura híbrida.

- 45 **[0005]** Existen numerosas dificultades en el uso de amplificación de ADN en estos contextos. La amplificación del ADN unicelular, el ADN de un pequeño número de células o de pequeñas cantidades de ADN, por PCR, puede fallar por completo. Esto a menudo se debe a la contaminación del ADN, la pérdida de la célula, su ADN o la accesibilidad del ADN durante la reacción de amplificación. Otras fuentes de error que pueden surgir al medir el ADN fetal mediante amplificación y análisis de microarrays incluyen errores de transcripción introducidos por la ADN polimerasa donde un nucleótido particular se copia incorrectamente durante la PCR y errores de lectura de microarrays debido a la hibridación imperfecta en la matriz. Otro problema es el abandono de alelos (ADO) definido como la incapacidad de amplificar uno de los dos alelos en una célula heterocigótica.

- 55 **[0006]** Existen muchas técnicas que proporcionan datos de genotipos. Algunos ejemplos incluyen lo siguiente. TAQMAN es una tecnología de genotipado única producida y distribuida por LIFE TECHNOLOGY. TAQMAN utiliza la reacción en cadena de la polimerasa (PCR) para amplificar secuencias de interés. Las ENSAYOS 500K de AFFYMETRIX y el sistema INFINIUM de ILLUMINA son matrices de genotipado que detectan la presencia de secuencias específicas de ADN en una gran cantidad de ubicaciones simultáneamente. HISEQ y MISEQ de ILLUMINA, y la plataforma ION TORRENT y SOLID de LIFE TECHNOLOGY permiten la secuenciación directa de una gran cantidad de secuencias de ADN individuales.

## SUMARIO

- [0007]** La presente invención proporciona el método *ex vivo* para establecer si un presunto padre es el padre biológico de un feto que se está gestando en una madre embarazada de acuerdo con las reivindicaciones 1-9.

- 65 **[0008]** En el presente documento se describen métodos para determinar la paternidad de un feto gestante de una manera no invasiva. Un método para establecer si un supuesto padre es el padre biológico de un feto que se está

gestando en una madre embarazada incluye obtener material genético del supuesto padre, obtener una muestra de sangre de la madre embarazada, realizar mediciones genotípicas, en una pluralidad de loci polimórficos, sobre el material genético del supuesto padre, obteniendo mediciones genotípicas, en la pluralidad de loci polimórficos, del material genético de la madre embarazada, haciendo mediciones genotípicas en una muestra mixta de ADN procedente de la muestra de sangre de la madre embarazada, donde la muestra mixta de ADN comprende ADN fetal y ADN materno, determinando, en una computadora, la probabilidad de que el presunto padre es el padre biológico del feto que se gesta en la madre embarazada usando las mediciones genotípicas hechas a partir del ADN del presunto padre, las mediciones genotípicas obtenidas de la madre embarazada y las mediciones genotípicas realizadas en la muestra mixta de ADN, y estableciendo si el presunto padre es el padre biológico del feto usando la probabilidad determinada de que el presunto padre sea el padre biológico del feto.

**[0009]** Los loci polimórficos comprenden polimorfismos de nucleótido único. La muestra mixta de ADN puede comprender ADN que provenía de ADN flotante libre en una fracción de plasma de la muestra de sangre de la madre embarazada. La muestra mixta de ADN puede comprender sangre completa materna o una fracción de sangre nuclear que contiene células nucleadas. La fracción de células nucleadas que contienen sangre materna puede enriquecerse para células de origen fetal.

**[0010]** La determinación de si el supuesto padre es el padre biológico puede incluir el cálculo de una estadística de prueba para el supuesto padre y el feto, en donde la estadística de prueba indica un grado de similitud genética entre el supuesto padre y el feto, y en donde la estadística de prueba se basa en las mediciones genotípicas realizadas a partir del ADN del presunto padre, las mediciones genotípicas realizadas a partir de la muestra mixta de ADN y las mediciones genotípicas obtenidas a partir del ADN de la madre embarazada, calculando la distribución de una estadística de prueba para una pluralidad de individuos que no están genéticamente relacionados con el feto, donde cada estadística de prueba calculada indica un grado de similitud genética entre un individuo no relacionado de la pluralidad de individuos que no están relacionados con el feto y el feto, en donde la estadística de prueba se basa en mediciones genotípicas hechas de ADN del individuo no relacionado, las mediciones genotípicas realizadas a partir de la muestra mixta de ADN, y mediciones genotípicas obtenidas del ADN de la madre embarazada, calculando una probabilidad de que la estadística de prueba calculada para el presunto padre y el feto sea parte de la distribución de la estadística de prueba calculada para la pluralidad de individuos no relacionados y el feto, y determinando la probabilidad de que el presunto padre es el padre biológico del feto usando la probabilidad de que la estadística de prueba calculada para el presunto padre sea parte de la distribución de la estadística de prueba calculada para la pluralidad de individuos no relacionados y el feto. En una realización, establecer si un presunto padre es el padre biológico del feto también implica establecer que el presunto padre es el padre biológico del feto al rechazar una hipótesis de que el presunto padre no está relacionado con el feto si la probabilidad de que el presunto padre es el padre biológico del feto está por encima de un umbral superior, o establece que el presunto padre no es el padre biológico del feto al no rechazar una hipótesis de que el presunto padre no está relacionado con el feto si la probabilidad de que el presunto padre sea el padre biológico del feto está por debajo de un umbral inferior, o no establece si un presunto padre es el padre biológico del feto si la probabilidad se encuentra entre el umbral inferior y el umbral superior, o si la probabilidad no se determinó con suficiente confianza.

**[0011]** La determinación de la probabilidad de que el supuesto padre sea el padre biológico del feto puede incluir la obtención de frecuencias poblacionales de alelos para cada locus en la pluralidad de loci polimórficos, creando una partición de posibles fracciones de ADN fetal en la muestra mixta de ADN que van desde un límite inferior de la fracción fetal hasta un límite superior de la fracción fetal, calculando una probabilidad de que el presunto padre sea el padre biológico del feto dadas las mediciones genotípicas obtenidas del ADN de la madre, las mediciones genotípicas hechas del ADN del supuesto padre, las mediciones genotípicas realizadas a partir de la muestra mixta de ADN, para cada una de las posibles fracciones fetales en la partición, determinando la probabilidad de que el supuesto padre sea el padre biológico del feto combinando las probabilidades calculadas de que el supuesto padre es el biológico padre del feto para cada una de las posibles fracciones fetales en la partición, calculando una probabilidad de que el supuesto padre no es el padre biológico del feto dadas las mediciones genotípicas hechas a partir de ADN de la madre, las mediciones genotípicas hechas a partir de la muestra mixta de ADN, las frecuencias de población de alelos obtenidas; para cada una de las posibles fracciones fetales en la partición, y determinar la probabilidad de que el presunto padre no sea el padre biológico del feto combinando las probabilidades calculadas de que el presunto padre no es el padre biológico del feto para cada uno de las posibles fracciones de fetos en la partición.

**[0012]** El cálculo de la probabilidad de que el supuesto padre es el padre biológico del feto y el cálculo de la probabilidad de que el supuesto padre no es el padre biológico del feto también puede incluir calcular, para cada uno de la pluralidad de loci polimórficos, un riesgo de datos de secuencia observados en un locus particular usando un modelo de respuesta de plataforma, una o una pluralidad de fracciones en la posible partición de fracciones fetales, una pluralidad de razones de alelos para la madre, una pluralidad de relaciones de alelos para el supuesto padre y una pluralidad de alelos relaciones para el feto, calculando una probabilidad de que el supuesto padre sea el padre biológico combinando la probabilidad de los datos de secuencia observados en cada locus polimórfico sobre todas las fracciones fetales en la partición, sobre las proporciones de alelos maternos en el conjunto de loci polimórficos, sobre las supuestas proporciones de alelos del padre en el conjunto de loci polimórficos, y sobre las relaciones de alelos fetales en el conjunto de loci polimórficos, calculando la probabilidad de que el supuesto padre no sea el padre biológico al combinar la probabilidad de los datos de secuencia observados en cada locus polimórfico sobre todas las

fracciones fetales en la partición, sobre las proporciones de alelos maternos en el conjunto de loci polimórficos, sobre las frecuencias de población para el conjunto de loci polimórficos, y sobre las proporciones de alelos fetales en el conjunto de loci polimórficos, calculando una probabilidad de que el supuesto padre sea el padre biológico en función de la probabilidad de que el supuesto padre sea el padre biológico, y calcule una probabilidad de que el supuesto padre no sea el padre biológico basado en la probabilidad de que el presunto padre no sea el padre biológico.

[0013] El cálculo de la probabilidad de que el presunto padre sea el padre biológico en función de la probabilidad de que el presunto sea el padre biológico se realiza utilizando una estimación de máxima probabilidad, o una técnica máxima *a posteriori*. Establecer si un supuesto padre es el padre biológico de un feto también puede incluir establecer que el supuesto padre es el padre biológico si la probabilidad calculada de que el supuesto padre es el padre biológico del feto es significativamente mayor que la probabilidad calculada de que el supuesto padre no es el padre biológico, o establecer que el presunto padre no es el padre biológico del feto si la probabilidad calculada de que el presunto padre es el padre biológico es significativamente mayor que la probabilidad calculada de que el presunto padre no es el padre biológico. En una realización, los loci polimórficos corresponden a cromosomas que tienen una alta probabilidad de ser disómicos.

[0014] La partición de posibles fracciones de ADN fetal puede contener sólo una fracción fetal, y donde la fracción fetal se determina por una técnica tomada de la lista que consiste en PCR cuantitativa, PCR digital, PCR dirigida, sondas de circularización, otros métodos de amplificación de ADN, captura por sondas de hibridación, otros métodos de enriquecimiento preferencial, microarrays de SNP, microarrays de ADN, secuenciación, otras técnicas para medir alelos polimórficos, otras técnicas para medir alelos no polimórficos, medir alelos polimórficos que están presentes en el genoma del padre pero no presente en el genoma de la madre, midiendo alelos no polimórficos que están presentes en el genoma del padre pero no presentes en el genoma de la madre, midiendo alelos que son específicos del cromosoma Y, comparando la cantidad medida de alelos heredados paternalmente a la cantidad medida de alelos heredados por vía materna, estimaciones de máxima verosimilitud, técnicas máximas *a posteriori*, y combinaciones de los mismos. En el caso de que la partición de posibles fracturas fetales contenga solo una fracción fetal, la fracción fetal puede determinarse utilizando un método que comprende realizar mediciones genotípicas en una pluralidad de loci polimórficos de la muestra mixta de ADN que comprende ADN de un individuo diana y ADN de un segundo individuo; obtener datos genotípicos en la pluralidad de loci polimórficos del segundo individuo; y determinar, en una computadora, la fracción de ADN del individuo diana presente en la muestra mixta usando las mediciones genotípicas de la muestra mixta de ADN, los datos genotípicos del segundo individuo, una técnica de estimación probabilística.

[0015] El material genético del supuesto padre puede ser obtenido a partir de tejido seleccionado del grupo que consiste en: sangre, tejido somático, espermatozoides, pelo, muestra bucal, la piel, otras muestras forenses, y combinaciones de los mismos. En una realización, se calcula una confianza para la determinación establecida de si el supuesto padre es el padre biológico del feto. La fracción de ADN fetal en la muestra mixta de ADN puede enriquecerse utilizando un método seleccionado del grupo que consiste en: selección de tamaño, PCR mediada por ligadura universal, PCR con tiempos de extensión cortos, otros métodos de enriquecimiento y combinaciones de los mismos.

[0016] La obtención de mediciones genotípicas del material genético de la madre embarazada puede incluir realizar mediciones genotípicas en una muestra de material genético de la madre embarazada que consiste esencialmente en material genético materno. La obtención de mediciones genotípicas del material genético de la madre embarazada puede incluir inferir qué mediciones genotípicas de las mediciones genotípicas realizadas en la muestra mixta de ADN son probablemente atribuibles al material genético de la madre embarazada, y usar esas mediciones genotípicas que se infirieron que son atribuibles al material genético de la madre como las medidas genotípicas obtenidas. El método también puede incluir tomar una decisión clínica basada en la determinación de paternidad establecida. La decisión clínica puede ser interrumpir un embarazo.

[0017] La toma de mediciones genotípicas se puede hacer por la medición de material genético utilizando una técnica o tecnología seleccionada del grupo que consiste en sondas de candado, la inversión de sondas moleculares, otras sondas circularización, la genotipificación de microarrays, ensayos de genotipificación de SNP, microarrays basados en chip, microarrays basados en perlas, otros microarrays SNP, otros métodos de genotipado, secuenciación de ADN de Sanger, secuenciación de pirosecuenciación, secuenciación de alto rendimiento, secuenciación dirigida usando sondas de circularización, secuenciación dirigida usando captura por sondas de hibridación, secuenciación de terminador de colorante reversible, secuenciación por ligadura, secuenciación por hibridación, otros métodos de secuenciación de ADN, otras plataformas de genotipado de alto rendimiento, hibridación fluorescente *in situ* (FISH), hibridación genómica comparativa (CGH), CGH de matriz y múltiplos o combinaciones de los mismos.

[0018] La toma de mediciones genotípicas puede realizarse en material genético que se amplifica y/o preferentemente se enriquece antes de medirse usando una técnica o tecnología que se selecciona del grupo que consiste en: Reacción de Cadena de Polimerasa (PCR), PCR mediado por ligando, cebador de oligonucleótido degenerativo PCR, amplificación dirigida, PCR, miniPCR, amplificación universal por PCR, amplificación de desplazamiento múltiple (MDA), PCR específica de alelo, técnicas de amplificación específicas de alelo, métodos de amplificación lineal, ligadura de ADN de sustrato seguido de otro método de amplificación, amplificación de puente, sondas de candado, sondas de circularización, sondas de captura por hibridación y combinaciones de las mismas.

[0019] El método puede incluir también la generación de un informe que comprende la paternidad establecido del feto. En una realización, el informe puede revelar la paternidad establecida del feto generado usando un método descrito aquí.

[0020] En el presente documento se describen métodos para determinar la fracción de ADN que se origina a partir de un individuo diana que está presente en una mezcla de ADN que contiene ADN del individuo diana, y también ADN de al menos otro individuo. De acuerdo con los aspectos ilustrados aquí, en una realización, un método para determinar una fracción de ADN de un individuo diana presente en una muestra mixta de ADN que comprende ADN del individuo diana y ADN de un segundo individuo puede incluir realizar mediciones genotípicas en una pluralidad de loci polimórficos de la muestra mixta de ADN, obteniendo datos genotípicos en la pluralidad de loci polimórficos del segundo individuo y determinando, en una computadora, la fracción del ADN del individuo diana presente en la muestra mixta utilizando las mediciones genotípicas de la muestra mixta de ADN, los datos genotípicos del segundo individuo y las técnicas de estimación probabilística.

[0021] La obtención de datos genotípicos del segundo individuo incluye realizar mediciones genéticas a partir de ADN que consiste esencialmente en ADN del segundo individuo. La obtención de datos genotípicos del segundo individuo puede incluir inferir qué mediciones genotípicas de las mediciones genotípicas realizadas en la muestra mixta de ADN probablemente sean atribuibles al material genético del segundo individuo, y usar esas mediciones genotípicas que se infiere que son atribuibles al material genético del segundo individuo como las medidas genotípicas obtenidas.

[0022] Inferir los datos genotípicos del individuo relacionado también puede incluir el uso de frecuencias de población de alelos en los loci. La fracción determinada de ADN de un individuo diana se expresa como una probabilidad de fracciones de ADN. En una realización, las mediciones genotípicas realizadas a partir de la muestra mixta comprenden mediciones genotípicas realizadas secuenciando el ADN en la muestra mixta. El ADN en la muestra mixta se enriquece preferentemente en la pluralidad de loci polimórficos antes de realizar mediciones genotípicas a partir de la muestra mixta de ADN. En una realización, los loci polimórficos comprenden polimorfismos de un solo nucleótido.

[0023] La determinación de la fracción también puede incluir la determinación de una probabilidad de una pluralidad de fracciones de ADN del individuo diana presente en la muestra mixta de ADN, determinando la fracción seleccionando la fracción de la pluralidad de fracciones con la mayor probabilidad. La determinación de la fracción también puede incluir la determinación de una probabilidad de una pluralidad de fracciones de ADN del individuo diana presente en la muestra mixta de ADN, utilizando una técnica de estimación de máxima verosimilitud para determinar la fracción más probable y determinar la fracción seleccionando la fracción que se determinó que era lo más probable.

[0024] El individuo diana puede ser un feto gestante en una madre embarazada, y el segundo individuo puede ser la madre embarazada. El método también puede incluir el uso de un modelo de plataforma que relaciona datos genotípicos medidos en los loci polimórficos, y el uso de una tabla que relaciona los genotipos maternos con los genotipos infantiles. La determinación también utiliza mediciones genotípicas en una pluralidad de loci polimórficos medidos en el ADN del padre del feto. Es posible que el método no utilice datos genotípicos del padre del feto. Es posible que el método no utilice loci en el cromosoma Y. El informe puede revelar una paternidad establecida del feto determinada usando un método divulgado aquí para determinar la fracción de ADN fetal presente en el plasma materno. El informe puede revelar un estado de ploidía del feto determinado usando un método divulgado aquí para determinar la fracción de ADN fetal presente en el plasma materno.

## BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0025] Las formas de realización descritas en la presente se explicará adicionalmente con referencia a los dibujos adjuntos, en donde como estructuras se denominan por números similares en todas las diversas vistas. Los dibujos mostrados no son necesariamente a escala, sino que generalmente se hace hincapié en ilustrar los principios de las realizaciones descritas actualmente.

**La Figura 1** muestra la distribución de las intensidades alélicas de dos contextos parentales medidos en plasma materno.

**La Figura 2** muestra la distribución de la estadística de prueba relacionada con la paternidad para 200 hombres no relacionados y el padre biológico.

**La Figura 3** muestra dos distribuciones de relaciones de intensidad para 200 hombres no relacionados y el padre biológico. Cada gráfico corresponde a diferentes canales de entrada.

**La Figura 4** muestra las curvas de frecuencia de distribución acumulativa (cdf) para la relación de correlación entre las mediciones genotípicas fetales y las mediciones genotípicas parentales para tres casos.

**La Figura 5** muestra histogramas de la relación de correlación entre las mediciones genotípicas fetales y las mediciones genotípicas parentales para tres casos.

**La Figura 6** muestra un histograma de la estadística de prueba de paternidad para 35 muestras en comparación con una distribución gaussiana idealizada de estadísticas de prueba para 800 hombres no relacionados.

**La Figura 7** muestra un ejemplo de un informe que revela una exclusión de paternidad.

**La Figura 8** muestra un ejemplo de un informe que revela una inclusión de paternidad.

La Figura 9 muestra un ejemplo de un informe que revela un resultado indeterminado.

## DESCRIPCIÓN DETALLADA

**[0026]** Se describe un método para determinar si un supuesto padre es o no el padre biológico de un feto que se está gestando en una madre embarazada. El método puede incluir obtener material genético del supuesto padre y obtener una muestra de sangre de la madre embarazada. El método puede incluir realizar mediciones genotípicas del supuesto padre y la madre embarazada, y realizar mediciones genotípicas en el ADN flotante libre (ffADN, es decir, cfADN) que se encuentra en el plasma de la madre embarazada. El método puede incluir la obtención de datos genotípicos para un conjunto de SNP de la madre y el supuesto padre del feto; realizar mediciones genotípicas para el conjunto de SNP en una muestra mixta que comprende ADN del individuo diana y también ADN de la madre del individuo diana. El método puede incluir el uso de las mediciones genotípicas para determinar, en una computadora, la probabilidad de que el supuesto padre sea el padre biológico del feto gestando en la madre embarazada. El método puede incluir el uso de los datos genotípicos de la madre embarazada y el supuesto padre para determinar una distribución alélica esperada para las mediciones genotípicas de la mezcla de ADN fetal/materno si el supuesto padre fuera el padre biológico del feto. El método puede incluir el uso de los datos genotípicos de la madre embarazada y los datos genotípicos de una pluralidad de individuos que se sabe que no son el padre para determinar una distribución alélica esperada para las mediciones genotípicas de la mezcla de ADN fetal/materno si el supuesto padre no es el padre biológico del feto. El método puede involucrar el cálculo de las probabilidades de que el supuesto padre sea el padre biológico del feto dadas las distribuciones alélicas esperadas y las mediciones reales de ADN del plasma materno. La serie de pasos descritos en el método puede dar como resultado una transformación del material genético de la madre embarazada y el supuesto padre para producir y determinar la identidad correcta del padre biológico de un feto gestante prenatalmente y de manera no invasiva. La determinación de la probabilidad de que el presunto padre sea el padre biológico puede incluir llamar o establecer al presunto padre como el padre biológico si la probabilidad de que el padre sea excluido de la distribución alélica creada usando la pluralidad de individuos no relacionados está por encima de un umbral. Determinar la probabilidad de que el presunto padre sea el padre biológico puede incluir llamar al presunto padre como no el padre biológico si la probabilidad de que el presunto padre sea excluido de la distribución alélica creada usando la pluralidad de individuos no relacionados está por debajo de un umbral. La determinación de paternidad se puede hacer asumiendo inicialmente que el presunto padre es de hecho el padre del niño; si el supuesto padre es incorrecto, los genotipos del niño no se ajustarán a estas predicciones, y la suposición inicial se considera incorrecta; sin embargo, si los genotipos secundarios se ajustan a las predicciones, entonces se considera que la suposición es correcta. Por lo tanto, la prueba de paternidad considera qué tan bien el ffADN observado se ajusta a los genotipos infantiles predichos por los genotipos del supuesto padre. En una realización, se puede generar un informe electrónico o físico que indique la determinación de paternidad.

**[0027]** La determinación de paternidad puede hacerse usando mediciones genéticas de ADN de libre flotación (ffADN) encontrado en la sangre materna, y la información del genotipo de la madre y el supuesto padre. El método general podría aplicarse a las mediciones de ADNff utilizando una variedad de plataformas, como microarrays SNP, secuenciación de alto rendimiento sin diana o secuenciación dirigida. Los métodos discutidos aquí abordan el hecho de que el ADN fetal flotante se encuentra en el plasma materno a concentraciones bajas pero desconocidas y es difícil de detectar. La prueba de paternidad puede comprender la evaluación de las mediciones de ADNf y la probabilidad de que hayan sido generadas por el supuesto padre, en función de sus genotipos. Independientemente de la plataforma de medición, la prueba puede basarse en los genotipos medidos en ubicaciones polimórficas. En algunas realizaciones los posibles alelos en cada locus polimórfico pueden ser generalizados a A y B, y opcionalmente C, D y/o E, etc.

**[0028]** Este método puede implicar el uso de datos de medición de alelos desde una pluralidad de loci. Los loci pueden ser polimórficos. Algunos o la mayoría de los loci pueden ser polimórficos. Los loci polimórficos pueden ser polimorfismos de un solo nucleótido (SNP). Algunos o la mayoría de los loci polimórficos pueden ser heterocigotos. Puede que no sea necesario determinar qué loci son heterocigotos antes de la prueba.

**[0029]** Un método dado a conocer en el presente documento pueden utilizar técnicas de enriquecimiento selectivo que preservan las frecuencias alélicas relativas que están presentes en la muestra original de ADN en cada locus polimórfico de un conjunto de loci polimórficos. La técnica de amplificación y/o enriquecimiento selectivo puede implicar técnicas de PCR como mini-PCR o PCR mediada por ligadura, captura de fragmentos por hibridación o sondas de circularización como sondas de inversión molecular. Los métodos para la amplificación o el enriquecimiento selectivo pueden implicar el uso de cebadores de PCR u otras sondas donde, tras la hibridación correcta con la secuencia diana, el extremo 3-cebador o 5-cebador de una sonda de nucleótidos se separa del sitio polimórfico del alelo por un pequeño número de nucleótidos. Las sondas en las que la región de hibridación está diseñada para hibridarse con un sitio polimórfico pueden excluirse. Estas realizaciones son mejoras sobre otros métodos que implican amplificación dirigida y/o enriquecimiento selectivo en el sentido de que preservan mejor las frecuencias alélicas originales de la muestra en cada locus polimórfico, ya sea que la muestra sea una muestra genómica pura de un solo individuo o una mezcla de individuos.

**[0030]** Un método descrito en este documento PCR dirigidas eficientes altamente multiplexados para amplificar el ADN seguido de secuenciación de alto rendimiento para determinar las frecuencias de alelos en cada locus diana.

Una técnica que permite que la PCR dirigida altamente multiplexada funcione de manera altamente eficiente implica el diseño de cebadores que es poco probable que se hibridicen entre sí. Las sondas de PCR pueden seleccionarse creando un modelo termodinámico de interacciones potencialmente adversas o interacciones no intencionadas, entre al menos 500, al menos 1.000, al menos 5.000, al menos 10.000, al menos 20.000, al menos 50.000 o al menos 100.000 potenciales pares de cebadores, o entre cebadores y muestra de ADN, y luego usar el modelo para eliminar diseños que son incompatibles con otros diseños en el conjunto, o con la muestra de ADN. Otra técnica que permite que la PCR dirigida altamente multiplexada se realice de una manera altamente eficiente es usar un enfoque de anidamiento parcial o completo para la PCR dirigida. El uso de uno o una combinación de estos enfoques permite la multiplexación de al menos 300, al menos 800, al menos 1.200, al menos 4.000 o al menos 10.000 cebadores en un solo grupo con el ADN amplificado resultante que comprende una mayoría de moléculas de ADN que, cuando se secuencian, se asignarán a loci específicos. El uso de uno o una combinación de estos enfoques permite la multiplexación de una gran cantidad de cebadores en un único conjunto con el ADN amplificado resultante que comprende más del 50%, más del 80%, más del 90%, más del 95%, más del 98%, o más del 99% de moléculas de ADN que se asignan a loci específicos.

**[0031]** Un método descrito en este documento puede implicar la determinación de si la distribución de las mediciones de los alelos observados es indicativa de una inclusión de paternidad o exclusión usando una estimación de máxima verosimilitud (MLE) técnica. El uso de una técnica de estimación de máxima verosimilitud es diferente y una mejora significativa sobre los métodos que utilizan la técnica de rechazo de hipótesis única en que las determinaciones resultantes se realizarán con una precisión significativamente mayor. Una razón es que las técnicas de rechazo de hipótesis única no contienen información sobre la hipótesis alternativa. Otra razón es que la técnica de máxima verosimilitud permite la determinación de umbrales de corte óptimos para cada muestra individual. Otra razón es que el uso de una técnica de máxima verosimilitud permite el cálculo de una confianza para cada determinación de paternidad. La capacidad de hacer un cálculo de confianza para cada determinación permite al profesional saber qué llamadas son precisas y cuáles tienen más probabilidades de estar equivocadas. Se puede combinar una amplia variedad de métodos con una técnica de estimación de máxima verosimilitud para mejorar la precisión de las llamadas de ploidía. Un método divulgado aquí puede implicar estimar la fracción fetal de ADN en la muestra mixta y usar esa estimación para calcular tanto la llamada de paternidad (determinación) como la confianza de la llamada de paternidad.

**[0032]** Un método descrito puede implicar el cálculo de una estadística de prueba que es indicativo del grado de relación entre un primer individuo y un segundo individuo, medidas genotípicas dadas en una pluralidad de loci polimórficos para el primer individuo, y medidas genotípicas en una pluralidad de loci polimórficos para una mezcla de ADN donde la mezcla de ADN comprende ADN del segundo individuo y un afín individual. El primer individuo puede ser un supuesto padre, el segundo individuo puede ser un feto gestante y el individuo relacionado puede ser la madre del feto. La estadística de prueba puede calcularse para el feto, la madre y una pluralidad de individuos que se sabe que no están relacionados con el feto, generando así una distribución de la métrica para individuos no relacionados. La estadística de prueba también se puede calcular para el feto, la madre y el supuesto padre. Se puede usar una sola prueba de rechazo de hipótesis para determinar si la estadística de prueba calculada usando los datos genotípicos del supuesto padre es parte de la distribución de estadísticas de prueba calculadas usando los datos genotípicos de los individuos no relacionados. Si se encuentra que la estadística de prueba calculada usando los datos genotípicos del presunto padre es parte de la distribución de las estadísticas de prueba calculadas usando los datos genotípicos de los individuos no relacionados, entonces se puede excluir la paternidad, es decir, se puede determinar que el presunto padre no es relacionado con el feto. Si se encuentra que la estadística de prueba calculada usando los datos genotípicos del supuesto padre no forma parte de la distribución de las estadísticas de prueba calculadas usando los datos genotípicos de los individuos no relacionados, entonces se puede incluir la paternidad, es decir, se puede determinar que el supuesto padre es relacionado con el feto.

**[0033]** La determinación de la paternidad implica la determinación de la probabilidad de los datos genotípicos medidos dados dos hipótesis posibles: la hipótesis de que el presunto padre es el padre biológico del feto, y la hipótesis de que el presunto padre no es el padre biológico del feto. Luego se puede calcular una probabilidad para cada una de las hipótesis dados los datos, y la paternidad se puede establecer en función de la probabilidad de cada una de las dos hipótesis. La determinación puede utilizar mediciones genéticas realizadas en el plasma materno, mediciones genéticas realizadas en el ADN del presunto padre y, opcionalmente, datos genotípicos maternos. En una realización, los datos genotípicos maternos pueden inferirse de las mediciones genotípicas realizadas en el plasma materno. La probabilidad puede determinarse usando una partición del rango de posibles fracciones fetales; El rango de fracciones fetales podría ser de 0,01% a 99,9%, y la malla puede tener incrementos que varían de 10% a 1%, de 1% a 0,1% y menores de 0,1%. La partición de posibles fracciones fetales puede ser del 2% al 30%, y los incrementos son de aproximadamente el 1%. En una realización, la malla podría ser continua, y las probabilidades podrían integrarse en los intervalos en lugar de combinarse. La probabilidad puede determinarse usando solo una fracción fetal, donde esa fracción fetal puede determinarse usando cualquier método apropiado. Para cada posible fracción fetal en la malla, se puede calcular la probabilidad de los datos dadas las dos hipótesis. Para la hipótesis de que el supuesto padre es el padre biológico, los genotipos del supuesto padre pueden usarse en el cálculo de la probabilidad, mientras que para la hipótesis de que el supuesto padre no es el padre biológico, los datos de frecuencia de alelos basados en la población también pueden usarse en el cálculo de la probabilidad. Se pueden usar los contextos principales y un modelo de plataforma para calcular la probabilidad de que los datos den una hipótesis. Las probabilidades se pueden combinar sobre todas las fracciones fetales en la partición, sobre todos los genotipos maternos y sobre todos los

genotipos paternos. Los genotipos parentales pueden ser probabilísticos (por ejemplo, en un SNP dado, un padre puede tener el genotipo GT con un 99% de probabilidad, GG con un 0,5% de probabilidad y TT con un 0,5% de probabilidad; en otra realización, los genotipos parentales pueden tener un valor (por ejemplo, en un SNP dado, un padre tiene el genotipo GT). Los términos probabilidad y verosimilitud pueden ser intercambiables, como en el lenguaje común; en otras realizaciones, los dos términos pueden no ser intercambiables y pueden leerse como los leerían expertos en la técnica en las estadísticas.

**[0034]** En algunos métodos conocidos en la técnica, la fracción fetal se determina usando las mediciones realizadas en loci que se encuentran exclusivamente en el genotipo paterno, por ejemplo, loci que se encuentran exclusivamente en el cromosoma Y, o el gen Rhesus D. Desafortunadamente, estos métodos requieren que el feto sea masculino (en el caso de que los loci se encuentren exclusivamente en el cromosoma Y) o que se pueda identificar un gen o conjunto de genes antes de las mediciones del ADN donde esos genes están presentes en el genotipo paterno, y no presente en el genotipo materno. Una complicación adicional es que, en el contexto de las pruebas de paternidad, no se sabe si el presunto padre es el padre biológico y, por lo tanto, con la excepción de los loci específicos del cromosoma Y, no es posible determinar qué loci puede estar presente en el padre y no en la madre. Por lo tanto, en el contexto de las pruebas de paternidad, actualmente no es posible determinar la fracción fetal cuando el feto es una mujer y cuando el feto es un hombre. La fracción fetal solo puede determinarse mediante el uso de loci específicos del cromosoma Y. En una realización, se describe un método en el presente documento para determinar la fracción de ADN fetal que está presente en la mezcla de ADN que comprende ADN materno y fetal. En una realización, el método puede determinar la fracción fetal de ADN fetal que está presente en la mezcla de ADN que comprende ADN materno y fetal utilizando mediciones genotípicas de cromosomas autosómicos. En una realización, el método puede determinar la fracción fetal de ADN fetal que está presente en la mezcla de ADN que comprende ADN materno y fetal independientemente del sexo del feto. En una realización, el método puede determinar la fracción fetal de ADN fetal que está presente en la mezcla de ADN que comprende ADN materno y fetal, independientemente de qué genes puedan tener la madre y el supuesto padre. El método instantáneo no requiere que el feto sea masculino, o que se pueda identificar un locus o loci que estén presentes en el padre y no en la madre. El método instantáneo no requiere que se conozca el genotipo paterno. El método instantáneo no requiere que se conozca el genotipo materno, ya que se puede inferir de las mediciones realizadas en el ADN en el plasma materno, que comprende una mezcla de ADN fetal y materno.

**[0035]** La distribución de los loci polimórficos se puede modelar utilizando una distribución binomial. La distribución de loci polimórficos se puede modelar utilizando una distribución beta-binomial. Al usar la distribución beta-binomial como modelo para la distribución de alelos, se puede modelar con mayor precisión las medidas probables de alelos que cuando se usan otras distribuciones; esto puede resultar en determinaciones de paternidad más precisas.

**[0036]** Un método descrito en este documento puede tener en cuenta la tendencia de que los datos sean ruidosos y contienen errores por fijación de una probabilidad a cada medición. El uso de técnicas de máxima verosimilitud para elegir la hipótesis correcta del conjunto de hipótesis que se hicieron utilizando los datos de medición con estimaciones probabilísticas adjuntas hace que sea más probable que se descarten las mediciones incorrectas, y las mediciones correctas se utilizarán en los cálculos que conducen a la determinación de la paternidad. Para ser más precisos, este método reduce sistemáticamente la influencia de los datos medidos incorrectamente en la determinación de la paternidad. Esta es una mejora con respecto a los métodos en los que se supone que todos los datos son igualmente correctos o los métodos en los que los datos periféricos se excluyen arbitrariamente de los cálculos que conducen a una determinación de paternidad. En una realización, los SNP individuales se ponderan por la varianza de medición esperada en función de la calidad del SNP y la profundidad de lectura observada; Esto puede resultar en un aumento en la precisión de la estadística resultante, resultando en un aumento de la precisión de la llamada de paternidad significativamente, especialmente en casos límite.

**[0037]** Los métodos descritos en este documento son particularmente ventajosos cuando se utilizan en muestras en las que una pequeña cantidad de ADN está disponible, o cuando el porcentaje de ADN fetal es bajo. Esto se debe a la tasa de abandono de alelo correspondientemente más alta que puede ocurrir cuando solo hay una pequeña cantidad de ADN disponible y/o la tasa de abandono de alelo fetal correspondientemente más alta cuando el porcentaje de ADN fetal es bajo en una muestra mixta de ADN fetal y materno. Una alta tasa de abandono de alelos, lo que significa que un gran porcentaje de los alelos no se midió para el individuo diana, da como resultado cálculos de fracción fetal poco precisos y determinaciones de paternidad poco precisas. Los métodos descritos en este documento permiten una determinación precisa de la 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%, inferior al 6%, inferior al 4% e incluso inferior al 3%.

**[0038]** Es posible determinar la paternidad de un individuo basándose en mediciones cuando el ADN de ese individuo se mezcla con el ADN de un individuo relacionado. La mezcla de ADN puede ser el ADN de libre flotación encontrado en el plasma materno, que puede incluir ADN de la madre, con genotipo conocido, y que puede mezclarse con ADN del feto, con genotipo desconocido. La paternidad del feto se puede determinar observando las medidas reales y determinando la probabilidad de paternidad dados los datos observados. Un método divulgado en este documento podría usarse en situaciones en las que hay una cantidad muy pequeña de ADN presente, como en situaciones forenses, donde hay una o unas pocas células disponibles (típicamente menos de diez células, menos de veinte



células, menos de 40 células, menos de 100 células, o una cantidad equivalente de ADN.) Un método descrito en este documento podría usarse en situaciones en las que el ADN está altamente fragmentado, como el ADNff encontrado en el plasma. Un método divulgado en este documento puede servir para realizar llamadas de paternidad desde una pequeña cantidad de ADN que no está contaminada por otro ADN, pero donde la llamada de paternidad es muy difícil debido a la pequeña cantidad de ADN. Las mediciones genéticas utilizadas como parte de estos métodos podrían realizarse en cualquier muestra que comprenda ADN o ARN, por ejemplo, pero sin limitarse a: sangre, plasma, fluidos corporales, orina, cabello, lágrimas, saliva, tejido, piel, uñas, blastómeros, embriones, líquido amniótico, muestras de vellosidades coriónicas, heces, bilis, linfa, moco cervical, semen u otras células o materiales que comprenden ácidos nucleicos. En una realización, un método descrito en este documento podría ejecutarse con métodos de detección de ácido nucleico tales como secuenciación, microarrays, qPCR, PCR digital u otros métodos utilizados para medir ácidos nucleicos. En algunas realizaciones, un método divulgado en el presente documento implica calcular, en una computadora, relaciones de alelos en la pluralidad de loci polimórficos a partir de las mediciones de ADN realizadas en las muestras procesadas. Un método divulgado en el presente documento puede implicar calcular, en una computadora, relaciones alélicas o distribuciones alélicas en una pluralidad de loci polimórficos a partir de las mediciones de ADN realizadas en las muestras procesadas junto con cualquier combinación de otras mejoras descritas en esta divulgación.

[0039] Además discusión de estos puntos se puede encontrar en este documento en otros lugares.

#### Pruebas prenatales de paternidad no invasivas (NPPT)

[0040] El proceso de pruebas de paternidad prenatales no invasivas implica una serie de pasos. Algunos de los pasos pueden incluir: (1) obtener el material genético del feto; (2) enriquecer el material genético del feto que puede estar en una muestra mixta, *ex vivo*; (3) amplificar el material genético, *ex vivo*; (4) enriquecer preferentemente loci específicos en el material genético, *ex vivo*; (5) medir el material genético, *ex vivo*; y (6) analizar los datos genotípicos, en una computadora, y *ex vivo*. Los métodos para reducir la práctica de estos seis y otros pasos relevantes se describen aquí. Al menos algunos de los pasos del método no se aplican directamente en el cuerpo. En una realización, la presente descripción se refiere a métodos de tratamiento y diagnóstico aplicados a tejidos y otros materiales biológicos aislados y separados del cuerpo. Al menos algunos de los pasos del método se ejecutan en una computadora.

[0041] La presente divulgación puede permitir que un clínico determine el estado genético de un feto, específicamente su biológica relación a otro individuo, que está gestando en una madre de una manera no invasiva de tal manera que la salud del bebé no se pone en riesgo por la recolección del material genético del feto, y que la madre no está obligada a someterse a un procedimiento invasivo.

[0042] Avances tecnológicos modernos han dado como resultado la capacidad de medir grandes cantidades de información genética a partir de una muestra genética utilizando métodos tales como alta secuenciación de rendimiento y matrices de genotipado. Los métodos descritos en este documento permiten a un médico aprovechar más las grandes cantidades de datos disponibles y hacer un diagnóstico más preciso de la identidad genética fetal. Un método basado en la informática puede dar lugar a determinaciones de paternidad de mayor precisión que los métodos actualmente conocidos en la técnica. Los detalles de una serie de realizaciones se dan a continuación. Las diferentes realizaciones pueden implicar diferentes combinaciones de los pasos mencionados anteriormente. Varias combinaciones de las diferentes realizaciones de los diferentes pasos pueden usarse indistintamente.

[0043] Una muestra de sangre se puede tomar de una madre embarazada, y el ADN de libre flotación en el plasma de la madre de la sangre, que contiene una mezcla de ambos ADN de origen materno, y el ADN de origen fetal, se aísla y se utiliza para determinar el estado de ploidía del feto. Un método divulgado en el presente documento puede implicar el enriquecimiento preferencial de esas secuencias de ADN en una mezcla de ADN que corresponde a alelos polimórficos de manera que las proporciones de alelos y/o las distribuciones de alelos permanezcan razonablemente consistentes tras el enriquecimiento. El método puede implicar amplificar el ADN aislado usando la amplificación del genoma completo (WGA). Un método divulgado en el presente documento puede implicar una amplificación basada en PCR dirigida de modo que un alto porcentaje de las moléculas resultantes correspondan a loci dirigidos. Un método divulgado en el presente documento puede implicar la secuenciación de una mezcla de ADN que contiene tanto ADN de origen materno como ADN de origen fetal. El método puede implicar medir el ADN amplificado utilizando un microarray diseñado para detectar secuencias de ácido nucleico, como una matriz SNP. Un método divulgado en el presente documento puede implicar el uso de distribuciones de alelos medidas para determinar la paternidad de un feto que se está gestando en una madre. Un método divulgado en el presente documento puede implicar informar el estado de paternidad determinado a un médico. Un método divulgado en el presente documento puede implicar tomar una acción clínica, por ejemplo, realizar pruebas invasivas de seguimiento tales como muestreo de vellosidades coriónicas o amniocentesis, prepararse para el nacimiento de un niño o una terminación electiva de un feto.

[0044] Esta solicitud hace referencia a la Solicitud de servicios públicos de Estados Unidos N° de serie 11/603,406, presentada el 28 de noviembre de 2006 (Publicación de Estados Unidos N°: 20070184467); Solicitud de servicios públicos de los Estados Unidos, número de serie 12/076,348, presentada el 17 de marzo de 2008 (publicación de los Estados Unidos: 20080243398); Solicitud de utilidad PCT N° de serie PCT/US09/52730, presentada el 4 de agosto de 2009 (N° de publicación PCT: WO/2010/017214); Solicitud de utilidad PCT, número de serie PCT/US10/050824,

presentada el 30 de septiembre de 2010 (publicación de PCT N°: WO/2011/041485), solicitud de utilidad de EE.UU., número de serie 13/110,685, presentada el 18 de mayo de 2011, y Solicitud de utilidad de EE.UU. de número de serie 13/300,235, presentada el 18 de noviembre de 2011. Parte del vocabulario utilizado en esta presentación puede tener antecedentes en estas referencias. Algunos de los conceptos descritos en este documento pueden entenderse mejor a la luz de los conceptos encontrados en estas referencias.

#### *Cribado de ADN fetal libre flotante que comprende sangre materna*

**[0045]** Los métodos descritos en el presente documento se pueden usar para ayudar a determinar si un niño, feto, u otro individuo diana es genéticamente relacionado con otro individuo. Esto puede hacerse en casos en los que el material genético del individuo diana se encuentra en presencia de una cantidad de material genético de otro individuo. El método puede usarse para ayudar a determinar si un feto está genéticamente unido a un supuesto padre utilizando el ADN fetal flotante que se encuentra en la sangre materna, junto con una muestra genética del padre y, opcionalmente, de la madre. El feto puede haberse originado de un óvulo de un donante de óvulos, de modo que el feto no está relacionado genéticamente con la madre en donde se está gestando. El método puede ser aplicable en casos donde la cantidad de ADN diana está en cualquier proporción con el ADN no diana; por ejemplo, el ADN diana podría representar entre el 0,000001 y el 99,999999% del ADN presente. El ADN contaminante no diana podría ser de una pluralidad de individuos; es ventajoso cuando se conocen datos genéticos de algunos o todos los individuos no diana relevantes, o donde están disponibles muestras genéticas de dichos individuos relacionados. Un método divulgado en el presente documento puede usarse para determinar datos genotípicos de un feto de sangre materna que contiene ADN fetal. También se puede usar en un caso donde hay múltiples fetos en el útero de una mujer embarazada, o donde puede haber otro ADN contaminante en la muestra, por ejemplo, de otros hermanos ya nacidos.

**[0046]** Esta técnica puede hacer uso del fenómeno de las células sanguíneas fetales que acceden a la circulación materna a través de las vellosidades de placenta. Por lo general, solo un número muy pequeño de células fetales ingresa a la circulación materna de esta manera (no lo suficiente como para producir una prueba de Kleihauer-Betke positiva para la hemorragia fetal-materna). Las células fetales se pueden clasificar y analizar mediante una variedad de técnicas para buscar secuencias de ADN particulares, pero sin los riesgos que los procedimientos invasivos tienen inherentemente. Esta técnica también puede hacer uso del fenómeno del ADN fetal flotante libre que obtiene acceso a la circulación materna mediante la liberación de ADN después de la apoptosis del tejido placentario donde 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 tan altas como 30-40% de ADN fetal.

**[0047]** La sangre puede ser extraída de 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. Además, también puede haber células sanguíneas fetales nucleadas que comprenden ADN de origen fetal, además de muchas células sanguíneas de origen materno, que típicamente no contienen ADN nuclear. Existen muchos métodos conocidos en la técnica para aislar el ADN fetal o crear fracciones enriquecidas en el ADN fetal. Por ejemplo, se ha demostrado que la cromatografía crea ciertas fracciones que están enriquecidas en el ADN fetal.

**[0048]** Una vez que la muestra de sangre materna, plasma, u otro fluido, elaborado de una manera relativamente no invasiva, y que contiene una cantidad de ADN fetal, ya sea celular o libre flotante, ya sea enriquecidos en su proporción con el ADN materno, o en su proporción original, está disponible, uno puede genotipar el ADN encontrado en dicha muestra. La sangre se puede extraer usando una aguja para extraer sangre de una vena, por ejemplo, la vena basílica. El método descrito en este documento puede usarse para determinar los datos genotípicos del feto. Por ejemplo, se puede usar para determinar el estado de ploidía en uno o más cromosomas, se puede usar para determinar la identidad de uno o un conjunto de SNP, incluidas las inserciones, deleciones y translocaciones. Se puede usar para determinar uno o más haplotipos, incluido el padre de origen de una o más características genotípicas. También se puede usar para determinar el grado de relación entre el feto y otro individuo.

**[0049]** Obsérvese que este método funciona con cualesquiera ácidos nucleicos que se pueden utilizar para cualquier genotipado y/o secuenciación de métodos, tales como la plataforma ILLUMINA INFINIUM ARRAY, AFFYMETRIX GENECHIP, ILLUMINA GENOME ANALYZER, o LIFE TECHNOLOGIES' SOLID SYSTEM, junto con los datos genotípicos medidos a partir de ellos. Esto incluye ADN flotante extraído del plasma o amplificaciones (por ejemplo, amplificación del genoma completo, PCR) del mismo; ADN genómico de otros tipos de células (por ejemplo, linfocitos humanos de sangre completa) o amplificaciones de los mismos. Para la preparación del ADN, cualquier método de extracción o purificación que genere ADN genómico adecuado para una de estas plataformas también funcionará. Este método podría funcionar igualmente bien con muestras de ARN. En una realización, el almacenamiento de las muestras se puede hacer de una manera que minimice la degradación (por ejemplo, por debajo del punto de congelación, a aproximadamente -20°C o a una temperatura más baja).

#### *Apoyo parental*

**[0050]** Algunas formas de realización pueden usarse en combinación con el método PARENTAL SUPPORT™ (PS), cuyas formas de realización se describen en la Solicitud de EE.UU. N° 11/603,406 (Publicación de EE.UU. N°: 20070184467), Solicitud de EE.UU. N° 12/076,348 (número de publicación de EE.UU.: 20080243398), solicitud de

EE.UU. 13/110,685, solicitud de PCT PCT/US09/52730 (número de publicación de PCT: WO/2010/017214), número de solicitud de PCT PCT/US10/050824 (número de publicación de PCT: WO/2011/041485), Solicitud PCT N° PCT/US2011/037018 (Publicación PCT N° WO/2011/146632), y Solicitud PCT N° PCT/US2011/61506. PARENTAL SUPPORT™ es un enfoque basado en la informática que se puede utilizar para analizar datos genéticos. Los métodos descritos aquí pueden considerarse como parte del método PARENTAL SUPPORT™. En algunas realizaciones, el método PARENTAL SUPPORT™ es una colección de métodos que pueden usarse para determinar los datos genéticos de un individuo diana, con alta precisión, de una o una pequeña cantidad de células de ese individuo, o de una mezcla de ADN que consiste en ADN del individuo diana y ADN de uno o una pluralidad de otros individuos, específicamente para determinar alelos relacionados con la enfermedad, otros alelos de interés, el estado de ploidía de uno o una pluralidad de cromosomas en el individuo diana, y/o la extensión de relación de otro individuo con el individuo diana. PARENTAL SUPPORT™ puede referirse a cualquiera de estos métodos. PARENTAL SUPPORT™ es un ejemplo de un método basado en informática.

[0051] El método PARENTAL SUPPORT™ hace uso de los 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 imperfecta medición del ADN diana, y posiblemente de uno o más individuos relacionados, junto con frecuencias de cruce basadas en la población, para reconstruir, *en silicio*, el genotipo en una pluralidad de alelos, y/o el estado de paternidad de un embrión o de cualquier célula diana, y el ADN diana en la ubicación de loci clave con un alto grado de confianza. 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 meiosis y la medición imperfecta del ADN diana, para crear hipótesis sobre qué datos genéticos pueden esperarse para diferentes situaciones, para calcular la probabilidad de cada una de las situaciones dadas los datos genéticos observados, determinando así qué situación es más probable. En algunas realizaciones, la situación que se cuestiona puede incluir si el individuo diana ha heredado un haplotipo de interés relacionado con la enfermedad, si el individuo diana ha heredado un haplotipo vinculado de fenotipo, si el individuo diana tiene uno o más cromosomas aneuploides, y/o si el individuo diana está relacionado con un individuo de interés y cuál puede ser el grado de relación. El método PARENTAL SUPPORT™ permite la limpieza de datos genéticos ruidosos. PARENTAL SUPPORT™ puede ser particularmente relevante cuando solo una pequeña fracción del material genético disponible proviene del individuo diana (por ejemplo, NPD o NPPT) y donde las mediciones directas de los genotipos son inherentemente ruidosas debido a la señal de ADN contaminante de otro individuo. El método PARENTAL SUPPORT™ es capaz de reconstruir secuencias aleloides diploides ordenadas de alta precisión en el embrión, junto con el número de copias de segmentos de cromosomas, a pesar de que las mediciones diploides no ordenadas convencionales pueden caracterizarse por altas tasas de abandono de alelos, abandonos, variables sesgos de amplificación y otros errores. El método puede emplear tanto un modelo genético subyacente como un modelo subyacente de error de medición. El modelo genético puede determinar tanto las probabilidades alélicas en cada SNP como las probabilidades de cruce entre los SNP. Las probabilidades alélicas pueden modelarse en cada SNP en función de los datos obtenidos de los padres y las probabilidades de cruce modelo entre SNP en función de los datos obtenidos de la base de datos HapMap, tal como fue desarrollado por el Proyecto Internacional HapMap. Dado el modelo genético subyacente adecuado y el modelo de error de medición, se puede usar la estimación *máxima a posteriori* (MAP), con modificaciones para la eficiencia computacional, para estimar los valores de alelo ordenados correctos en cada SNP en el embrión.

#### Definiciones

#### [0052]

*Polimorfismo de un solo nucleótido (SNP)* se refiere a un solo nucleótido que puede diferir entre los genomas de dos miembros de la misma especie. El uso del término no debe implicar ningún límite en la frecuencia con que ocurre cada variante.

*Secuencia* se refiere a una secuencia de ADN o una secuencia genética. Puede referirse a la estructura física primaria 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 la cadena complementaria a 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 particular de interés en el ADN de un individuo, que puede referirse a un SNP, el sitio de una posible inserción o eliminación, o el sitio de alguna otra variación genética relevante. Los SNP vinculados a la enfermedad también pueden referirse a loci vinculados a la enfermedad.

*Alelo polimórfico*, también "locus polimórfico", se refiere a un alelo o locus donde el genotipo varía entre los individuos dentro de una especie determinada. Algunos ejemplos de alelos polimórficos incluyen polimorfismos de un solo nucleótido, repeticiones cortas en tándem, deleciones, duplicaciones e inversiones.

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

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

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

típicamente *en silico*, sin embargo, también es posible considerar los nucleótidos físicos en una secuencia como datos genéticos codificados químicamente. Se puede decir que los datos genotípicos están "en", "de", "a", "de" o "en" la(s) persona(s). Los datos genotípicos pueden referirse a mediciones de salida de una plataforma de genotipado donde esas mediciones se realizan en material genético.

*Material genético* también "*muestra genética*" se refiere a la materia física, como el tejido o la sangre, de uno o más individuos que comprenden ADN o ARN.

*Confianza* se refiere a la probabilidad estadística de que el SNP, alelo, conjunto de alelos, llamada de ploidía o llamada de paternidad es correcto.

*Aneuploidía* se refiere al estado donde el número incorrecto de cromosomas está presente en una célula. En el caso de una célula humana somática, puede referirse al caso en donde una célula no contiene 22 pares de cromosomas autosómicos y un par de cromosomas sexuales. En el caso de un gameto humano, puede referirse al caso donde una célula no contiene uno de cada uno de los 23 cromosomas. En el caso de un solo tipo de cromosoma, puede referirse al caso en donde están presentes más o menos de dos copias de cromosomas homólogos pero no idénticos, o donde hay dos copias de cromosomas presentes que se originan del mismo padre.

*Cromosoma* puede referirse a una sola copia cromosómica, lo que significa una sola molécula de ADN de la cual hay 46 en una célula somática normal; un ejemplo es "el cromosoma 18 derivado de la madre". El cromosoma también puede referirse a un tipo de cromosoma, de los cuales hay 23 en una célula somática humana normal; un ejemplo es el "cromosoma 18".

*Monosomía* se refiere al estado en donde una célula solo contiene uno de un tipo de cromosoma.

*Disomía* se refiere al estado en donde una célula contiene dos de tipo cromosómico.

*Disomía uniparental* se refiere al estado donde una célula contiene dos de un tipo de cromosoma, y donde ambos cromosomas se originan de uno de los padres.

*Trisomía* se refiere al estado en donde una célula contiene tres de un tipo de cromosoma.

*Estado del material genético* o simplemente "estado genético" puede referirse a la identidad de un conjunto de SNP en el ADN, a los haplotipos por fases del material genético o a la secuencia del ADN, incluidas las inserciones, eliminaciones, repeticiones y mutaciones. También puede referirse al estado de ploidía de uno o más cromosomas, segmentos cromosómicos, o un conjunto de segmentos cromosómicos.

*Establecer la paternidad* o "determinar la paternidad" se refiere a establecer o determinar que un supuesto padre es o no el padre biológico de un feto gestante, o determinar o establecer la probabilidad de que un supuesto padre sea el padre biológico del feto.

*Determinación de paternidad* se refiere a la determinación de que el presunto padre es o no el padre biológico del feto. Una determinación de paternidad es el resultado de establecer, llamar o determinar la paternidad.

*Paternidad* se refiere a la identidad del padre biológico de un individuo.

*Inclusión de paternidad* se refiere al establecimiento de que un supuesto padre es el padre biológico de un feto.

*Exclusión de paternidad* se refiere a establecer que un supuesto padre no es el padre biológico de un feto.

*Presunto padre* se refiere a un hombre cuya relación paterna con un feto está en cuestión.

*Padre biológico* de un individuo se refiere al hombre cuyo material genético fue heredado por el individuo.

*Relación alélica* se refiere a la relación entre la cantidad de cada alelo en un locus polimórfico que está presente en una muestra o en un individuo. Cuando la muestra se mide por secuenciación, la relación alélica puede referirse a la relación de lecturas de secuencia que se asignan a cada alelo en el locus. Cuando la muestra se mide mediante un método de medición basado en la intensidad, la relación de alelos puede referirse a la relación de las cantidades de cada alelo presente en ese locus según lo estimado por el método de medición.

*Distribución alélica*, o "distribución de recuento de alelos" se refiere a la cantidad relativa de cada alelo que está presente para cada locus en un conjunto de loci. Una distribución alélica puede referirse a un individuo, a una muestra o a un conjunto de mediciones realizadas en una muestra. En el contexto de la secuenciación, la distribución alélica se refiere al número o número probable de lecturas que se asignan a un alelo particular para cada alelo en un conjunto de loci polimórficos. Las mediciones de alelos pueden tratarse probabilísticamente, es decir, la probabilidad de que un alelo dado esté presente para una secuencia de lectura dada es una fracción entre 0 y 1, o pueden tratarse de forma binaria, es decir, cualquier lectura dada se considera ser exactamente cero o una copia de un alelo particular.

*Sesgo alélico* se refiere al grado en que la proporción medida de alelos en un locus heterocigoto es diferente a la proporción que estaba presente en la muestra original de ADN. El grado de sesgo alélico en un locus particular es igual a la relación alélica observada en ese locus, tal como se mide, dividido por la relación de alelos en la muestra de ADN original en ese locus. El sesgo alélico puede definirse como mayor que uno, de modo que si el cálculo del grado de sesgo alélico devuelve un valor,  $x$ , que es menor que 1, entonces el grado de sesgo alélico puede reexpresarse como  $1/x$ . El sesgo alélico puede deberse al sesgo de amplificación, el sesgo de purificación o algún otro fenómeno que afecta a diferentes alelos de manera diferente.

*Cebador*, también "sonda de PCR", se refiere a una sola molécula de ADN (un oligómero de ADN) o una colección de moléculas de ADN (oligómeros de ADN) donde las moléculas de ADN son idénticas, o casi, y donde el cebador contiene una región diseñada para hibridarse con un locus polimórfico dirigido y puede contener una secuencia de cebado diseñada para permitir la amplificación por 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 se genera mediante diversos métodos, como PCR o síntesis directa, y que pretende ser complementaria a una cadena de una secuencia de ADN diana específica en una muestra. Las sondas de captura híbridas exógenas pueden agregarse a una muestra preparada e hibridarse a través de un proceso de reticulación de deantura para formar dúplex de fragmentos endógenos exógenos. Estos dúplex pueden separarse físicamente de la muestra por varios medios.

*Lectura de secuencia* se refiere a datos que representan una secuencia de bases de nucleótidos que se midieron usando un método de secuenciación clonal. La secuenciación clonal puede producir datos de secuencia que representan uno solo, o clones, o grupos de una molécula de ADN original. Una lectura de secuencia también puede tener un puntaje de calidad asociado en cada posición de base de la secuencia que indica la probabilidad de que el nucleótido se haya llamado correctamente.

*Mapeo de una secuencia de lectura* es el proceso de determinar la ubicación de origen de una secuencia de lectura en la secuencia del genoma de un organismo particular. 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.

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

*Heterocigoto* se refiere a tener alelos diferentes en los correspondientes loci cromosómicos.

*Tasa de heterocigosidad* se refiere a la tasa de individuos en la población que tienen alelos heterocigotos en un locus dado. La tasa de heterocigosidad también puede referirse a la proporción esperada o medida de alelos, en un locus dado en un individuo, o una muestra de ADN.

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

*Datos haplotípicos*, también "datos por fases" o "datos genéticos ordenados", se refieren a datos de un solo cromosoma en un genoma diploide o poliploide, es decir, la copia segregada materna o paterna de un cromosoma en un genoma diploide.

*Fase* se refiere al acto de determinar los datos genéticos haplotípicos de un individuo dado datos genéticos diploides no ordenados (o poliploides). Puede referirse al acto de determinar cuál de dos genes en un alelo, para un conjunto de alelos que se encuentran en un cromosoma, están asociados con cada uno de los dos cromosomas homólogos en un individuo.

*Vatos por fases* se refieren a datos genéticos en los que se han determinado uno o más haplotipos.

*Fetal* se refiere a "del feto" o "de la región de la placenta que es genéticamente similar al feto". En una mujer embarazada, una parte de la placenta es genéticamente similar al feto, y el ADN fetal flotante que se encuentra en la sangre materna puede haberse originado en la parte de la placenta con un genotipo que coincide con el feto.

*ADN de origen fetal* se refiere al ADN que originalmente era parte de una célula cuyo genotipo era esencialmente equivalente al del feto. Tenga en cuenta que la información genética en la mitad de los cromosomas en un feto se hereda de la madre del feto; en algunas realizaciones, el ADN de estos cromosomas heredados de la madre que provienen de una célula fetal se considera "de origen fetal" y no "de origen materno".

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

*Niño* puede referirse a un embrión, un blastómero o un feto. Tenga en cuenta que en las realizaciones descritas actualmente, los conceptos descritos se aplican igualmente bien a las personas que son un niño nacido, un feto, un embrión o un conjunto de células de los mismos. El uso del término niño puede significar simplemente que el individuo al que se hace referencia como niño es la descendencia genética de los padres.

*Padre* se refiere a la madre o padre genético de un individuo. Un individuo generalmente tiene dos padres, una madre y un padre, aunque esto no necesariamente el caso, como en el quimerismo genético o cromosómico.

*Madre* puede referirse a la madre biológica de un individuo, y/o puede referirse a las mujeres que llevan al individuo cuando él/ella se gesta.

*Contexto parental* se refiere al estado genético de un SNP dado, en cada uno de los dos cromosomas relevantes para uno o ambos padres de la diana.

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

*Decisión clínica* se refiere a cualquier decisión de tomar o no una acción que tenga un resultado que afecte la salud o la supervivencia de un individuo. En el contexto de las pruebas de paternidad prenatales, una decisión clínica puede referirse a una decisión de abortar o no abortar un feto. Una decisión clínica también puede referirse a una decisión de realizar más pruebas o tomar medidas para prepararse para el nacimiento de un niño.

*Cuadro de diagnóstico* se refiere a una o una combinación de máquinas diseñadas para realizar uno o una pluralidad de aspectos de los métodos descritos en este documento. En una realización, el cuadro de diagnóstico puede colocarse en un punto de atención al paciente. En una realización, el cuadro de diagnóstico puede realizar una amplificación dirigida seguida de secuenciación. En una realización, el cuadro de diagnóstico puede funcionar sola o con la ayuda de un técnico.

*Método basado en informática* o "enfoque basado en informática" se refiere a un método que depende en

gran medida de las estadísticas 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 en uno o más cromosomas o el estado alélico en uno o más alelos al inferir estadísticamente el estado más probable, en lugar de medir directamente el estado físico, dado una gran cantidad de datos genéticos, por ejemplo, de una matriz molecular o secuenciación. En una realización de la presente divulgación, la técnica basada en informática puede ser una descrita en esta patente. En una realización de la presente descripción, puede ser PARENTAL SUPPORT™.

*Enriquecimiento preferencial de ADN* que corresponde a uno o una pluralidad de loci, o el enriquecimiento preferencial de ADN en uno o una pluralidad de loci, se refiere a cualquier método que da como resultado el porcentaje de moléculas de ADN en una mezcla de ADN posterior al enriquecimiento que corresponde a los loci son más altos que el porcentaje de moléculas de ADN en la mezcla de ADN de pre-enriquecimiento que corresponde a los loci. El método puede implicar la amplificación selectiva de moléculas de ADN que corresponden a los loci. El método puede implicar la eliminación de moléculas de ADN que no corresponden a los loci.

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

*Amplificación selectiva* puede referirse a un método que aumenta el número de copias de una molécula particular de ADN, o moléculas de ADN que corresponden a una región particular de ADN. También puede referirse a un método que aumenta el número de copias de una molécula de ADN dirigida particular, o una región de ADN dirigida más de lo que aumenta las moléculas o regiones de ADN no dirigidas. La amplificación selectiva puede ser un método de enriquecimiento preferencial.

*Secuencia de cebado universal* se refiere a una secuencia de ADN que se puede agregar a una población de moléculas de ADN diana, por ejemplo mediante ligadura, PCR o PCR mediada por ligadura. Una vez agregados a la población de moléculas objetivo, se pueden usar cebadores específicos para las secuencias de cebado universales para amplificar la población diana usando un solo par de cebadores de amplificación. Las secuencias de cebado universales generalmente no están relacionadas con las secuencias diana.

*Adaptadores universales*, o "adaptadores de ligadura" o "etiquetas de biblioteca" son moléculas de ADN que contienen una secuencia de cebado universal que se puede unir covalentemente al extremo 5-cebado y 3-cebado de una población de moléculas de ADN bicatenario diana. La adición de los adaptadores proporciona secuencias de cebado universales al extremo 5-cebado y 3-cebado de la población diana desde la cual puede tener lugar la amplificación por PCR, amplificando todas las moléculas de la población diana, utilizando un único par de cebadores de amplificación.

*Orientación* se refiere a un método utilizado para amplificar selectivamente o enriquecer de manera preferencial aquellas moléculas de ADN que corresponden a un conjunto de loci, en una mezcla de ADN.

*Hipótesis* se refiere a la posibilidad de que el supuesto padre sea el padre biológico del feto, o que el supuesto padre no sea el padre biológico del feto.

*Determinación, establecimiento y el cálculo* pueden usarse indistintamente.

#### Contextos parentales

**[0053]** El contexto parental se refiere al estado genético de un alelo dado, en cada uno de los dos cromosomas relevantes para uno o ambos de los dos padres de la diana. Tenga en cuenta que, en una realización, el contexto parental no se refiere al estado alélico de la diana, sino que se refiere al estado alélico de los padres. El contexto parental para un SNP dado puede consistir en cuatro pares de bases, dos paternos y dos maternos; pueden ser iguales o diferentes entre sí. Normalmente se escribe como " $m_1m_2|f_1f_2$ ", donde  $m_1$  y  $m_2$  son el estado genético del SNP dado en los dos cromosomas maternos, y  $f_1$  y  $f_2$  son el estado genético del SNP en los dos cromosomas paternos. En algunas realizaciones, el contexto parental puede escribirse como " $f_1f_2|m_1m_2$ ". Tenga en cuenta que los subíndices "1" y "2" se refieren al genotipo, en el alelo dado, del primer y segundo cromosoma; También tenga en cuenta que la elección de qué cromosoma está marcado como "1" y cuál está marcado como "2" puede ser arbitrario.

**[0054]** Obsérvese que en esta descripción, A y B se utilizan a menudo para representar genéricamente base de identidades de pares; A o B podrían representar igualmente C (citocina), G (guanina), A (adenina) o T (timina). Por ejemplo, si, en un alelo basado en SNP dado, el genotipo de la madre era T en ese SNP en un cromosoma, y G en ese SNP en el cromosoma homólogo, y el genotipo del padre en ese alelo es G en ese SNP en ambos cromosomas homólogos, se puede 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. Tenga en cuenta que, en teoría, cualquiera de los cuatro nucleótidos posibles podría ocurrir en un alelo dado y, por lo tanto, es posible, por ejemplo, que la madre tenga un genotipo de AT y que el padre tenga un genotipo de GC en un determinado alelo. Sin embargo, los datos empíricos indican que en la mayoría de los casos solo se observan dos de los cuatro pares de bases posibles en un alelo dado. Es posible, por ejemplo, cuando se usan repeticiones en tándem individuales, tener más de dos contextos parentales, más de cuatro e incluso más de diez. En esta descripción, la discusión supone que solo se observarán dos posibles pares de bases en un alelo dado, aunque las realizaciones descritas en el presente documento podrían modificarse para tener en cuenta los casos en los que esta suposición no es válida.

**[0055]** Un "contexto parental" puede referirse a un conjunto o subconjunto de SNPs diana que tienen el mismo contexto parental. Por ejemplo, si uno midiera 1.000 alelos en un cromosoma dado en un individuo diana, entonces el contexto AA|BB podría referirse al conjunto de todos los alelos en el grupo de 1.000 alelos donde el genotipo de la madre del

objetivo era homocigoto, y el genotipo del padre del objetivo es homocigoto, pero donde el genotipo materno y el genotipo paterno son diferentes en ese lugar. Si los datos de los padres no están en fase, y por lo tanto  $AB \neq BA$ , entonces hay nueve posibles contextos parentales:  $AA|AA$ ,  $AA|AB$ ,  $AA|BB$ ,  $AB|AA$ ,  $AB|AB$ ,  $AB|BB$ ,  $BB|AA$ ,  $BB|AB$  y  $BB|BB$ . Si los datos parentales están en fases, y por lo tanto  $AB = BA$ , entonces hay dieciséis contextos parentales diferentes 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 en un cromosoma, excluyendo algunos SNP en los cromosomas sexuales, tiene uno de estos contextos parentales. El conjunto de SNP en donde el contexto parental para uno de los padres es heterocigoto puede denominarse el contexto heterocigoto.

#### *Diferentes implementaciones de las realizaciones actualmente divulgadas*

**[0056]** Se describen métodos en el presente documento para determinar la paternidad de un individuo diana. El individuo diana puede ser un blastómero, un embrión o un feto. En la presente divulgación, un método para determinar la paternidad de un individuo puede incluir cualquiera de los pasos descritos en este documento, y combinaciones de los mismos: la fuente del material genético que se utilizará para determinar la paternidad del feto puede ser células fetales, como los glóbulos rojos fetales nucleados, aislados de la sangre materna. El método puede implicar obtener una muestra de sangre de la madre embarazada. En la presente descripción, el material genético que se utilizará para determinar la paternidad del feto puede liberar ADN flotante del plasma materno, donde el ADN flotante libre puede estar compuesto por una mezcla de ADN fetal y materno.

**[0057]** La fuente del material genético del feto pueden ser células fetales, tales como los glóbulos rojos fetales nucleados, aislados de la sangre materna. El método puede implicar obtener una muestra de sangre de la madre embarazada. El método puede implicar aislar un glóbulo rojo fetal utilizando técnicas visuales, basadas en la idea de que una determinada combinación de colores está asociada de forma exclusiva con los glóbulos rojos nucleados, y una combinación similar de colores no está asociada con ninguna otra célula presente en la sangre de la madre. La combinación de colores asociados con los glóbulos rojos nucleados puede incluir el color rojo de la hemoglobina alrededor del núcleo, cuyo color puede hacerse más distintivo por tinción, y el color del material nuclear que puede teñirse, por ejemplo, azul. Al aislar las células de la sangre materna y esparcirlas sobre un portaobjetos, y luego identificar aquellos puntos en los que se ve tanto rojo (de la hemoglobina) como azul (del material nuclear), se puede identificar la ubicación de las células de sangre roja nucleadas. Entonces se puede extraer esos glóbulos rojos nucleados usando un micromanipulador, usar técnicas de genotipado y/o secuenciación para medir aspectos del genotipo del material genético en esas células.

**[0058]** En una realización, se puede manchar los glóbulos rojos nucleados con un troquel que sólo emite fluorescencia en presencia de la hemoglobina fetal y no hemoglobina materna, y así eliminar la ambigüedad entre si un glóbulo rojo nucleado se deriva de la madre o el feto. Algunas realizaciones de la presente divulgación pueden implicar la tinción o de otro modo marcar material nuclear. Algunas realizaciones de la presente divulgación pueden implicar marcar específicamente material nuclear fetal usando anticuerpos específicos de células fetales.

**[0059]** Hay muchas otras maneras para aislar las células fetales de la sangre materna, o el ADN fetal de la sangre materna, o para muestras Enrich de material genético fetal en la presencia de material genético materno. Algunos de estos métodos se enumeran aquí, pero esto no pretende ser una lista exhaustiva. Algunas técnicas apropiadas se enumeran aquí por conveniencia: uso de anticuerpos marcados con fluorescencia o de otro modo, cromatografía de exclusión por tamaño, etiquetas de afinidad marcadas 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 agotamiento de CD45/14 y selección positiva para CD71 de células CD45/14 negativas, gradientes de Percoll simples o dobles con diferentes osmolalidades, o método de lectina específica de galactosa.

**[0060]** En algunas realizaciones, la muestra genética puede ser preparada, aislada y/o purificada. En algunas realizaciones, la muestra puede centrifugarse para separar varias capas. En algunas realizaciones, la preparación del ADN puede implicar amplificación, separación, purificación por cromatografía, purificación por electroforesis, filtración, separación líquido-líquido, aislamiento, precipitación, enriquecimiento preferencial, amplificación preferencial, amplificación dirigida o cualquiera de una serie de otras técnicas conocidas en la técnica o descritas aquí.

**[0061]** En algunas realizaciones, el método de la presente descripción puede implicar la amplificación de ADN. La amplificación del ADN, un proceso que transforma una pequeña cantidad de material genético en una mayor cantidad de material genético que comprende un conjunto similar de datos genéticos, puede realizarse mediante una amplia variedad de métodos, que incluyen, entre otros, la reacción en cadena de la polimerasa (PCR). Un método para amplificar el ADN es la amplificación del genoma completo (WGA). Hay varios métodos disponibles para WGA: PCR mediada por ligadura (LM-PCR), PCR de cebador oligonucleotídico degenerado (DOP-PCR) y amplificación de desplazamiento múltiple (MDA). En LM-PCR, las secuencias cortas de ADN llamadas adaptadores están ligadas a extremos romos de ADN. Estos adaptadores contienen secuencias de amplificación universal, que se utilizan para amplificar el ADN por PCR. En DOP-PCR, los cebadores aleatorios que también contienen secuencias de amplificación universal se usan en una primera ronda de recocido y PCR. Luego, se utiliza una segunda ronda de PCR para amplificar más las secuencias con las secuencias de cebadores universales. MDA utiliza la polimerasa phi-29, que es una enzima altamente específica y no procesadora que replica el ADN y se ha utilizado para el análisis de

células individuales. La amplificación del genoma completo de una sola célula se ha utilizado con éxito para una variedad de aplicaciones durante varios años. Existen otros métodos para amplificar el ADN de una muestra de ADN. La amplificación de ADN transforma la muestra inicial de ADN en una muestra de ADN que es similar en el conjunto de secuencias, pero de cantidades mucho mayores. En algunos casos, la amplificación puede no ser necesaria.

**[0062]** En algunas realizaciones, el ADN se puede amplificar usando una amplificación universal, como WGA o MDA. En algunas realizaciones, el ADN puede amplificarse mediante amplificación dirigida, por ejemplo usando PCR dirigida o sondas de circularización. En algunas realizaciones, el ADN puede enriquecerse preferentemente usando un método de amplificación dirigido, o un método que da como resultado la separación total o parcial del ADN deseado del ADN no deseado, tal como la captura mediante enfoques de hibridación. En algunas realizaciones, el ADN puede amplificarse usando una combinación de un método de amplificación universal y un método de enriquecimiento preferencial. Una descripción más completa de algunos de estos métodos se puede encontrar en otra parte de este documento.

**[0063]** Los datos genéticos del individuo diana y/o de la persona relacionada puede transformarse de un estado molecular a un estado electrónico mediante la medición del material genético apropiado el uso de herramientas y/o técnicas tomadas de un grupo que incluye, pero no limitado a: microarrays de genotipado y secuenciación de alto rendimiento. Algunos métodos de secuenciación de alto rendimiento incluyen la secuenciación de ADN de Sanger, la secuencia de pirosecuenciación, la plataforma ILLUMINA SOLEXA, ILLUMINA's GENOME ANALYZER o la plataforma de secuenciación 454 de APPLIED BIOSYSTEM, la plataforma de HELICOS's TRUE SINGLE MOLECULE SEQUENCING, el método de secuenciación de microscopios HALCYON MOLECULAR o cualquier otro método de secuencia de microscopio electrónico. Todos estos métodos transforman físicamente los datos genéticos almacenados en una muestra de ADN en un conjunto de datos genéticos que generalmente se almacenan en un dispositivo de memoria en el camino al procesamiento.

**[0064]** Datos genéticos de un individuo relevante se pueden medir mediante sustancias de análisis tomadas de un grupo que incluye, pero no limitado a: tejido diploide mayor del individuo, una o más células diploides de los uno o más haploides individuales, las células de la persona, 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 sangre materna, uno o más embriones creados a partir de un gameto del individuo relacionado, uno o más blastómeros tomados de dicho embrión, material genético extracelular encontrado en el individuo relacionado, material genético que se sabe que se originó del individuo relacionado y combinaciones del mismo.

**[0065]** En algunas realizaciones, la probabilidad de que un presunto padre es el padre biológico de un feto se puede calcular. En algunas realizaciones, la determinación de paternidad se puede usar para tomar una decisión clínica. Este conocimiento, típicamente almacenado como una disposición física de la materia en un dispositivo de memoria, puede transformarse en un informe. Entonces se puede actuar sobre el informe. Por ejemplo, la decisión clínica puede ser interrumpir el embarazo; alternativamente, la decisión clínica puede ser continuar el embarazo.

**[0066]** En una realización de la presente exposición, cualquiera de los métodos descritos en el presente documento pueden ser modificados para permitir múltiples objetivos para provenir de un mismo individuo diana, por ejemplo, múltiples extracciones de sangre de la misma madre embarazada. Esto puede mejorar la precisión del modelo, ya 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 sirvió como los datos primarios que se informaron, y el otro sirvió como datos para verificar dos veces los datos genéticos diana primarios. En una realización, una pluralidad de conjuntos de datos genéticos, cada uno medido a partir de material genético tomado del individuo diana, se consideran en paralelo, y por lo tanto ambos conjuntos de datos genéticos diana sirven para ayudar a determinar la paternidad del feto.

**[0067]** En una realización, el método puede ser utilizado para el propósito de la prueba de paternidad. Por ejemplo, dada la información genotípica basada en SNP de la madre, y de un hombre que puede o no ser el padre genético, y la información genotípica medida de la muestra mixta, es posible determinar si la información genotípica del hombre realmente representa ese padre genético real del feto gestante. Una manera simple de hacer esto es simplemente mirar los contextos donde la madre es AA y el posible padre es AB o BB. En estos casos, uno puede esperar ver que el padre contribuya la mitad (AA|AB) o la totalidad (AA|BB) del tiempo, respectivamente. Teniendo en cuenta la ADO esperada, es sencillo determinar si los SNP fetales que se observan están correlacionados o no con los del posible padre. Otros métodos para hacer una determinación de paternidad se describen en otra parte de este documento.

**[0068]** En una realización de la presente descripción, una madre embarazada quisiera determinar si un hombre es el padre biológico de su feto. Ella va a su médico, y le da una muestra de su sangre, y ella y su esposo le dan muestras de su propio ADN de hisopos. Un investigador de laboratorio genotipifica el ADN de los padres utilizando el protocolo MDA para amplificar el ADN de los padres y las matrices de ILLUMINA INFINIUM para medir los datos genéticos de los padres en una gran cantidad de SNP. Luego, el investigador hace girar la sangre, toma el plasma y aísla una muestra de ADN flotante mediante cromatografía de exclusión por tamaño. Alternativamente, el investigador usa uno o más anticuerpos fluorescentes, como uno específico de la hemoglobina fetal para aislar un glóbulo rojo fetal



nucleado. Luego, el investigador toma el material genético fetal aislado o enriquecido y lo amplifica utilizando una biblioteca de oligonucleótidos de 70 mers adecuadamente diseñados de tal manera que dos extremos de cada oligonucleótido corresponden a las secuencias flanqueantes a cada lado de un alelo diana. Tras la adición de una polimerasa, ligasa y los reactivos apropiados, los oligonucleótidos se sometieron a una circularización de llenado de huecos, capturando el alelo deseado. Se añadió una exonucleasa, inactivada por calor, y los productos se usaron directamente como plantilla para la amplificación por PCR. Los productos de PCR se secuenciaron en un ILLUMINA GENOME ANALYZER. Las lecturas de secuencia se usaron como entrada para el método PARENTAL SUPPORT™, que luego predijo el estado de ploidía del feto. El método determina que el supuesto padre no es el padre biológico del feto, y calcula una confianza en la determinación del 99,98%. Se genera un informe que revela tanto la determinación de paternidad como la confianza de la determinación.

**[0069]** En otra realización, una mujer que está embarazada desea saber si un hombre es el padre biológico de su feto. El obstetra toma una muestra de sangre de la madre y el 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 en 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 aún más de los datos genéticos almacenados molecularmente en datos genéticos almacenados electrónicamente ejecutando el material genético en un secuenciador de alto rendimiento para medir los genotipos parentales. La muestra de plasma se enriquece preferentemente en un conjunto de loci utilizando un método de PCR dirigido hemi-anidado de 5.000 plex. La mezcla de fragmentos de ADN se prepara en una biblioteca de ADN adecuada para la secuenciación. El ADN se secuencia luego usando 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 el hardware de la computadora. Se puede usar una técnica basada en la informática que incluye las realizaciones descritas actualmente, como PARENTAL SUPPORT™, para determinar la paternidad del feto. Esto puede implicar calcular, en una computadora, recuentos de alelos en la pluralidad de loci polimórficos a partir de las mediciones de ADN realizadas en la muestra enriquecida; y determinar la probabilidad de que el hombre sea el padre biológico de su feto. Se determina que la probabilidad de que el supuesto padre sea el padre biológico del feto es del 99,9999%, y la confianza de la determinación de paternidad se calcula en el 99,99%. Se imprime un informe o se envía electrónicamente al obstetra de la mujer embarazada, quien transmite la determinación a la mujer. La mujer, su esposo y el médico se sientan y discuten el informe.

**[0070]** En una realización, el material genético en bruto de la madre y el padre se transforma por medio de la amplificación a una cantidad de ADN que es similar en secuencia, pero más grande en cantidad. Luego, mediante un método de genotipado, los datos genotípicos codificados por los ácidos nucleicos se transforman en mediciones genéticas que pueden almacenarse física y/o electrónicamente en un dispositivo de memoria, como los descritos anteriormente. Los algoritmos relevantes que componen el algoritmo PARENTAL SUPPORT™, cuyas partes relevantes se analizan en detalle en este documento, se traducen a un programa informático, utilizando un lenguaje de programación. Luego, a través de la ejecución del programa de computadora en el hardware de la computadora, en lugar de ser bits y bytes físicamente codificados, dispuestos en un patrón que representa los datos de medición sin procesar, se transforman en un patrón que representa una determinación de alta confianza de la paternidad del feto. Los detalles de esta transformación dependerán de los datos en sí mismos y del lenguaje informático y del sistema de hardware utilizado para ejecutar el método descrito en este documento. Luego, los datos que están configurados físicamente para representar una determinación de paternidad de alta calidad del feto se transforman en un informe que puede enviarse a un profesional de la salud. Esta transformación puede llevarse a cabo utilizando una impresora o una pantalla de computadora. El informe puede ser una copia impresa, en papel u otro medio adecuado, o puede ser electrónico. En el caso de un informe electrónico, puede transmitirse, puede almacenarse físicamente en un dispositivo de memoria en una ubicación en la computadora accesible para el profesional de la salud; también puede mostrarse en una pantalla para que pueda leerse. En el caso de una visualización en pantalla, los datos pueden transformarse a un formato legible al provocar la transformación física de píxeles en el dispositivo de visualización. La transformación puede lograrse disparando electrones físicamente a una pantalla fosforescente, alterando una carga eléctrica que cambia físicamente la transparencia de un conjunto específico de píxeles en una pantalla que puede estar frente a un sustrato que emite o absorbe fotones. Esta transformación se puede lograr cambiando la orientación a nanoescala de las moléculas en un cristal líquido, por ejemplo, de fase nemática a colestérica o esméctica, en un conjunto específico de píxeles. Esta transformación puede lograrse mediante una corriente eléctrica que hace que se emitan fotones desde un conjunto específico de píxeles formados por una pluralidad de diodos emisores de luz dispuestos en un patrón significativo. Esta transformación se puede lograr por cualquier otra forma utilizada para mostrar información, como una pantalla de computadora, o algún otro dispositivo de salida o forma de transmitir información. El profesional de la salud puede actuar sobre el informe, de modo que los datos del informe se transformen en una acción. La acción puede ser continuar o interrumpir el embarazo, en cuyo caso un feto gestante se transforma en un feto no vivo. Alternativamente, uno puede transformar un conjunto de mediciones genotípicas en un informe que ayude a un médico a tratar a su paciente embarazada.

**[0071]** En algunas realizaciones, los métodos descritos en el presente documento pueden usarse en una edad muy temprana de gestación, por ejemplo tan pronto como cuatro semanas, tan pronto como cinco semanas, tan pronto como seis semanas, tan pronto como siete semanas, tan pronto como ocho semanas, tan pronto como nueve semanas, tan pronto como diez semanas, tan pronto como once semanas, y tan pronto como doce semanas.

**[0072]** Cualquiera de las realizaciones descritas en el presente documento puede implementarse en circuitos electrónicos digitales, circuitos integrados, ASIC especialmente diseñados (circuitos integrados específicos de la aplicación), hardware de computadora, firmware, software o en combinaciones de los mismos. El aparato de las realizaciones descritas actualmente puede implementarse en un producto de programa informático incorporado de forma tangible en un dispositivo de almacenamiento legible por máquina para su ejecución por un procesador programable; y los pasos del método de las realizaciones divulgadas actualmente pueden ser realizadas por un procesador programable que ejecuta un programa de instrucciones para realizar funciones de las realizaciones divulgadas actualmente operando datos de entrada y generando salida. Las realizaciones descritas actualmente pueden implementarse ventajosamente en uno o más programas informáticos que son ejecutables y/o interpretables en un sistema programable que incluye al menos un procesador programable, que puede tener un propósito especial o general, acoplado para recibir datos e instrucciones de, y para transmitir datos e instrucciones a un sistema de almacenamiento, al menos un dispositivo de entrada y al menos un dispositivo de salida. Cada programa de computadora puede implementarse en un lenguaje de programación orientado a objetos o de procedimientos de alto nivel o en lenguaje ensamblador o de máquina si se desea; y, en cualquier caso, el idioma puede ser un lenguaje compilado o interpretado. Un programa de computadora puede implementarse en cualquier forma, incluso como un programa independiente, o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Un programa de computadora puede implementarse para ser ejecutado o interpretado en una computadora o en varias computadoras en un sitio, o distribuido en múltiples sitios e interconectado por una red de comunicación.

**[0073]** Los medios de almacenamiento legibles por computadora, como se usan en este documento, se refieren al 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 computadora, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento legibles por computadora incluyen, entre otros, RAM, ROM, EPROM, EEPROM, memoria flash u otra tecnología de memoria de estado sólido, CD-ROM, DVD u otro almacenamiento óptico, casetes magnéticos, cinta magnética, almacenamiento en disco magnético o otros dispositivos de almacenamiento magnético, o cualquier otro medio físico o material que se pueda utilizar para almacenar de manera tangible la información o datos o instrucciones deseadas y al que pueda acceder una computadora o procesador.

#### *Enriquecimiento y secuenciación dirigidos*

**[0074]** El uso de una técnica para enriquecer una muestra de ADN en un conjunto de loci diana seguido de secuenciación como parte de un método para la llamada de alelos prenatales no invasivos o llamadas de ploidía puede conferir una serie de ventajas inesperadas. En algunas realizaciones de la presente divulgación, el método implica medir datos genéticos para su uso con un método basado en informática, tal como PARENTAL SUPPORT™ (PS). El resultado final de algunas de las realizaciones son los datos genéticos accionables de un embrión o un feto. Hay muchos métodos que pueden usarse para medir los datos genéticos del individuo y/o los individuos relacionados como parte de los métodos incorporados. En una realización, se describe un método para enriquecer la concentración de un conjunto de alelos dirigidos en el presente documento, el método comprende uno o más de los siguientes pasos: amplificación dirigida de material genético, adición de sondas de oligonucleótidos específicos de loci, ligadura de cadenas de ADN específicas, aislamiento de conjuntos de ADN deseado, eliminación de componentes no deseados de una reacción, detección de ciertas secuencias de ADN por hibridación y detección de la secuencia de una o una pluralidad de cadenas de ADN por métodos de secuenciación de ADN. En algunos casos, las cadenas de ADN pueden referirse al material genético diana, en algunos casos pueden referirse a cebadores, en algunos casos pueden referirse a secuencias sintetizadas, o combinaciones de las mismas. Estos pasos pueden llevarse a cabo en diferentes órdenes. Dada la naturaleza altamente variable de la biología molecular, generalmente no es obvio qué métodos, y qué combinaciones de pasos, funcionarán mal, bien o mejor en diversas situaciones.

**[0075]** Por ejemplo, un paso de amplificación universal de la ADN antes de la amplificación específica puede conferir varias ventajas, tales como eliminando el riesgo de cuellos de botella y reducir el sesgo alélico. El ADN puede mezclarse con una sonda de oligonucleótidos que puede hibridarse con dos regiones vecinas de la secuencia diana, una a cada lado. Después de la hibridación, los extremos de la sonda se pueden conectar agregando una polimerasa, un medio para la ligadura y cualquier reactivo necesario para permitir la circularización de la sonda. Después de la circularización, se puede agregar una exonucleasa para digerir el material genético no circularizado, seguido de la detección de la sonda circularizada. El ADN puede mezclarse con cebadores de PCR que pueden hibridarse con dos regiones vecinas de la secuencia diana, una a cada lado. Después de la hibridación, los extremos de la sonda pueden conectarse agregando una polimerasa, un medio para la ligadura y cualquier reactivo necesario para completar la amplificación por PCR. El ADN amplificado o no amplificado puede ser dirigido por sondas de captura híbridas que se dirigen a un conjunto de loci; después de la hibridación, la sonda puede localizarse y separarse de la mezcla para proporcionar una mezcla de ADN enriquecida en secuencias diana.

**[0076]** En algunas realizaciones, la detección del material genético diana se puede realizar de una manera multiplexada. El número de secuencias genéticas diana que pueden ejecutarse en paralelo puede variar de uno a diez, diez a cien, cien a mil, mil a diez mil, diez mil a cien mil, cien mil a un millón, o de un millón a diez millones. Tenga en cuenta que la técnica anterior incluye divulgaciones de reacciones de PCR multiplexadas exitosas que involucran

grupos de hasta aproximadamente 50 o 100 cebadores, y no más. Los intentos anteriores de multiplexar más de 100 cebadores por grupo han dado como resultado problemas significativos con reacciones secundarias no deseadas, como la formación de cebador-dímero.

**[0077]** En algunas realizaciones, este método puede ser utilizado para determinar el genotipo de una sola célula, un pequeño número de células, dos a cinco células, seis a diez células, diez a veinte células, veinte a cincuenta células, cincuenta a cien células, cien a mil células, o una pequeña cantidad de ADN extracelular, por ejemplo de uno a diez picogramos, de diez a cien picogramos, de cien picogramos a un nanogramo, de uno a diez nanogramos, de diez a cien nanogramos, o de cien nanogramos a un microgramo.

**[0078]** El uso de un método para orientar ciertos loci seguidos de la secuenciación como parte de un método para el alelo de llamada o ploidía que llama puede conferir un número de ventajas inesperadas. Algunos métodos por los cuales el ADN puede ser dirigido, o preferentemente enriquecido, incluyen el uso de sondas de circularización, sondas invertidas unidas (LIP, MIP), captura mediante métodos de hibridación como SURESELECT y estrategias de amplificación de PCR mediadas por PCR o ligadura.

**[0079]** En algunas realizaciones, un método de la presente descripción implica la medición de los datos genéticos para uso con un método basado, tal como PARENTAL SUPPORT™ (PS). PARENTAL SUPPORT™ es un enfoque basado en la informática para manipular datos genéticos, cuyos aspectos se describen en este documento. El resultado final de algunas de las realizaciones son los datos genéticos accionables de un embrión o un feto seguidos de una decisión clínica basada en los datos accionables. Los algoritmos detrás del método PS toman los datos genéticos medidos del individuo diana, a menudo un embrión o feto, y los datos genéticos medidos de individuos relacionados, y pueden aumentar la precisión con la que se conoce el estado genético del individuo diana. En una realización, los datos genéticos medidos se usan en el contexto de hacer determinaciones de paternidad durante el diagnóstico genético prenatal. Existen muchos métodos que pueden usarse para medir los datos genéticos del individuo y/o los individuos relacionados en los contextos mencionados anteriormente. Los diferentes métodos comprenden varios pasos, que a menudo implican la amplificación de material genético, la adición de sondas de oligonucleótidos, la ligadura de cadenas de ADN específicas, el aislamiento de conjuntos de ADN deseado, la eliminación de componentes no deseados de una reacción, la detección de ciertas secuencias de ADN por hibridación, detección de la secuencia de una o una pluralidad de cadenas de ADN mediante métodos de secuenciación de ADN. En algunos casos, las cadenas de ADN pueden referirse al material genético diana, en algunos casos pueden referirse a cebadores, en algunos casos pueden referirse a secuencias sintetizadas o combinaciones de las mismas. Estos pasos pueden llevarse a cabo en diferentes órdenes. Dada la naturaleza altamente variable de la biología molecular, generalmente no es obvio qué métodos, y qué combinaciones de pasos, funcionarán mal, bien o mejor en diversas situaciones.

**[0080]** Algunas realizaciones de la presente descripción implican el uso de "Linked Inverted Probes" (labios), que han sido descritos previamente en la literatura. LIP es un término genérico destinado a abarcar tecnologías que implican la creación de una molécula circular de ADN, donde las sondas están diseñadas para hibridarse con la región diana de ADN en cualquier lado de un alelo dirigido, de modo que se agreguen las polimerasas y/o ligasas apropiadas, y las condiciones apropiadas, tampones y otros reactivos, completarán la región de ADN invertida complementaria a través del alelo diana para crear un bucle circular de ADN que capture la información encontrada en el alelo diana. Los LIP también pueden denominarse sondas pre-circularizadas, sondas pre-circulares o sondas circulares. La sonda de LIP puede ser una molécula de ADN lineal entre 50 y 500 nucleótidos de longitud, y en una realización entre 70 y 100 nucleótidos de longitud; en algunas realizaciones, puede ser más largo o más corto que el descrito aquí. Otras realizaciones de la presente divulgación implican diferentes realizaciones de la tecnología de LIP, tales como las sondas de candado y las sondas de inversión molecular (MIP).

**[0081]** Un método para lugares específicos diana para la secuenciación es sintetizar sondas en las que el 3' y 5' de las sondas hibridan con ADN diana en lugares adyacentes a y a cada lado de la región diana, de una manera invertida, de modo que la adición de ADN polimerasa y ADN ligasa da como resultado la extensión desde el extremo 3', agregando bases a la sonda monocatenaria que son complementarias a la molécula diana (rellenado de espacio), seguido de la ligadura del nuevo extremo 3' al extremo 5' de la sonda original que da como resultado una molécula de ADN circular que puede aislarse posteriormente del ADN de fondo. Los extremos de la sonda están diseñados para flanquear la región de interés diana. Un aspecto de este enfoque se denomina comúnmente MIPS y se ha utilizado en conjunción con las tecnologías de la matriz para determinar la naturaleza de la secuencia llenada.

**[0082]** PCR mediada por la ligadura es el método de PCR utilizado para enriquecer preferentemente una muestra de ADN mediante la amplificación de uno o una pluralidad de loci en una mezcla de ADN, el método comprende: obtener un conjunto de pares de cebadores, donde cada cebador en el par contiene una secuencia específica diana y una secuencia no diana, donde la secuencia específica diana está diseñada para reconocer una región diana, una aguas arriba y una aguas abajo del sitio polimórfico; polimerización del ADN desde el extremo 3-cebado del cebador aguas arriba para llenar la región de cadena sencilla entre él y el extremo 5-cebado del cebador aguas abajo con nucleótidos complementarios a la molécula diana; ligación de la última base polimerizada del cebador aguas arriba a la base adyacente de 5-cebado del cebador aguas abajo; y amplificación de solo moléculas polimerizadas y ligadas usando las secuencias no diana contenidas en el extremo 5 del cebador aguas arriba y el extremo 3 del cebador aguas abajo. Los pares de cebadores diana distintos pueden mezclarse en la misma reacción. Las secuencias no diana sirven como

secuencias universales de modo que todos los pares de cebadores que se han polimerizado y ligado con éxito pueden amplificarse con un solo par de cebadores de amplificación.

[0083] En una realización, una muestra de ADN se puede enriquecer preferentemente utilizando una captura por el enfoque de hibridación. Algunos ejemplos de captura comercial mediante tecnologías de hibridación incluyen SURESELECT de AGILENT y TRUSEQ de ILLUMINA. En la captura por hibridación, un conjunto de oligonucleótidos que es complementario o mayormente complementario de las secuencias deseadas se hibrida con una mezcla de ADN y luego se separa físicamente de la mezcla. Una vez que las secuencias deseadas se han hibridado con los oligonucleótidos de direccionamiento, el efecto de eliminar físicamente los oligonucleótidos de direccionamiento es eliminar también las secuencias diana. Una vez que se eliminan los oligos hibridados, se pueden calentar por encima de su temperatura de fusión y se pueden amplificar. Algunas formas de eliminar físicamente los oligonucleótidos de direccionamiento es uniéndolos covalentemente los oligos de direccionamiento a un soporte sólido, por ejemplo, un cordón magnético o un chip. Otra forma de eliminar físicamente los oligonucleótidos de direccionamiento es uniéndolos covalentemente a un resto molecular con una fuerte afinidad por otro resto molecular. Un ejemplo de dicho par molecular es la biotina y la estreptavidina, como se usa en SURESELECT. Por lo tanto, las secuencias dirigidas podrían unirse covalentemente a una molécula de biotina, y después de la hibridación, se puede usar un soporte sólido con estreptavidina fijada para derribar los oligonucleótidos biotinilados, a los que se hibridan con las secuencias diana.

[0084] En algunas realizaciones, la PCR se puede utilizar para ubicaciones específicas diana del genoma. En muestras de plasma, el ADN original está altamente fragmentado (típicamente menos de 500 pb, con una longitud promedio inferior a 200 pb). En la PCR, los cebadores directos e inversos deben reconocerse en el mismo fragmento para permitir la amplificación. Por lo tanto, si los fragmentos son cortos, los ensayos de PCR también deben amplificar regiones relativamente cortas. El ensayo de PCR se puede generar en grandes cantidades, sin embargo, las interacciones entre diferentes ensayos de PCR hacen que sea difícil multiplexarlos más allá de aproximadamente cien ensayos. Se pueden usar varios enfoques moleculares complejos para aumentar el nivel de multiplexación, pero aún se puede limitar a menos de 100, quizás 200 o posiblemente 500 ensayos por reacción. Las muestras con grandes cantidades de ADN pueden dividirse entre múltiples reacciones secundarias y luego recombinarse antes de la secuenciación. Para las muestras donde la muestra general 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 puede referirse 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. Tenga en cuenta que si bien este método es particularmente útil en pequeñas cantidades de ADN donde otros métodos que implican dividirse en múltiples grupos pueden causar problemas significativos relacionados con el ruido estocástico introducido, este método aún proporciona el beneficio de minimizar el sesgo cuando se ejecuta en muestras de cualquier cantidad de ADN. En estas situaciones, se puede usar un paso de preamplificación universal para aumentar la cantidad total de muestra. Idealmente, este paso de preamplificación no debería alterar apreciablemente las distribuciones alélicas.

[0085] En general, para llevar a cabo la secuenciación dirigida de múltiples dianas (n) de una muestra (mayor que 50, mayor que 100, mayor que 500, o mayor que 1.000), se puede dividir la muestra en una serie de reacciones paralelas que amplificar una o un número menor de dianas individuales. Esto se ha realizado en placas de pocillos múltiples de PCR o se puede hacer en plataformas comerciales como la FLUIDIGM ACCESS ARRAY (48 reacciones por muestra en chips microfluídicos) o DROPLET PCR por RAIN DANCE TECHNOLOGY (100s a unos pocos miles de dianas). Desafortunadamente, estos métodos de división y agrupación son problemáticos para las muestras con una cantidad limitada de ADN, ya que a menudo no hay suficientes copias del genoma para garantizar que haya una copia de cada región del genoma en cada pocillo. Esto es un problema especialmente grave cuando los loci polimórficos están dirigidos, y se necesitan las proporciones relativas de los alelos en los loci polimórficos, ya que el ruido estocástico introducido por la división y la agrupación causará mediciones muy poco precisas de las proporciones de los alelos que estuvieron presentes en la muestra original de ADN. Aquí se describe un método para amplificar de manera efectiva y eficiente muchas reacciones de PCR que es aplicable a casos en los que solo está disponible una cantidad limitada de ADN. En una realización, el método puede aplicarse para el análisis de células individuales, fluidos corporales, mezclas de ADN tales como el ADN flotante libre que se encuentra en plasma materno, biopsias, muestras ambientales y/o forenses.

[0086] En una realización, la secuencia específica puede implicar uno, una pluralidad o todos los siguientes pasos. a) Generar y amplificar universalmente una biblioteca con secuencias adaptadoras en ambos extremos de los fragmentos de ADN. b) Dividir en múltiples reacciones después de la amplificación de la biblioteca. c) Realizar una amplificación de aproximadamente 100-plex, aproximadamente 1000-plex, o aproximadamente 10.000-plex de dianas seleccionadas utilizando un cebador "directo" específico de diana por objetivo y un cebador específico de etiqueta. d) Realizar una segunda amplificación de este producto usando cebadores específicos de diana "reverso" y uno (o más) cebadores específicos para una etiqueta universal que se introdujo como parte de los cebadores directos específicos de diana en la primera ronda. e) Dividir el producto en múltiples partes alícuotas y amplificar las subagrupaciones de dianas en reacciones individuales (por ejemplo, 50 a 500-plex, aunque esto puede usarse hasta un solo complejo). f) Agrupar productos de reacciones de subagrupaciones paralelos. Durante estas amplificaciones, los cebadores pueden llevar etiquetas compatibles de secuencia (longitud parcial o total) de modo que los productos puedan secuenciarse.

[0087] En una forma de realización, es posible mitigar las pérdidas potenciales en las etapas posteriores mediante la

amplificación de la totalidad o una fracción de la muestra de ADN libre de células original (cfADN). Hay varios métodos disponibles para amplificar todo el material genético en una muestra, aumentando la cantidad disponible para los procedimientos posteriores. En una realización, los fragmentos de ADN de PCR mediada por ligamiento (LMPCR) se amplifican por PCR después de la ligadura de uno de los adaptadores distintos, dos adaptadores distintos o muchos adaptadores distintos. En una realización, se usa la polimerasa phi-29 de amplificación de desplazamiento múltiple (MDA) para amplificar todo el ADN isotérmicamente. En DOP-PCR y variaciones, se utiliza cebado aleatorio para amplificar el material de ADN original. Cada método tiene ciertas características, como la uniformidad de la amplificación en todas las regiones representadas del genoma, la eficiencia de captura y amplificación del ADN original y el rendimiento de la amplificación en función de la longitud del fragmento.

**[0088]** El diseño de ensayo tradicional PCR resulta en pérdidas significativas de moléculas fetales distintas, pero las pérdidas pueden ser considerablemente reducidas por el diseño de los ensayos de PCR muy cortos, denominados ensayos de mini-PCR. El ADNc fetal en el suero materno está altamente fragmentado y los tamaños de los fragmentos se distribuyen de manera aproximadamente gaussiana con una media de aproximadamente 160 pb, una desviación estándar de aproximadamente 15 pb, un tamaño mínimo de aproximadamente 100 pb y un tamaño máximo de aproximadamente 220 bp. La distribución de las posiciones inicial y final del fragmento con respecto a los polimorfismos diana, aunque no necesariamente al azar, varía ampliamente entre objetivos individuales y entre todas las dianas colectivamente y el sitio polimórfico de un locus diana particular puede ocupar cualquier posición desde el principio hasta el final entre los diversos fragmentos que se originan en ese lugar. Tenga en cuenta que el término mini-PCR puede referirse igualmente a la PCR normal sin restricciones ni limitaciones adicionales.

**[0089]** Durante la PCR, la amplificación solo se producirá a partir de fragmentos de ADN de plantilla que comprenden sitios de cebador directo e inverso. Debido a que los fragmentos de ADNc fetal son cortos, la probabilidad de que estén presentes ambos sitios cebadores, la probabilidad de que un fragmento fetal de longitud L que comprenda tanto los sitios cebadores hacia adelante como hacia atrás sea la relación de la longitud del amplicón 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 amplificarán con éxito de aproximadamente 72%, 69%, 66%, 63%, 59% o 56%, respectivamente, de moléculas de fragmentos de plantilla disponibles. La longitud del amplicón es la distancia entre los extremos de 5 cebadores de los sitios de cebado directo e inverso. La longitud del amplicón que es más corta que la utilizada típicamente por los conocidos en la técnica puede dar como resultado mediciones más eficientes de los loci polimórficos deseados al requerir solo lecturas de secuencia cortas. En una realización, una fracción sustancial de los amplicones debe ser inferior a 100 pb, inferior a 90 pb, inferior a 80 pb, inferior a 70 pb, inferior a 65 pb, inferior a 60 pb, inferior a 55 pb, inferior a 50 pb, o menos de 45 pb.

**[0090]** Obsérvese que en los métodos conocidos en la técnica anterior, los ensayos cortos tales como los descritos en el presente documento se evitan por lo general debido a que no son necesarios y que imponen considerable restricción en el diseño de cebadores mediante la limitación de la longitud del cebador, el recocido de características, y la distancia entre el cebador directo e inverso.

**[0091]** PCR multiplex puede implicar una sola ronda de PCR en donde se amplifican todas las dianas o puede implicar una ronda de PCR seguida de una o más rondas de PCR anidada o alguna variante de PCR anidada. La PCR anidada consiste en 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 espurias al amplificar, en reacciones posteriores, solo aquellos productos de amplificación del anterior que tienen la secuencia interna correcta. La reducción de dianas de amplificación espurias mejora el número de mediciones útiles que se pueden obtener, especialmente en la secuenciación. La PCR anidada típicamente implica diseñar cebadores completamente internos a los sitios de unión del cebador anteriores, lo que necesariamente aumenta el tamaño mínimo de segmento de ADN requerido para la amplificación. Para muestras como el ADNc de plasma materno, en donde el ADN está altamente fragmentado, el tamaño de ensayo más grande reduce el número de moléculas de ADNc distintas a partir de las cuales se puede obtener una medición. En una realización, para compensar este efecto, uno puede usar un enfoque de anidamiento parcial en donde uno o ambos cebadores de la segunda ronda se superponen a los primeros sitios de unión que se extienden internamente un cierto número de bases para lograr una especificidad adicional mientras aumentan mínimamente el tamaño total del ensayo.

**[0092]** En una realización, una agrupación multiplex de los ensayos de PCR están diseñados para amplificar SNP potencialmente heterocigotos o 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 de PCR puede estar entre 50 y 200 ensayos de PCR, entre 200 y 1.000 ensayos de PCR, entre 1.000 y 5.000 ensayos de PCR, o entre 5.000 y 20.000 ensayos de PCR (50 a 200-plex, 200 a 1.000-plex, 1.000 a 5.000-plex, 5.000 a 20.000-plex, más de 20.000-plex respectivamente).

**[0093]** En una realización, una agrupación de ensayos de PCR 100-plex a 500-plex, 500-plex a 1000-plex, 1000-plex a 2000-plex, 2000-plex a 5000-plex, 5000-plex a 10000-plex, 10.000-plex a 20.000-plex, 20.000-plex a 50.000-plex, o 50.000-plex a 100.000-plex se crea de tal manera que los cebadores directo e inverso tienen colas que corresponden a las secuencias directas e inversas requeridas por un instrumento de secuenciación de alto rendimiento como HISEQ, GAIIIX o MISEQ disponibles de ILLUMINA. Además, se incluye 5-cebado a las colas de secuenciación es una

secuencia adicional que se puede utilizar como sitio de cebado en una PCR posterior para agregar secuencias de códigos de barras de nucleótidos a los amplicones, lo que permite la secuenciación múltiple de múltiples muestras en un solo carril del instrumento de secuenciación de alto rendimiento. En una realización, se crea un grupo de ensayo de PCR de 10.000 plex de tal manera que los cebadores inversos tienen colas que corresponden a las secuencias inversas requeridas por un instrumento de secuenciación de alto rendimiento. Después de la amplificación con el primer ensayo de 10.000 plex, se puede realizar una amplificación por PCR posterior usando otro grupo de 10.000 plex que tiene cebadores directos parcialmente anidados (por ejemplo, 6 bases anidadas) para todas las dianas y un cebador inverso correspondiente a la cola de secuencia inversa incluida en la primera ronda. Esta ronda posterior de amplificación parcialmente anidada con solo un 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 espurios. Las etiquetas de secuenciación se pueden agregar a los adaptadores de ligadura adjuntos y/o como parte de las sondas de PCR, de modo que la etiqueta sea parte del amplicón final.

**[0094]** La fracción fetal afecta al rendimiento de la prueba; es más difícil determinar la paternidad correcta en muestras con una fracción fetal más baja. Hay varias formas de enriquecer la fracción fetal del ADN que se encuentra en el plasma materno. La fracción fetal puede incrementarse mediante el método de LM-PCR descrito anteriormente, ya discutido, así como mediante la eliminación selectiva de fragmentos maternos largos. En la realización, los fragmentos más largos se eliminan usando técnicas de selección de tamaño. En una realización, antes de la amplificación por PCR multiplex de los loci diana, se puede llevar a cabo una reacción de PCR multiplex adicional para eliminar selectivamente fragmentos largos y en gran parte maternos correspondientes a los loci diana en la PCR multiplex subsiguiente. Los cebadores adicionales están diseñados para recoger un sitio a una mayor distancia del polimorfismo de lo que se espera que esté presente entre los fragmentos de ADN fetal libres de células. Estos cebadores pueden usarse en una reacción de PCR multiplex de un ciclo antes de la PCR multiplex de los loci polimórficos diana. Estos cebadores distales están etiquetados con una molécula o resto que puede permitir el reconocimiento selectivo de los fragmentos de ADN etiquetados. En una realización, estas moléculas de ADN pueden modificarse covalentemente con una molécula de biotina que permite la eliminación del ADN bicatenario recién formado que comprende estos cebadores después de un ciclo de PCR. El ADN bicatenario formado durante esa primera ronda es probablemente de origen materno. La eliminación del material híbrido se puede lograr mediante el uso de perlas de estreptavidina magnéticas. Existen otros métodos de etiquetado que pueden funcionar igualmente bien. En una realización, los métodos de selección de tamaño pueden usarse para enriquecer la muestra para cadenas más cortas de ADN; por ejemplo, aquellos de menos de aproximadamente 800 pb, menos de aproximadamente 500 pb o menos de aproximadamente 300 pb. La amplificación de fragmentos cortos puede proceder como de costumbre.

**[0095]** En algunas realizaciones, el ADN diana puede originarse a partir de células individuales, a partir de muestras de ADN que consiste en menos de una copia de la meta del genoma, a partir de bajas cantidades de ADN, a partir de ADN de origen mixto (por ejemplo, plasma de embarazo: ADN de placenta y materna; plasma y tumores de pacientes con cáncer: mezcla entre ADN sano y cáncer, trasplante, etc.), de otros fluidos corporales, de cultivos celulares, de sobrenadantes de cultivo, de muestras forenses de ADN, de muestras antiguas de ADN (por ejemplo, insectos atrapados en ámbar), de otras muestras de ADN, y combinaciones de los mismos.

**[0096]** En algunas realizaciones, los métodos descritos en este documento pueden usarse para amplificar y/o detectar SNPs, individuales repeticiones en tándem (STRs), número de copias, la metilación de nucleótidos, los niveles de mRNA, otros tipos de niveles de expresión de ARN, otras características genéticas y/o epigenéticas. Los métodos de mini-PCR descritos en el presente documento pueden usarse junto con la secuenciación de próxima generación; se puede utilizar con otros métodos de aguas abajo, tales como microarrays, a contar por PCR digital en tiempo real PCR, análisis de espectrometría de masas etc.

**[0097]** En alguna realización, los métodos de amplificación mini-PCR descritos en este documento pueden usarse como parte de un método para cuantificación precisa de las poblaciones minoritarias. Puede usarse para cuantificación absoluta utilizando calibradores de espiga. Se puede usar para la mutación/menor cuantificación de alelo a través de secuenciación muy profunda, y puede ejecutarse en un multiplexado muy de moda. Puede usarse para pruebas estándar de paternidad e identidad de parientes o antepasados, en humanos, animales, plantas u otras criaturas. Se puede usar para pruebas forenses. Se puede usar para el genotipo rápido y el análisis de número de copias (CN), en cualquier tipo de material, por ejemplo, líquido amniótico y CVS, esperma, producto de la concepción (POC). Se puede usar para análisis de células individuales, como el genotipado en muestras biopsiadas de embriones. Se puede usar para el análisis rápido de embriones (en menos de uno, uno o dos días de la biopsia) mediante secuenciación dirigida mediante mini-PCR.

**[0098]** PCR altamente multiplexados a menudo pueden resultar en la producción de una muy alta proporción de ADN producto resultante de reacciones secundarias no productivas, tales como la formación de dímeros del cebador. En una realización, los cebadores particulares que tienen más probabilidades de causar reacciones secundarias improductivas pueden eliminarse de la biblioteca de cebadores para dar una biblioteca de cebadores que dará como resultado una mayor proporción de ADN amplificado que se mapea en el genoma. El paso de eliminar los cebadores problemáticos, es decir, aquellos cebadores que son particularmente propensos a reafirmar los dímeros, inesperadamente ha permitido niveles de multiplexación de PCR extremadamente altos para su posterior análisis por secuenciación. En sistemas como la secuenciación, donde el rendimiento se degrada significativamente por los

dímeros de cebador y/u otros productos dañinos, se ha logrado una multiplexación mayor que 10, mayor que 100 y mayor que 100 veces mayor que otras multiplexiones descritas. Tenga en cuenta que esto se opone a los métodos de detección basados en sondas, por ejemplo, microarrays, TaqMan, PCR, etc., donde un exceso de dímeros de cebadores no afectará el resultado de manera apreciable. También tenga en cuenta que la creencia general en la técnica es que la PCR de multiplexación para la secuenciación se limita a aproximadamente 100 ensayos en el mismo pozo.

**[0099]** Hay un número de maneras para elegir cebadores para una biblioteca en donde la cantidad de cebadores-dímero de no-mapeo se minimizan u otros productos cebadores de travesura. Los datos empíricos indican que un pequeño número de cebadores "malos" son responsables de una gran cantidad de reacciones secundarias de cebadores-dímero sin mapeo. La eliminación de estos cebadores "malos" puede aumentar el porcentaje de lecturas de secuencia que se asignan a loci específicos. Una forma de identificar los cebadores "malos" es observar los datos de secuenciación del ADN que se amplificaron mediante amplificación dirigida; esos dímeros de cebadores que se ven con mayor frecuencia se pueden eliminar para dar una biblioteca de cebadores que es significativamente menos probable que produzca ADN de producto secundario que no se mapee al genoma. También hay programas disponibles públicamente que pueden calcular la energía de unión de varias combinaciones de cebadores, y eliminar aquellos con la energía de unión más alta también proporcionará una biblioteca de cebadores que es significativamente menos probable que produzca ADN de producto secundario que no se mapee en el genoma.

**[0100]** Tenga en cuenta que existen otros métodos para determinar qué sondas de PCR tienen más probabilidad de formar dímeros. En una realización, el análisis de un conjunto de ADN que se ha amplificado usando un conjunto no optimizado de cebadores puede ser suficiente para determinar cebadores problemáticos. Por ejemplo, el análisis puede hacerse usando secuenciación, y se determina que los dímeros que están presentes en la mayor cantidad son los que tienen más probabilidades de formar dímeros, y se pueden eliminar.

**[0101]** Para seleccionar ubicaciones diana, se puede comenzar con un grupo de supuestos diseños de pares de cebadores y crear un modelo termodinámico de interacciones potencialmente adversas entre pares de cebadores, y luego usar el modelo para eliminar diseños que son incompatibles con otros diseños en el grupo.

**[0102]** Hay muchos flujos de trabajo que son posibles cuando se realiza la PCR; se describen algunos flujos de trabajo típicos de los métodos descritos aquí. Los pasos descritos aquí no pretenden excluir otros posibles pasos ni implica que ninguno de los pasos descritos aquí sean necesarios para que el método funcione correctamente. Se conoce un gran número de variaciones de parámetros u otras modificaciones en la literatura, y se pueden realizar sin afectar la esencia de la divulgación. A continuación se proporciona un flujo de trabajo generalizado particular seguido de una serie de posibles variantes. Las variantes generalmente se refieren a posibles reacciones de PCR secundarias, por ejemplo, diferentes tipos de anidamiento que se pueden realizar (paso 3). Es importante tener en cuenta que las variantes se pueden hacer en diferentes momentos, o en diferentes órdenes que los descritos explícitamente en este documento.

1. El ADN en la muestra puede tener adaptadores de ligadura, a menudo denominados etiquetas de biblioteca o etiquetas de adaptador de ligadura (LT), adjuntas, donde los adaptadores de ligadura contienen una secuencia de cebado universal, seguida de una amplificación universal. Esto se puede hacer usando un protocolo estándar diseñado para crear bibliotecas de secuencia después de la fragmentación. La muestra de ADN puede tener un extremo romo y luego se puede agregar una A en el extremo 3'. Se puede agregar y ligar un adaptador en Y con un saliente en T. Se pueden usar otros extremos adhesivos que no sean un voladizo A o T. Se pueden agregar otros adaptadores, por ejemplo adaptadores de ligadura en bucle. Los adaptadores pueden tener una etiqueta diseñada para la amplificación por PCR.

2. Amplificación específica diana (STA): la preamplificación de cientos a miles a decenas de miles e incluso cientos de miles de dianas pueden multiplexarse en una reacción. El STA normalmente se ejecuta de 10 a 30 ciclos, aunque puede ejecutarse de 5 a 40 ciclos, de 2 a 50 ciclos e incluso de 1 a 100 ciclos. Los cebadores se pueden seguir, por ejemplo, para un flujo de trabajo más simple o para evitar la secuenciación de una gran proporción de dímeros. Tenga en cuenta que, por lo general, los dímeros de ambos cebadores que llevan la misma etiqueta no se amplificarán ni secuenciarán de manera eficiente. Se pueden llevar a cabo entre 1 y 10 ciclos de PCR; se pueden llevar a cabo entre 10 y 20 ciclos de PCR; se pueden llevar a cabo entre 20 y 30 ciclos de PCR; se pueden llevar a cabo entre 30 y 40 ciclos de PCR; se pueden realizar más de 40 ciclos de PCR. La amplificación puede ser una amplificación lineal. El número de ciclos de PCR puede optimizarse para obtener un perfil de profundidad de lectura (DOR) óptimo. Diferentes perfiles de DOR pueden ser deseables para diferentes propósitos. Es deseable una distribución más uniforme de las lecturas entre todos los ensayos; si el DOR es demasiado pequeño para algunos ensayos, el ruido estocástico puede ser demasiado alto para que los datos sean demasiado útiles, mientras que si la profundidad de lectura es demasiado alta, la utilidad marginal de cada lectura adicional es relativamente pequeña.

**[0103]** Las colas de cebador pueden mejorar la detección de ADN fragmentado de bibliotecas etiquetadas universalmente. Si la etiqueta de la biblioteca y las colas del cebador contienen una secuencia homóloga, se puede mejorar la hibridación (por ejemplo, la temperatura de fusión ( $T_m$ ) se reduce) y los cebadores se pueden extender si solo una parte de la secuencia diana del cebador está en el fragmento de ADN de muestra. Se pueden usar 13 o más

pares de bases específicas diana. Se pueden utilizar de 10 a 12 pares de bases específicas diana. Se pueden utilizar de 8 a 9 pares de bases específicas diana. Se pueden utilizar de 6 a 7 pares de bases específicas diana. STA puede realizarse en ADN preamplificado, por ejemplo, MDA, RCA, otras amplificaciones de genoma completo o PCR universal mediada por adaptador. El STA se puede realizar en muestras enriquecidas o agotadas de ciertas secuencias y poblaciones, por ejemplo, por selección de tamaño, captura de dianas, degradación dirigida.

3. Puede ser posible realizar PCR multiplex secundarias o reacciones de extensión del cebador para aumentar la especificidad y reducir los productos indeseables. Por ejemplo, la anidación completa, semi-anidación, hemi-anidación y/o subdivisión en reacciones paralelas de grupos de análisis más pequeños son todas técnicas que pueden usarse para aumentar la especificidad. Los experimentos han demostrado que dividir una muestra en tres reacciones de 400 plex dio como resultado un producto de ADN con mayor especificidad que una reacción de 1.200 plex con exactamente los mismos cebadores. De manera similar, los experimentos han demostrado que dividir una muestra en cuatro reacciones de 2.400-plex dio como resultado un producto de ADN con mayor especificidad que una reacción de 9.600-plex con exactamente los mismos cebadores. Puede ser posible usar cebadores específicos de dianas y etiquetas específicas de la misma y opuesta direccionalidad.

4. Puede ser posible amplificar una muestra de ADN (dilución, purificada o no) producida por una reacción STA usando cebadores específicos de etiqueta y "amplificación universal", es decir, para amplificar muchas o todas las dianas preamplificadas y etiquetadas. 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.

**[0104]** Estos métodos pueden usarse para el análisis de cualquier muestra de ADN, y son especialmente útiles cuando la muestra de ADN es particularmente pequeña, o cuando es una muestra de ADN donde el ADN se origina en más de un individuo, como en el caso del plasma materno. Estos métodos se pueden usar en muestras de ADN, como un número único o pequeño de células, ADN genómico, ADN plasmático, bibliotecas de plasma amplificado, bibliotecas de sobrenadante apoptótico amplificado u otras muestras de ADN mixto. Estos métodos pueden usarse en el caso en que células de diferente constitución genética puedan estar presentes en un solo individuo, como con cáncer o trasplantes.

**[0105]** La amplificación por PCR multiplex puede implicar el uso de varios tipos de protocolos de anidación, por ejemplo: mini-PCR semi-anidada, mini-PCR completamente anidada, mini-PCR heminada, mini-PCR triplemente anidada, PCR mini-anidada unilateral, mini-PCR unilateral o mini-PCR semianidada inversa.

#### *Cuadro de diagnóstico*

**[0106]** La presente descripción puede comprender un cuadro de diagnóstico que es capaz de llevar a cabo parcial o completamente aspectos de los métodos descritos en esta descripción. El cuadro de diagnóstico puede ubicarse en el consultorio de un médico, en el laboratorio de un hospital o en cualquier lugar adecuado razonablemente próximo al punto de atención del paciente. El cuadro puede ejecutar los aspectos del método de una manera totalmente automatizada, o puede requerir que un técnico complete uno o varios pasos manualmente. En una realización, el cuadro puede poder analizar los datos genotípicos medidos en el plasma materno. En una realización, el cuadro puede estar vinculado a medios para transmitir los datos genotípicos medidos usando el cuadro de diagnóstico a una instalación de cómputo externa que luego puede analizar los datos genotípicos y posiblemente también generar un informe. El cuadro de diagnóstico puede incluir una unidad robótica que sea capaz de transferir muestras acuosas o líquidas de un recipiente a otro. Puede comprender varios reactivos, tanto sólidos como líquidos. Puede comprender un secuenciador de alto rendimiento. Puede comprender una computadora.

#### *Kit de cebadores*

**[0107]** Se puede formular un kit que comprenda una pluralidad de cebadores diseñados para lograr los métodos descritos en esta descripción. Los cebadores pueden ser cebadores externos hacia adelante e inversos, cebadores internos hacia adelante e inversos como se divulga aquí; podrían ser cebadores que han sido diseñados para tener baja afinidad de unión a otros cebadores en el kit como se describe en la sección sobre diseño de cebadores; podrían ser sondas de captura híbridas o sondas pre-circulares como se describe en las secciones relevantes, o alguna combinación de las mismas. En una realización, se puede formular un kit para determinar el estado de ploidía de un cromosoma diana en un feto gestante diseñado para usarse con los métodos descritos en este documento, comprendiendo el kit una pluralidad de cebadores internos hacia adelante y opcionalmente la pluralidad de cebadores inversos internos, y opcionalmente cebadores directos externos y cebadores inversos externos, donde cada uno de los cebadores está diseñado para hibridarse con la región de ADN inmediatamente aguas arriba y/o aguas abajo de uno de los sitios polimórficos en el cromosoma diana, y opcionalmente cromosomas adicionales. El kit de cebador puede usarse en combinación con el cuadro de diagnóstico descrito en otra parte de este documento.

#### *Estimaciones de máxima verosimilitud*

**[0108]** Muchos métodos conocidos en la técnica para detectar la presencia o ausencia de un fenotipo o genotipo, por



ejemplo, una anomalía cromosómica, una afección médica o una relación de paternidad implican el uso de una prueba de rechazo de hipótesis única, donde se mide la métrica que está directamente relacionada o correlacionada con la condición, y si la métrica está en un lado de un umbral dado, se determina que la condición está presente, mientras que si la métrica cae en el otro lado del umbral, la condición está decidida a estar ausente. Una prueba de rechazo de hipótesis única solo analiza la distribución nula al decidir entre las hipótesis nula y alternativa. Sin tener en cuenta la distribución alternativa, no se puede estimar la probabilidad de cada hipótesis dados los datos observados y, por lo tanto, no se puede calcular una confianza en la llamada. Por lo tanto, con una prueba de rechazo de hipótesis única, se obtiene una respuesta de sí o no sin una estimación de la confianza asociada con el caso específico.

**[0109]** El método descrito en este documento puede ser capaz de detectar la presencia o ausencia del fenotipo o genotipo, por ejemplo, una anomalía cromosómica, una condición médica o una relación de paternidad, utilizando un método de máxima verosimilitud. Esta es una mejora sustancial sobre un método que utiliza una técnica de rechazo de hipótesis única, ya que el umbral para llamar ausencia o presencia de la condición se puede ajustar según corresponda para cada caso. Esto es particularmente relevante para las técnicas de diagnóstico que tienen como objetivo determinar la paternidad de un feto en gestación a partir de los datos genéticos disponibles de la mezcla de ADN fetal y materno presente en el libre ADN flotante encontrado en plasma materno. El método de estimación de máxima verosimilitud puede usar las distribuciones alélicas asociadas con cada hipótesis para estimar la probabilidad de los datos condicionados en cada hipótesis. Estas probabilidades condicionales se pueden convertir a una llamada de hipótesis y confianza. De manera similar, el método de estimación de máximo a posteriori usa las mismas probabilidades condicionales que la estimación de máxima verosimilitud, pero también incorpora los antecedentes de la población al elegir la mejor hipótesis y determinar la confianza.

**[0110]** Por lo tanto, el uso de una técnica de estimación de máxima verosimilitud (MLE), o la técnica de máximo a posteriori (MAP) estrechamente relacionada da dos ventajas, primero aumenta la posibilidad de una llamada correcta, y también permite que una confianza sea calculada para cada llamada. En una realización, la selección de la llamada de paternidad correspondiente a la hipótesis con la mayor probabilidad se lleva a cabo utilizando estimaciones de probabilidad máxima o estimaciones a posteriori máximas. Un método se describe para la determinación de la paternidad de un feto en gestación que consiste en tomar cualquier método actualmente conocido en la técnica que utiliza una única técnica de rechazo de hipótesis y reformulándola tal que utiliza una técnica MLE o MAP.

#### *Un método para la determinación de la paternidad*

**[0111]** El ADNff está típicamente presente en una fracción baja en una mezcla con ADN materno. En algunas realizaciones, la madre tiene un genotipo conocido o se puede medir o inferir el genotipo materno. Típicamente, la fracción de ADN fetal que se encuentra en el plasma materno está entre 2 y 20%, aunque en diferentes condiciones este porcentaje puede variar de aproximadamente 0,01% a aproximadamente 50%. Se puede utilizar una micromatriz u otra técnica que proporcione datos de intensidad por alelo para medir el plasma materno. La secuenciación se puede utilizar para medir el ADN contenido en el plasma materno. En estos casos, la medición de la intensidad del alelo o el recuento de lectura de secuencia en un alelo particular es una suma de las señales maternas y fetales. Suponiendo que la proporción de la mezcla del ADN del niño con la madre es  $r$  a 1, el número relativo de alelos en un locus consiste en 2 alelos de la madre y 2 alelos del niño. Los loci pueden comprender polimorfismos de un solo nucleótido. La **Tabla 1** muestra el número relativo de cada alelo en la mezcla para una selección de contextos parentales informativos.

**Tabla 1:** Número de alelos por contexto

Contexto pariente	A en mezcla	B en mezcla
AA AA	$2 + 2r$	0
AA BB	$2 + r$	R
AB AA	$1 + 2r$	1
AA AB	$2+2r$ o $2 + r$	0 o $r$
BB AA	$r$	$2 + r$
BB BB	0	$2 + 2r$

**[0112]** Tenga en cuenta que la elección de los cuatro contextos anteriores como informativos no pretende incluir todos los contextos que pueden ser informativos. Se puede usar cualquier combinación de contextos, y hay una cantidad significativa de información que se puede encontrar en las mediciones genómicas de cualquier contexto.

**[0113]** Incluso en presencia de una tasa de abandono de alelo significativa del niño, puede haber una distinción clara entre la señal donde está presente un alelo y la señal donde no está presente un alelo. Por ejemplo, considere las mediciones del alelo A de los SNP en los pares de contexto BB|BB y BB|AA y las mediciones del alelo B de los SNP en los pares de contexto AA|AA y AA|BB. En cada caso, no debe haber señal presente en el primer contexto y debe haber señal presente en el segundo contexto, donde los alelos del niño no se hayan retirado. Sin embargo, si el supuesto padre es incorrecto, a veces habrá señal presente en ambos contextos. Por lo tanto, la distribución de las mediciones de SNP debe ser diferente, dependiendo de si el supuesto padre es correcto o no.

[0114] La diferencia típicamente será más observable en el extremo de alta señal de la distribución, porque estos serán los SNP donde hay una mayor probabilidad de tener contribuciones de ADN del niño. Esta diferencia se puede observar comparando valores de percentiles altos de las distribuciones de las mediciones de SNP. Ejemplos de posibles métodos de percentiles son el rango más cercano, la interpolación lineal entre los rangos más cercanos y el percentil ponderado.

[0115] Por ejemplo, definir  $X_1$  como el conjunto de medidas de SNP del alelo A en el contexto BB|BB y  $X_2$  como el conjunto de medidas de alelo A en el contexto BB|AA, de todos los cromosomas. Si el supuesto padre es correcto, entonces el 99° valor percentil de  $X_1$  será significativamente menor que el 99° valor percentil de  $X_2$ . Si el supuesto padre es incorrecto, el 99° valores percentil de las dos distribuciones estará más cerca. Nota que cualquier percentil se puede utilizar igualmente bien, por ejemplo, 95° percentil, 90° percentil, 85° percentil o 80° percentil. En una realización, para un canal de medición particular,  $X_1$  puede definirse como las mediciones del contexto sin señal y  $X_2$  puede definirse como las mediciones del contexto donde la madre y el padre son homocigotos, y solo los alelos del padre proporcionan una señal (inherente a través del niño).

[0116] Definir  $p_1$  como el 99° (o 95°, 90° etc.) percentil de los datos  $X_1$  y  $p_2$  como el 99° (o 95°, 90°, etc.) percentil de los datos  $X_2$ . Definir la estadística de prueba  $t$  como  $p_1/p_2$ . Tenga en cuenta que otras funciones de  $p_1$  y  $p_2$  que demuestran la diferencia en los valores pueden usarse igualmente bien. El valor de  $t$  variará dependiendo de la cantidad de ADN infantil presente en la muestra, que no se conoce. Por lo tanto, los umbrales de clasificación para  $t$  no pueden calcularse a priori.

[0117] La estadística de prueba  $t$  para una sola muestra puede compararse con una distribución generada a partir de los genotipos de muchos individuos que se sabe que no son el padre, utilizando el siguiente procedimiento. Suponga que los genotipos de un gran conjunto de individuos no relacionados están disponibles.

1. Para cada individuo no relacionado, suponga que es el padre y calcule el valor de la estadística de prueba  $t$ .
2. Que  $T_u$  sea el conjunto de medidas  $t$  de los machos no relacionados. Ajuste una distribución a  $T_u$ . Esta es la distribución de  $t$  para la muestra particular, bajo la hipótesis nula. La hipótesis nula es "los genotipos no provienen del padre del niño presente en la muestra". La distribución  $P_u(t)$  podría ser un ajuste de máxima probabilidad o un método de momentos ajustados a una distribución conocida, por ejemplo, una distribución gaussiana, un ajuste de densidad kernel utilizando una función del núcleo, por ejemplo, gaussiano, cuadrado, triángulo, etc., o cualquier otro método de ajuste de distribución apropiado.
3. Considerar los genotipos del supuesto padre y calcular la estadística de prueba correspondiente  $t_c$ .
4. Se espera que el verdadero padre resulte en un valor menor de  $t$  que un individuo no relacionado. La probabilidad de que un padre produzca  $t_c$  no relacionado o un valor más periférico es la función de densidad acumulada de  $P_u$  evaluada en  $t_c$ . Por lo tanto, el valor  $p$  para rechazar la hipótesis nula viene dado por lo siguiente:

$$p = \int_0^{t_c} P_u(t) dt$$

[0118] Si  $p$  cae por debajo de un umbral de significancia  $\alpha$ , entonces la hipótesis de que el padre alega que es un individuo no relacionado puede rechazarse con significancia  $\alpha$ . Si  $p$  es mayor que  $\alpha$ , entonces la hipótesis nula no puede ser rechazada, y el supuesto padre puede no estar relacionado con el niño presente en la muestra. El umbral de significancia  $\alpha$  define la probabilidad de que un individuo no relacionado pueda clasificarse como el padre correcto. Por ejemplo, con un umbral de  $\alpha$  igual a 0,01, el uno por ciento de los individuos no relacionados podría identificarse como el padre correcto.

[0119] Varios métodos pueden ser considerados para la combinación de datos de los canales de alelo A y B. Por ejemplo, dos métodos simples son exigir que el valor  $p$  de todos los canales esté por debajo de un umbral, o exigir que el valor  $p$  de cualquier canal esté por debajo del umbral.

[0120] El método de prueba de paternidad puede suponer que el ADN del niño está presente a una concentración suficiente para distinguir entre los SNP que tienen o no señal del niño. En ausencia de una concentración de ADN infantil suficiente, este método puede informar "padre incorrecto" porque los alelos heredados paternalmente esperados no se miden en el plasma materno. Se describe un método que puede confirmar la presencia de ADN infantil suficiente antes de aplicar la prueba de paternidad. La confirmación de presencia infantil se basa en el cálculo de una estadística de prueba que es proporcional a la concentración de ADN infantil, pero que no requiere genotipos del padre. Si la estadística de la prueba está por encima del umbral requerido, entonces la concentración de ADN infantil es suficiente para realizar la prueba de paternidad.

[0121] Considerar el conjunto de SNP (de todos los cromosomas combinados) donde el genotipo madre es AA y se mide el canal B. Se espera una señal solo en el subconjunto de SNP donde el genotipo secundario contiene una B, pero estos SNP no pueden identificarse a priori sin los genotipos padre, que no están disponibles. En cambio, considerar las frecuencias de las poblaciones de SNP  $\{f_i\}$  donde  $f_i$  es el número medio de muestra de Bs en el genotipo

de SNP  $i$ , basado en una gran muestra de población. Tenga en cuenta que la mayoría de los SNP donde el genotipo madre es AA tendrá  $f_i$  menor que 0,5, pero la distribución de  $f_i$  en estos SNP se extiende casi a uno. Considere dos conjuntos de SNP,  $S_1$  y  $S_2$ , donde  $S_1 = \{i: f_i < T_L\}$  y  $S_2 = \{i: f_i > T_H\}$ . Los umbrales  $T_L$  y  $T_H$  se establecen de modo que se espera que muy pocos SNP en  $S_1$  tengan una B y muchos SNP en  $S_2$  tengan una B, y cada conjunto tiene una población suficiente. El algoritmo puede usar  $T_L = 0,05$  y  $T_H = 0,7$ , mientras que otros valores para  $T_L$  y  $T_H$  podrían funcionar igual o mejor. Que  $y_i$  sea la medida del canal B de SNP  $i$ ,  $Y_1 = \{y_i: i \in S_1\}$  e  $Y_2 = \{y_i: i \in S_2\}$ . Las distribuciones de  $Y_1$  e  $Y_2$  serán muy similares porque la mayoría de los SNP en ambas distribuciones no tendrán señal. Sin embargo, se espera que algunos números no triviales de SNP en  $S_2$  tengan señal de niños y se espera que muy pocos SNPs en  $S_1$  tengan señal de niño. Por lo tanto, la cola de  $Y_2$  debe extenderse a una intensidad mayor que la cola de  $Y_1$ . Que  $p_i$  sea un percentil cercano a 1, por ejemplo, el percentil 99. En presencia de ADN infantil suficiente, el percentil  $p_i$  de  $Y_2$  debería ser significativamente más alto que el percentil  $p_i$  de  $Y_1$ . Por lo tanto, las estadísticas de prueba se pueden definir de la siguiente manera.

$$S = \text{percentil}(Y_2; p_i) - \text{percentil}(Y_1; p_i)$$

[0122] La estadística de prueba puede normalizarse mediante una variedad de métodos para intentar explicar las diferencias de amplificación entre las matrices. La normalización podría hacerse por cromosoma. La normalización se podría hacer por arreglo. La normalización se puede hacer por secuencia de ejecución.

[0123] El siguiente cálculo muestra cómo los umbrales  $T_L$  y  $T_H$  pueden distinguir el efecto del ADN infantil en una muestra materna particular, en base a números aproximados de SNP y tasas de deserción. La **Tabla 3** muestra algunos datos de

[0124] El método puede implicar los siguientes supuestos:

- Las frecuencias de población se calculan a partir de un conjunto de datos de población grande, por ejemplo, más de 500 individuos, más de 1.000 individuos, más de 5.000 individuos o más de 20.000 individuos, y el número de SNP en cada contexto proviene de un ejemplo de madre y padre.
- No hay SNPs donde la madre es AA y el padre es BB (en realidad, estos son aproximadamente el 8 por ciento de los SNPs de la madre AA)
- La mitad de los SNPs donde el padre tiene B dan como resultado el niño B.
- La tasa de abandono del niño es del 90 por ciento.
- Las mediciones con señal secundaria serán más altas que las mediciones sin señal secundaria

[0125] La **Tabla 3** muestra algunos datos de un caso de paternidad particular usando el método divulgado con los parámetros anteriores. No se espera que la medición del percentil 98 de  $S_1$  incluya ningún SNP con señal secundaria presente. Se espera que la medición del percentil 98 de  $S_2$  incluya alrededor de 50 SNP con señal secundaria presente. La diferencia entre los dos debe reflejar la cantidad de señal secundaria.

**Tabla 3:** Datos relativos a la determinación de paternidad

conjunto	definición	núm SNPs en conjunto	$f_i$ promedio en conjunto	núm SNPs padre B	núm SNPs señal niño	fracción de conjunto, señal de niño
$S_1$	$f_i < 0,05$	13300	0,012	171	9	0,0007
$S_2$	$f_i > 0,7$	3000	0,79	2370	119	0,039

[0126] La **Figura 1** muestra la distribución de datos de intensidad de alelos para contextos AA|AA y AA|BB de una muestra de plasma materno recogida a las 38 semanas. El alelo B se mide. Tenga en cuenta que la distribución AA|BB se extiende significativamente más alta que la distribución AA|AA, lo que muestra que el alelo B (que solo está presente en el genoma del niño) está presente en el contexto AA|BB pero no en el contexto AA|AA. La **Figura 2** proviene de la misma muestra de plasma materno que la **Figura 1**, y muestra la distribución de la estadística de prueba  $t$  para el alelo B usando los genotipos de 200 individuos no relacionados. Se muestran dos distribuciones (las dos curvas): el ajuste Gaussiano de máxima probabilidad y una distribución del núcleo. El valor  $t_c$  para el padre biológico está marcado con una estrella. El valor  $p$  es inferior a  $10^{-7}$  para la hipótesis nula de que el supuesto padre no está relacionado con el niño.

[0127] La **Tabla 4** presenta los resultados de 8 muestras de sangre materna en diferentes etapas del embarazo. Se calcula un valor  $p$  basado en los datos medidos de cada canal (alelo A y alelo B) para el padre correcto. Si se requiere que los valores  $p$  de ambos canales estén por debajo de 0,01, entonces dos muestras se clasifican incorrectamente.

Si solo se requiere un canal para pasar el umbral, entonces todas las muestras se clasifican correctamente. Se puede usar cualquier cantidad de métricas y umbrales para confirmar o excluir la paternidad. Por ejemplo, uno podría usar un valor p de corte de 0,02, 0,005, 0,001 o 0,00001; de manera similar, uno podría exigir que uno o ambos valores p del canal estén por debajo de un umbral dado, o uno podría tener dos umbrales diferentes para los diferentes valores p del canal.

**Tabla 4:** Valores p para dos canales para ocho determinaciones de paternidad.

	Semanas embarazo	Valor p (Y)	Valor p (X)
10	11	$2,3 \times 10^{-7}$	$<10^{-7}$
	16	0,013	$<10^{-7}$
	17	$<10^{-7}$	$<10^{-7}$
	17	$<10^{-7}$	0,0002
	20	$<10^{-7}$	$<10^{-7}$
15	28	0,14	0,0048
	38	$<10^{-7}$	$<10^{-7}$
	38	$<10^{-7}$	$<10^{-7}$

**[0128]** La **Tabla 4** muestra los valores P para la hipótesis nula de que el padre correcto es un individuo no relacionado. Cada fila corresponde a una muestra de sangre materna diferente, y la muestra genética paterna correspondiente. Las mediciones genéticas realizadas en 200 machos no relacionados se utilizaron como control. La curva en la **Figura 3** muestra la distribución de las relaciones de intensidad para 200 hombres no relacionados, y la estrella representa la relación de intensidad para el padre biológico. Estos datos se toman de un caso en donde se extrajo sangre de una madre que estaba embarazada de 11 semanas.

**[0129]** La **Figura 4** muestra las curvas de frecuencia de distribución acumulativa (cdf) para la relación de correlación entre las mediciones genotípicas fetales y las mediciones genotípicas parentales para tres casos: (1) donde tanto la madre embarazada como el supuesto padre son los padres biológicos del feto ("correcto", curva más a la derecha), (2) donde la madre embarazada es la madre biológica del feto, pero el supuesto padre no es el padre biológico del feto ("uno equivocado", curva media) y (3) donde ni la madre embarazada ni el supuesto padre son los padres biológicos del feto ("dos equivocados", curva más a la izquierda). Las curvas de cdf son la relación de correlación entre los datos genotípicos del embrión, calculados a partir de los datos medidos en una sola célula, y los datos genotípicos de los padres asumidos cuando cero, uno o dos de los padres asumidos son en realidad los padres genéticos del feto. Tenga en cuenta que las etiquetas para "correcto" y "dos incorrectos" se invierten. La **Figura 5** muestra histogramas para los mismos tres casos. Tenga en cuenta que este histograma está compuesto por más de 1000 casos en los que uno o ambos padres son incorrectos. El histograma de la tasa de correlación medida entre los datos genotípicos del feto, medidos en una sola célula, y los datos genotípicos de los padres asumidos cuando cero, uno o dos de los padres asumidos son en realidad los padres genéticos del feto.

**[0130]** Treinta y cinco resultados de paternidad se muestran en la **Figura 6** usando el método instantáneo para pruebas de paternidad. Se analizaron con muestras recolectadas de mujeres embarazadas con edades gestacionales que van de 9 a 40 semanas. La curva roja a la derecha representa una distribución gaussiana normalizada de la estadística de prueba de paternidad para 800 hombres no relacionados. La distribución de los machos no relacionados es diferente para cada caso; Aquí se utiliza una distribución normalizada para fines de visualización.

**[0131]** Las barras azules representan la estadística de prueba normalizada para el (presunto) padre correcto. Está claro que los padres correctos están claramente separados de los machos no relacionados. Tenga en cuenta que la estadística de prueba normalizada se aproxima crudamente a las desviaciones estándar, por lo tanto, "-5" en el gráfico a continuación es aproximadamente 5 desviaciones estándar de la media. Por lo tanto, todos los padres correctos asumidos en esta cohorte han sido confirmados como los padres correctos con un significado de al menos 99,9999%.

**[0132]** El conocimiento de los haplotipos de los padres podrían usarse para aumentar la exactitud de la prueba. Por ejemplo, si los dos haplotipos del padre son conocidos para un segmento dado de un cromosoma, entonces el conocimiento de qué SNP están presentes para los casos en que no hay abandono se puede utilizar para determinar qué SNP deben esperarse para aquellos casos en los que puede haber deserción. Por ejemplo, imagine un conjunto de tres SNP que están vinculados, es decir, están ubicados muy juntos en el mismo cromosoma y donde los contextos de la madre y el supuesto padre son: AA|AB, AA|AB, AA|BA. Tenga en cuenta que cuando el genotipo de un padre está en fases, entonces  $AB \neq BA$ , ya que la primera de las dos letras representa los alelos en el primer haplotipo, y la segunda letra representa los alelos en el segundo haplotipo. Ahora imagine que para esos tres SNP, se mide un nivel significativo del alelo B para los tres; en este caso, la posibilidad de que el presunto padre sea el padre correcto es baja, porque los dos haplotipos del padre son A, A, B y B, B, A, mientras que el genotipo fetal medido es positivo para B en los tres SNP, y la madre solo pudo haber contribuido con una A. Si el genotipo del padre no fue escalonado, no sería posible descartar a este supuesto padre dado este conjunto de medidas. Esa determinación de los haplotipos del padre puede determinarse dadas las mediciones de ADN genómico diploide junto con las mediciones genéticas haploides realizadas en uno o más espermatozoides. El uso de más de un espermatozoide puede permitir la determinación de los haplotipos con mayor precisión, así como la cantidad de cruces que pueden haber ocurrido, para

cada uno de los cromosomas, junto con su ubicación, durante la meiosis que formó el esperma. Un método para lograr esta fase paterna se puede encontrar con mayor detalle en las cuatro solicitudes de patente de Rabinowitz a las que se hace referencia en este documento.

[0133] La determinación de paternidad se puede hacer exclusivamente usando mediciones SNP, y no se usan datos de repeticiones en tándem individuales. La determinación de la paternidad se puede hacer exclusivamente utilizando mediciones SNP y STR. Los datos de SNP pueden medirse utilizando microarrays de SNP, o pueden medirse por secuenciación. La secuenciación puede no ser dirigida, o puede ser dirigida, por ejemplo, mediante el uso de sondas de circularización que están dirigidas a un conjunto de loci polimórficos, o puede ser utilizada mediante el uso de captura mediante técnicas de hibridación. Los datos genéticos pueden medirse mediante una combinación de métodos; por ejemplo, los datos genéticos de los padres pueden medirse en un microarray SNP, mientras que el ADN aislado del suero materno puede medirse usando secuenciación dirigida donde las sondas de hibridación de captura se usan para apuntar a los mismos SNP que se encuentran en el microarray SNP. Se puede usar una combinación de los siguientes tipos de datos para determinar si el supuesto padre es el padre biológico del feto: datos SNP, datos STR, datos cruzados, datos de microdelección, datos de inserción, datos de translocación u otros datos genéticos.

[0134] El método puede comprender la generación de un informe que divulga la paternidad establecida del feto u otro individuo diana. El informe puede generarse con el propósito de comunicar la determinación de paternidad. El informe puede comprender una probabilidad de que el supuesto padre sea el padre biológico del feto. Algunos ejemplos de dicho informe se muestran dentro; La Figura 7 es un ejemplo de un informe que revela una exclusión de paternidad, la Figura 8 es un ejemplo de un informe que revela una inclusión de paternidad y la Figura 9 es un ejemplo de un informe que indica un resultado indeterminado. El informe puede comprender un gráfico que contiene una distribución de una métrica relacionada con la paternidad para una pluralidad de individuos no relacionados con respecto a un feto y madre dados (mostrados en gris curva), y una indicación de la métrica para el supuesto padre (se muestra como un triángulo). La distribución de los machos no relacionados es diferente para cada caso; en estos tres informes, aquí se utiliza una distribución real de la estadística de prueba para el feto y los machos no relacionados. El informe también puede contener una indicación de que el presunto padre es más probable que forme parte de la distribución de individuos no relacionados (por ejemplo, la Figura 7) y, por lo tanto, se establece que el presunto padre no es el padre biológico del feto; el hecho de que el triángulo esté en la región de exclusión de paternidad del gráfico indica que se trata de una exclusión de paternidad. El informe también puede contener una indicación de que es más probable que el presunto padre no sea parte de la distribución de la métrica de paternidad para personas no relacionadas (por ejemplo, la Figura 8), y que el presunto padre es el padre biológico del feto; el hecho de que el triángulo esté en la región de inclusión de paternidad del gráfico indica que se trata de una inclusión de paternidad. El informe también puede contener una indicación de que las mediciones son indeterminadas (por ejemplo, la Figura 9); El hecho de que el triángulo esté en la región de "resultado indeterminado" del gráfico indica que no se llegó a ninguna conclusión con respecto al establecimiento de la paternidad del feto.

[0135] La determinación de si el presunto padre es el padre biológico del feto se realiza sin el uso de repeticiones en tándem individuales (ROS). La precisión de la determinación de la paternidad se puede aumentar mediante la eliminación gradual de los genotipos de los padres. Los genotipos de uno o más de los padres pueden ser escalonados con el uso de material genético de uno o más individuos relacionados con ese padre. El individuo relacionado con el padre puede ser los padres padre, madre, hermano, hijo, hija, hermano, hermana, tía, tío, gemelo, clon, un gameto del padre y combinaciones de los mismos.

#### Otro método para la determinación de la paternidad

[0136] El plasma materno y, opcionalmente, el otro material genético se pueden medir mediante secuenciación, por ejemplo, utilizando secuenciadores de alto rendimiento como HISEQ o MISEQ de ILLUMINA, o el ION TORRENT de LIFE TECHNOLOGIES.

[0137] Las pruebas de paternidad no invasivas pueden realizarse en una muestra de sangre materna si hay una concentración suficiente de ADN fetal flotante. En general, la fracción de ADN fetal en la mayoría de los casos, el plasma materno estará entre el 2 y el 20 por ciento, aunque puede ser tan bajo como 0,01% o tan alto como 40%, dependiendo en parte de la edad gestacional. Ya se ha demostrado que este rango de fracción fetal es suficiente para la prueba de paternidad mediante un método de rechazo de hipótesis única que utiliza microarrays SNP. La secuenciación de alto rendimiento es una plataforma mucho más precisa que permite el modelado matemático de la respuesta de medición esperada en cada SNP, para combinaciones de genotipos de madre e hijo. El plasma materno y, opcionalmente, el otro material genético se puede medir mediante secuenciación, por ejemplo, utilizando secuenciadores de alto rendimiento como HISEQ o MISEQ de ILLUMINA, o ION TORRENT de LIFE TECHNOLOGIES. La confianza en las inclusiones o exclusiones de paternidad se puede calcular utilizando teorías de probabilidad y/o estimación.

[0138] El método para las pruebas de paternidad puede incluir lo siguiente. Para un supuesto padre, se puede calcular la probabilidad de los datos de secuenciación, derivados del plasma, con respecto a las dos hipótesis diferentes: (1) el supuesto padre es el padre (biológico) correcto (Hc) y (2) el supuesto padre no es el padre (biológico) correcto (Hw). La hipótesis de que tiene la probabilidad más alta o se elige entonces a posteriori. Este enfoque se puede combinar

con un modelo de plataforma que relaciona la relación de alelos en el plasma con el número observado de alelos secuenciados A y B. Con el modelo de plataforma disponible, es posible obtener probabilidades probabilísticas de los alelos secuenciados A y B para cada ubicación de SNP para cada hipótesis.

[0139] Una complicación es que la cantidad de fracción fetal en el plasma materno puede variar entre individuos y con el tiempo. El método puede explicar esta variabilidad. Hay varias formas de abordar este tipo de variabilidad. El método puede estimar explícitamente la fracción fetal; El método puede poner a priori en la cantidad desconocida e integra sobre todos los valores posibles. El método puede usar un previo que es como una distribución uniforme de 0 a algún umbral, por ejemplo, 40%. Cualquier previo puede funcionar en teoría. Las probabilidades de varias fracciones secundarias pueden calcularse, ya sea en un espacio continuo o en una partición finita e integrarse o sumar en el rango, respectivamente.

[0140] Considere el plasma materno con fracción fetal,  $f$ , y un único SNP donde la relación alélica esperada presente en el plasma es  $r$  (basada en los genotipos materno y fetal). La proporción de alelos esperada puede definirse como la fracción esperada de alelos A en el ADN combinado materno y fetal. Para genotipo materno  $g_m$  y el genotipo de niño  $g_c$ , la relación esperada de alelo está dada por la ecuación 1, suponiendo que los genotipos se representan como proporciones de alelo también.

$$r = fg_c + (1 - f)g_m \quad (1)$$

[0141] La observación en el SNP comprende el número de lecturas mapeadas con cada alelo presente,  $n_a$  y  $n_b$ , que suma a la profundidad de lectura  $d$ . Suponga que se han aplicado medidas de control de calidad a las probabilidades de mapeo de manera que los mapeos y las observaciones de alelos puedan considerarse correctos. Un modelo simple para la probabilidad de observación es una distribución binomial que supone que cada una de las lecturas  $d$  se extrae independientemente de un gran grupo que tiene una relación de alelo  $r$ . La ecuación 2 describe este modelo.

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

[0142] Cuando los genotipos materno y fetal son todos A o todos B, la relación alélica esperada en plasma será 0 o 1, y  $p_{bino}$  será No estar bien definido. Además, esto no es deseable porque a veces se observan alelos inesperados en la práctica. El modelo binomial se puede ampliar de varias maneras. Es posible utilizar una relación de alelo corregida  $r = 1/(n_a + n_b)$  para permitir que se tenga en cuenta una pequeña cantidad del alelo inesperado. Es posible usar datos de entrenamiento para modelar la tasa de alelos inesperados que aparecen en cada SNP, y usar este modelo para corregir la relación de alelos esperada. Cuando la relación alélica esperada no es 0 o 1, la relación alélica observada puede no converger a la relación alélica esperada debido al sesgo de amplificación u otros fenómenos. La relación de alelos se puede modelar como una distribución beta centrada en la relación de alelos esperada, lo que lleva a una distribución beta-binomial para  $P(n_a, n_b | r)$  que tiene una mayor varianza que el binomio.

[0143] Un modelo de plataforma general para la respuesta en un único SNP puede definirse como  $F(a, b, g_c, g_m, f)$  (3), o la probabilidad de observar  $n_a = a$  y  $n_b = b$  dado los genotipos maternos y fetales, que también depende de la fracción fetal a través de la ecuación 1.

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

[0144] Obsérvese que puede ser factible para simplificar la fórmula (3) por acondicionado en función de  $g_c$ ,  $g_m$  y  $f$  por ejemplo, mediante el uso de  $r$  como se define en (1) y el ejemplo binomial en (2). La ecuación para  $F$  podría entonces escribirse

$$F(a, b, g_c, g_m, f) = P(n_a = a, n_b = b | g_c, g_m, f) = P(n_a = a, n_b = b | r(g_c, g_m, f)) \quad (4)$$

[0145] En general, la forma funcional de  $F$  puede ser una distribución binomial, distribución beta-binomial, distribución multivariada de Pólya, una distribución empírica estimada a partir de datos de entrenamiento o funciones similares como se discutió anteriormente. La forma funcional de  $F$  puede tomar diferentes formas dependiendo de la hipótesis del número de copias en el cromosoma en cuestión.

*Un método para el cálculo de la fracción fetal*

[0146] La determinación de la fracción de ADN fetal que está presente en la fracción mixta de ADN puede ser una

parte integral de cualquier método para la determinación no invasiva de paternidad prenatal, llamada de ploidía o llamada de alelo. La fracción fetal de la muestra mixta se puede determinar utilizando los datos genotípicos de la madre, los datos genotípicos del padre y los datos genotípicos medidos de la muestra mixta que contiene ADN materno y fetal. En el contexto de las pruebas de paternidad, y también en menor medida en el caso de llamadas de ploidía, se desconoce la identidad del padre y, por lo tanto, los datos genotípicos del padre biológico de un feto pueden no estar disponibles. En estos casos, es importante contar con un método para la determinación de la fracción fetal que no requiera el genotipo del padre biológico del feto. Aquí se describen varios métodos para realizar la estimación de la fracción fetal. Estos métodos se describen de manera general, de modo que son apropiados cuando el genotipo del padre biológico está disponible y cuando no lo está.

**[0147]** Para un cromosoma particular, supongamos que estamos considerando  $N$  SNPs, para los cuales tenemos los siguientes datos:

- Un conjunto de mediciones de secuencia de plasma NR  $S = (S_1, \dots, S_{NR})$ . Donde tenemos  $(A, B)$  cuenta para los alelos  $A$  y  $B$  para cada SNP,  $s$  se puede escribir como  $s = ((a_1, b_1), \dots, (a_N, b_N))$ , donde  $a_i$  es el recuento  $a$  en SNP  $i$ ,  $b_i$  es el recuento  $b$  en SNP  $i$ , y  $\sum_{i=1:N} (a_i + b_i) = NR$
- Datos primarios que consisten en:

- Información del genotipo: madre  $G_m = (G_{m1}, \dots, G_{mN})$ , padre  $G_f = (G_{f1}, \dots, G_{fN})$ , donde  $G_{mi}, G_{fi} \in (AA, AB, BB)$ ; y/o
- Mediciones de datos de secuencia: mediciones madre NRM  $s_m = (s_{m1}, \dots, s_{mnr})$ , mediciones padre NRF  $S_f = (S_{f1}, \dots, S_{fnr})$ . Similar a la simplificación anterior, si tenemos  $(A, B)$  cuenta en cada SNP  $S_m = ((a_{m1}, b_{m1}), \dots, (a_{mN}, b_{mN}))$ ,  $S_f = ((a_{f1}, b_{f1}), \dots, (a_{fN}, b_{fN}))$

Colectivamente, los datos de madre, padre e hijo pueden denotarse como  $D = (G_m, G_f, S_m, S_f, S)$ . Los datos genotípicos de ambos padres pueden estar disponibles; datos genotípicos de solo la madre pueden estar disponibles; datos genotípicos de solo el padre pueden estar disponibles; Los datos genotípicos de ninguno de los padres pueden estar disponibles. Los datos genotípicos maternos pueden inferirse de los datos genotípicos medidos en la muestra mixta. Tenga en cuenta que, en general, los datos primarios son deseables y aumentan la precisión del algoritmo, pero no son obligatorios.

**[0148]** La estimación de la fracción secundaria  $f$  es la fracción secundaria esperada dados los datos:

$$\hat{f} = E(cfr|D) = \int f * P(f|D)df$$

**[0149]** Se puede dividir el intervalo de posibles fracciones secundarias en un conjunto  $C$  de puntos finamente espaciados y realizar los cálculos en cada punto que reducen la ecuación anterior a:

$$\hat{f} = E(f|D) = \sum_{f \in C} f * P(f|D)$$

$P(f|D)$  es la probabilidad de una fracción secundaria particular  $f$  dados los datos  $D$ . Se puede deducir aún más usando la regla de Bayes:

$$P(f|D) \sim P(D|f) * P(f)$$

donde  $P(f)$  es el peso previo de una fracción secundaria particular. Esto puede derivarse de información previa no uniforme (uniforme) y puede ser proporcional al espacio entre las fracciones secundarias candidatas en el conjunto  $C$ .

**[0150]**  $P(D|f)$  es la probabilidad de datos dados dada la fracción secundaria particular, derivada bajo los supuestos de número de copia particular sobre los cromosomas relevantes. Podemos suponer una disomía en los cromosomas utilizados. La probabilidad de los datos en todos los SNP es el producto de la probabilidad de datos en los SNP individuales.

$$P(D|f) = P(D|f, H) = \prod_i P(D|f, H, i)$$

Donde  $i$  denota un SNP particular, para SNP  $i$  tenemos:

$$P(D|f, H, i) = \sum_{g_m, g_f, g_c} P(D|g_m, g_f, g_c, f, H, i) * P(g_c|g_m, g_f, H) * P(g_m|i) * P(g_f|i)$$

donde  $g_m$  son posibles genotipos de la madre verdaderos,  $g_f$  son posibles genotipos del padre verdaderos,  $g_c$  son posibles genotipos hijos y  $g_m, g_f, g_c \in \{AA, AB, BB\}$ .

**[0151]**  $P(g_m|i)$  es la probabilidad previa general de genotipo materno  $g_m$  en SNP  $i$ , en base a la frecuencia de población conocida en SNP  $i$ , denotada  $pA_i$ . En particular:

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

lo mismo para  $p(f|i)$ , probabilidad de genotipo del padre.

**[0152]** Que  $P(g_c|g_m, g_f, H)$  denotar es la probabilidad de obtener un verdadero genotipo infantil =  $c$ , dados los padres  $m$ ,  $f$ , y asumiendo la hipótesis  $H$ , que podemos calcular fácilmente. Por ejemplo, para una disomía:

padres		P(c m,f,disomía)		
M	F	AA	AB	BB
AA	AA	1	0	0
AB	AA	0,5	0,5	0
BB	AA	0	1	0
AA	AB	0,5	0,5	0
AB	AB	0,25	0,5	0,25
BB	AB	0	0,5	0,5
AA	BB	0	1	0
AB	BB	0	0,5	0,5
BB	BB	0	0	1

**[0153]** Que  $P(D|g_m, g_f, g_c, h, i, f)$  la probabilidad de datos dada  $D$  en SNP  $i$ , dado el verdadero genotipo  $m$  de la madre, el verdadero genotipo  $f$  del padre, el verdadero genotipo  $c$  del niño, la hipótesis  $H$  para el número de copia y la fracción  $f$  del niño. Se puede desglosar en probabilidad de datos de madre, padre e hijo de la siguiente manera:

$$P(D|g_m, g_f, g_c, H, f, i) = P(s_m|g_m, i)P(G_m|g_m, i)P(s_f|g_f, i)P(G_f|g_f, i)P(s|g_m, g_c, H, f, i)$$

**[0154]** La probabilidad de datos de genotipo de madre ilumina  $g_{mi}$  en SNP  $i$  en comparación con el genotipo verdadero  $g_m$ , suponiendo que los genotipos de ilumina sean correctos, es simplemente:

$$P(G_m|g_m, i) = \begin{cases} 1 & g_{mi} = g_m \\ 0 & g_{mi} \neq g_m \end{cases}$$

**[0155]** La probabilidad de datos de la secuencia madre en SNP  $i$ , en el caso de recuentos  $S_{mi} = (a_{mi}, b_{mi})$ , sin ruido adicional o sesgo involucrado, puede ser la probabilidad binomial definida como:  $P(S_{mi}|i) = P_{X|m}(a_{mi})$  donde  $X|m \sim \text{Binom}(p_m(A), a_{mi}+b_{mi})$  con  $p_m(A)$  definido como:

m	AA	AB	BB	A	B	ninguna llamada
p(A)	1	0,5	0	1	0	0,5

Se aplica una ecuación similar para las probabilidades paternas.

**[0156]** Tenga en cuenta que es posible obtener una respuesta sin los datos principales, especialmente sin los datos principales. Por ejemplo, si no hay datos de genotipo padre  $F$  disponibles, se puede usar  $P(G_f|g_f, i) = 1$ . Si no hay datos de secuencia del padre  $S_f$  está disponible, se puede usar  $P(S_f|g_f, i) = 1$ . La información de diferentes cromosomas se puede agregar usando promedios, promedio ponderado o una función similar.

*Otro método para el cálculo de la fracción fetal*

**[0157]** Aquí se describe otro método para determinar la fracción de ADN fetal en una mezcla de ADN. Una versión de una estimación de máxima verosimilitud de la fracción fetal  $f$  para una prueba de paternidad, prueba de ploidía u otro



propósito, puede derivarse sin el uso de información paterna. Defina  $S_0$  como el conjunto de SNP con genotipo materno 0 (AA),  $S_{0,5}$  como el conjunto de SNP con genotipo materno 0,5 (AB) y  $S_1$  como el conjunto de SNP con genotipo materno 1 (BB). Los posibles genotipos fetales en  $S_0$  son 0 y 0,5, lo que resulta en un conjunto de posibles proporciones

alélicas  $R_0(f) = \{0, \frac{f}{2}\}$ . De forma similar,  $R_{0,5} = \{0,5-f, 0,5, 0,5+f\}$  y  $R_1(f) = \{1-\frac{f}{2}, 1\}$ . Todos o cualquier subconjunto de los conjuntos  $S_0$ ,  $S_{0,5}$  y  $S_1$  se puede utilizar para derivar una estimación de fracción secundaria.

**[0158]** Defina  $N_{a0}$  y  $N_{b0}$  como los vectores formados por recuentos de secuencias para SNP en  $S_0$ ,  $N_{a0,5}$  y  $N_{b0,5}$  de manera similar para  $S_{0,5}$ , y  $N_{a1}$  y  $N_{b1}$  de manera similar para  $S_1$ . La estimación de máxima verosimilitud  $f$  de  $f$ , usando todos los conjuntos de genotipos maternos, se define mediante la ecuación 4.

$$\hat{f} = \arg \max_f P(N_{a0}, N_{b0}|f) P(N_{a0,5}, N_{b0,5}|f) P(N_{a1}, N_{b1}|f) \quad (4)$$

**[0159]** Suponiendo que los recuentos de alelos en cada SNP están condicionados independientemente de la relación de alelos plasmáticos del SNP, las probabilidades pueden expresarse como productos sobre el SNP en cada conjunto:

$$P(N_{a0}, N_{b0}|f) = \prod_{S \in S_0} P(n_{as}, n_{bs}|f)$$

$$P(N_{a1}, N_{b1}|f) = \prod_{S \in S_1} P(n_{as}, n_{bs}|f) \quad (5)$$

donde  $n_{as}$ ,  $n_{bs}$  son los recuentos de SNPs s.

**[0160]** La dependencia de  $F$  es a través de los conjuntos de posibles relaciones de alelo  $R_0(f)$ ,  $R_{0,5}(f)$  y  $R_1(f)$ . La probabilidad de SNP  $P(n_{as}, n_{bs}|f)$  puede aproximarse asumiendo el genotipo de máxima verosimilitud condicionado por  $f$ . A una fracción fetal y una profundidad de lectura razonablemente altas, la selección del genotipo de máxima verosimilitud será de alta confianza. Por ejemplo, con una fracción fetal del 10 por ciento y una profundidad de lectura de 1.000, considere un SNP donde la madre tenga el genotipo 0. Las proporciones de alelo esperadas son 0 y 5 por ciento, lo que será fácilmente distinguible a una profundidad de lectura suficientemente alta. La sustitución del genotipo infantil estimado en la ecuación 5 da como resultado la ecuación completa (6) para la estimación de la fracción fetal.

$$\hat{f} = \arg \max_f \left[ \frac{\prod_{S \in S_0} \left( \max_{r_s \in R_0(f)} P(n_{as}, n_{bs}|r_s) \right) \prod_{S \in S_{0,5}} \left( \max_{r_s \in R_{0,5}(f)} P(n_{as}, n_{bs}|r_s) \right)}{\prod_{S \in S_1} \left( \max_{r_s \in R_1(f)} P(n_{as}, n_{bs}|r_s) \right)} \right]$$

**[0161]** La fracción fetal debe estar en el rango  $[0, 1]$  y, por lo tanto, la optimización puede implementarse fácilmente mediante una búsqueda unidimensional restringida.

**[0162]** Otro método sería la de suma sobre los genotipos posibles en cada SNP, dando como resultado la siguiente expresión (7) para  $P(n_a, n_b|f)$  para un SNP en  $S_0$ . La probabilidad previa  $P(r)$  podría suponerse uniforme sobre  $R_0(f)$ , o podría basarse en las frecuencias de la población. La extensión a los grupos  $S_{0,5}$  y  $S_1$  es trivial.

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

Derivación de probabilidades.

**[0163]** Se puede calcular una confianza a partir de las probabilidades de datos de las dos hipótesis  $H_{\#}$ , es decir, el supuesto padre es el padre biológico y  $H_{\#f}$ , es decir, el supuesto padre no es el padre biológico. El objetivo es calcular  $P(H|D)$ , es decir, la probabilidad de la hipótesis de los datos dados, para cada hipótesis e inferir qué hipótesis es más probable. Esto se puede hacer usando la regla de Bayes:  $P(H|D) \sim P(D|H) * P(H)$  donde  $P(H)$  es el peso previo de la hipótesis, y donde  $P(D|H)$  es la probabilidad de datos dada la hipótesis.

**[0164]** Considere  $P(D|H, f)$ , es decir, la probabilidad de que los datos den hipótesis para una fracción secundaria particular. Si hay disponible una distribución en la fracción secundaria, es posible derivar

$$P(D|H) = \int P(D, f|H) df$$

y, además,

$$P(D|H) = \int P(D|H, f)P(f|H)df$$

5 **[0165]** Tenga en cuenta que  $P(f|H)$  es independiente de la hipótesis, es decir,  $P(f|H) = P(f)$  ya que la fracción secundaria es la misma independientemente de si el presunto padre es el padre biológico o no, y se puede elegir cualquier  $P(f)$  anterior razonable, por ejemplo, un uniforme anterior de 0 a 50% de fracción infantil. Es posible usar solo una fracción secundaria,  $f$ . En este caso,

10

$$P(D|H) = P(D|H, \hat{f})$$

**[0166]** Considere la probabilidad  $P(D|H, f)$ .

La probabilidad de cada hipótesis se deriva

15 en función del modelo de respuesta, la fracción fetal estimada, los genotipos maternos, los supuestos genotipos padre, las frecuencias de población de alelos, los recuentos de alelos plasmáticos y los SNP. Deje que  $D$  represente los datos como se definió anteriormente.

**[0167]** Se puede suponer que la observación en cada SNP está condicionada de forma independiente a la relación de alelo plasmático, por lo tanto, la probabilidad de una hipótesis de paternidad es el producto de las probabilidades en los SNP:

20

$$P(D|H, f) = \prod_{\text{SNPs } i} P(D|H, f, i)$$

25

**[0168]** Las siguientes ecuaciones describen cómo se puede derivar la probabilidad de un único SNP  $i$  y una sola fracción de hijo  $f$ . La ecuación 8 es una expresión general para la probabilidad de cualquier hipótesis  $H$ , que luego se desglosará en los casos específicos de  $H_{tf}$  y  $H_{wf}$ . Tenga en cuenta que los genotipos,  $g_m$ ,  $g_{tf}$ ,  $g_{df}$  y  $g_c$ , toman valores en  $\{AA, AB, BB\}$  que se traduce en  $\{0, 0.5, 1\}$  donde  $AA = 0$ ,  $AB = 0.5$ ,  $BB = 1$ . Además,  $g_{tf}$  denota los genotipos del verdadero padre y  $g_{df}$  denota los genotipos representados por los datos proporcionados para el padre. En el caso de  $H_{tf}$ ,  $g_{tf}$  y  $g_{df}$  son equivalentes.

30

$$P(D|f, H, i) = \sum_{g_m, g_{tf}, g_{df}, g_c \in \{0, 0.5, 1\}} P(D|g_m, g_{df}, g_c, f, H, i) * P(g_c|g_m, g_{tf}, H) * P(g_m|i) * P(g_{tf}|i) * P(g_{df}|i) \quad (8)$$

35

**[0169]** En el caso de la hipótesis  $H_{tf}$ , el supuesto padre es el padre biológico y los genotipos fetales se heredan de los genotipos maternos y los presuntos genotipos del padre. La ecuación anterior se simplifica a:

40

$$P(D|f, H = H_{tf}, i) = \sum_{g_m, g_{tf}, g_c \in \{0, 0.5, 1\}} P(D|g_m, g_{tf}, g_c, f, H = H_{tf}, i) * P(g_c|g_m, g_{tf}, H = H_{tf}) * P(g_m|i) * P(g_{tf}|i)$$

45

50

Además,

$$P(g_c|g_m, g_{tf}, H = H_{tf}) = P(g_c|g_m, g_{tf})$$

55

y

$$P(D|g_m, g_{tf}, g_c, f, H = H_{tf}, i) = P(s_m|g_m, i)P(G_m|g_m, i)P(s_f|g_{tf}, i)P(G_f|g_{tf}, i)P(s|g_m, g_c, f, i) \quad (9)$$

60

**[0170]** En el caso de  $H_{wf}$ , el supuesto padre no es el padre biológico. Se puede generar una estimación de los verdaderos genotipos del padre utilizando las frecuencias de población en cada SNP. Por lo tanto, las probabilidades de los genotipos infantiles pueden determinarse por los genotipos maternos conocidos y las frecuencias de la

65

población, es decir, los datos no proporcionan información adicional sobre los genotipos del padre biológico. En este caso, la ecuación anterior no se simplifica aún más y permanece como:

$$P(D|f, H = H_{wf}, i) = \sum_{g_m, g_{tf}, g_{df}, g_c \in \{0,0.5,1\}} P(D|g_m, g_{df}, g_c, f, H = H_{wf}, i) * P(g_c|g_m, g_{tf}, H = H_{wf}) * P(g_m|i) * P(g_{tf}|i) * P(g_{df}|i)$$

Además,

$$P(g_c|g_m, g_{tf}, H = H_{wf}) = P(g_c|g_m, g_{tf})$$

donde la única información sobre  $g_{tf}$  son los antecedentes de la población y:

$$P(D|g_m, g_{tf}, g_c, f, H = H_{wf}, i) = P(s_m|g_m, i)P(G_m|g_m, i)P(s_f|g_{df}, i)P(G_f|g_{df}, i)P(s|g_m, g_c, f, i) \quad (10)$$

**[0171]** En ambas expresiones de las probabilidades,  $P(D|f, H, i)$ , es decir, para ambas hipótesis, el modelo de respuesta,  $P(s|g_m, g_{df}, g_c, f, H)$  está generalizado. Se mencionan ejemplos específicos en otras partes del documento en discusiones sobre modelos de plataforma generales. Algunos ejemplos para el modelo de reposo incluyen la distribución binomial, la distribución beta-binomial, la distribución multivariada de Póya, una distribución empírica estimada a partir de datos de entrenamiento o funciones similares como se discutió anteriormente.

**[0172]** La confianza  $C_p$  en la paternidad correcta se puede calcular a partir de las probabilidades  $P(D|H_{tf})$  y  $P(D|H_{wf})$ . Este cálculo puede calcularse utilizando la regla de Bayes de la siguiente manera:

$$C_p = \frac{P(D|H_{tf})P(H_{tf})}{P(D|H_{tf})P(H_{tf}) + P(D|H_{wf})P(H_{wf})}$$

o escrito para un caso más específico como producto sobre SNP de las dos probabilidades:

$$C_p = \frac{\prod_i P(D|H_{tf}, G_{ms}, G_{tf}, f)}{\prod_s P(n_{as}, n_{bs}|H_{t}, G_{ms}, G_{tf}, f) + \prod_s P(n_{as}, n_{bs}|H_{f}, G_{ms}, G_{tf}, f)}$$

La confianza puede calcularse de la siguiente manera:

$$C_p = \frac{P(D|H_{tf})}{P(D|H_{tf}) + P(D|H_{wf})}$$

También son posibles otras funciones razonables de las probabilidades.

## Sección experimental

### *Experimento 1*

**[0173]** Veintiún mujeres embarazadas con la paternidad confirmada y edades gestacionales entre 6 y 21 semanas se inscribió. Los participantes donaron sangre voluntariamente como parte de nuestro programa de investigación aprobado por el IRB, y fueron extraídos de centros de FIV, oficinas de obstetricia y la población general en diferentes lugares de los EE.UU. ADN libre de células (ADNf) aislado del plasma materno, junto con el ADN de la madre y supuesto padre, se amplificaron y midieron utilizando una matriz SNP. Se usó un método informático divulgado en este documento para excluir o incluir la paternidad de 21 padres correctos y 36.400 padres incorrectos al comparar a cada supuesto padre con una distribución de referencia generada a partir de un conjunto de más de 5.000 individuos no relacionados. 20 de 21 muestras tenían suficiente ADN fetal para devolver los resultados. Veinte de veinte (100%) de las inclusiones de paternidad fueron correctas. 36.382 de 36.382 exclusiones de paternidad fueron correctas (100%), con 18 "sin llamadas" debido a la similitud genética intermedia. No hubo llamadas equivocadas.

**[0174]** La población estaba compuesta por parejas que donaron su sangre para la investigación prenatal. Las mujeres tuvieron que tener embarazos únicos, estar en el primer o segundo trimestre y haber confirmado la paternidad. Se

recogieron muestras de sangre de mujeres usando tubos de sangre SIN CÉLULAS (STRECK) que contienen conservantes de glóbulos blancos, y se recogieron muestras genéticas del padre, ya sea como una muestra de sangre (EDTA) o bucal. Se obtuvo el consentimiento informado por escrito de todos los participantes, y las muestras genéticas se obtuvieron de pacientes inscritos en un estudio aprobado por el IRB.

**[0175]** La sangre de la madre se centrifugó para aislar la capa leucocitaria y el suero. El ADN genómico en el manto y la supuesta capa leucocitaria paterna y el ADN en el suero materno se prepararon para el análisis y se ejecutaron en matrices SNL CYTO12 ILLUMINA INFINIUM utilizando protocolos estándar. Brevemente, se aisló el ADN del suero usando el kit QIAGEN CIRCULATING NUCLEIC ACID y se eluyó en 45 µl de tampón de acuerdo con las instrucciones del fabricante. Se usó eluato de veinte microlitros en una reacción de terminación roma en 1x tampón NEB 4, dNTP 0,42 mM y ADN polimerasa T4 2,5 U (NEW ENGLAND BIOLABS), se incubó a 20°C durante 30 minutos, luego a 75°C durante 15 minutos. Se agregaron tres microlitros de mezcla de ligadura (0,5 µl 10xNEB 4, 1 µl 10 mM ATP, 1 µl T4 PNK (NEW ENGLAND BIOLABS), 0,5 µl T4 ADN Ligasa (NEW ENGLAND BIOLABS)) y las muestras se incubaron a 16 C durante 24 horas, luego 75 C por 15 min. La muestra se transfirió al ensayo estándar ILLUMINA INFINIUM junto con el ADN genómico materno y el supuesto padre paterno. En resumen, 24 µl de ADN se amplificaron al genoma completo a 37°C durante 20-24 horas seguido de fragmentación y precipitación. El precipitado se resuspendió luego en tampón de hibridación, se desnaturalizó por calor y se transfirió a matrices Cyto12 SNP usando un TECAN EVO. Las matrices se incubaron a 48°C durante al menos 16 horas, se tiñeron con X (INFINIUM II CHEMISTRY) y se lavaron en TECAN EVO, y finalmente se escanearon. Las intensidades de matriz se extrajeron usando BEADSTIDUO (ILLUMINA).

**[0176]** El método de la informática descrito generó una estadística de prueba que mide el grado de similitud genética entre el feto y otro individuo. Esta estadística de prueba se calculó tanto para el presunto padre como para un conjunto de más de 5.000 individuos no relacionados. Una sola prueba de rechazo de hipótesis determinó si la estadística calculada para el supuesto padre podría excluirse de la distribución formada por los individuos de referencia no relacionados. Si el supuesto padre podía ser rechazado del conjunto no relacionado, entonces se producía una inclusión de paternidad; de lo contrario, se excluyó la paternidad. Para las 20 muestras con suficiente ADN, la prueba de paternidad se realizó contra 20 padres correctos, y para 1,820 padres incorrectos seleccionados al azar.

**[0177]** Se solicitó una inclusión de paternidad cuando el valor p de la estadística de prueba del presunto padre sobre la distribución de individuos no relacionados fue inferior a  $10^{-4}$ . Esto significa que, en teoría, no se espera que más de uno de cada 10.000 individuos no relacionados muestren tanta similitud genética con el feto. Se llamó una "no llamada" cuando el valor p está entre  $10^{-4}$  y 0,02. Se realizó una llamada de "ADN fetal insuficiente" cuando el ADN fetal constituía menos del 2% del ADN plasmático. El conjunto de individuos no relacionados utilizados para generar la distribución esperada estaba compuesto por individuos de una amplia variedad de antecedentes raciales, y la determinación de inclusión o exclusión de paternidad se recalculó para conjuntos de individuos no relacionados de diferentes razas, incluida la raza indicada para el supuesto padre. Los resultados de inclusión y exclusión fueron generados automáticamente por el algoritmo, y no fue necesaria la intervención humana.

**[0178]** En conclusión, se ensayaron veinte uno muestras de sangre materna con la paternidad conocida. Veinte de las 21 muestras arrojaron resultados, mientras que una tenía ADN fetal insuficiente para el análisis; esta muestra fue extraída de una mujer a las 8 semanas de edad gestacional. Veinte de veinte (100%) resultados confirmaron la paternidad correcta, cada uno con un valor p de  $<10^{-4}$ . Cada una de las 20 muestras con suficiente fracción fetal se analizó contra un conjunto aleatorio de 1.820 padres incorrectos, para un total de 36,400 pruebas de paternidad individuales. 36.382 de estos análisis arrojaron un resultado; 36.382 de 36.382 (100%) correctamente tenían la paternidad excluida con un valor p de mayor que  $10^{-4}$ , y 18 de 36.400 (0,05%) fueron llamados "ninguna llamada", con un valor de p entre  $10^{-4}$  y 0,02. No hubo exclusiones o inclusiones incorrectas de paternidad.

**[0179]** Nueve de 21 muestras habían confirmado la paternidad debido al control de la fertilización durante la FIV con la paternidad correcta confirmada después de la fertilización a través del diagnóstico genético previo a la implantación. Doce muestras confirmaron la paternidad mediante pruebas de paternidad independientes de ADN genómico fetal/infantil, realizadas por el Centro de Diagnóstico de ADN, Fairfield, Ohio.

## Experimento 2

**[0180]** En un experimento, se prepararon cuatro muestras de plasma materno y se amplificaron usando un protocolo de 9,600-plex hemi-anidado. Las muestras se prepararon de la siguiente manera: se centrifugaron entre 15 y 40 ml de sangre materna para aislar la capa leucocitaria y el plasma. El ADN genómico en la madre y se preparó a partir de la capa leucocitaria y el ADN paterno se preparó a partir de una muestra de sangre o muestra de saliva. El ADN libre de células en el plasma materno se aisló usando el kit de QIAGEN CIRCULATING NUCLEIC ACID y se eluyó en 45 µl de tampón TE según las instrucciones del fabricante. Los adaptadores de ligadura universal se añadieron al final de cada molécula de 35 µl de ADN de plasma purificado y las bibliotecas se amplificaron durante 7 ciclos usando cebadores específicos del adaptador. Las bibliotecas se purificaron con cuentas AGENCOURT AMPURE y se eluyeron en 50 µl de agua.

**[0181]** 3 µl del ADN se amplificó con 15 ciclos de STA (95°C durante 10 min para la activación inicial de la polimerasa,

a continuación, 15 ciclos de 95°C durante 30 s; 72°C durante 10 s; 65°C durante 1 min; 60°C por 8 min; 65°C por 3 min y 72°C por 30s; y una extensión final a 72°C por 2 min) usando una concentración de cebador 14,5 nM de 9600 cebadores inversos etiquetados específicos del objetivo y un cebador directo específico de adaptador de biblioteca a 500 nM.

**[0182]** El protocolo de PCR hemi-anidado implicaba una segunda amplificación de una dilución del primer producto de STA durante 15 ciclos de STA (95°C durante 10 minutos para la activación inicial de la polimerasa, luego 15 ciclos de 95°C durante 30 segundos; 65°C por 1 min; 60°C por 5 min; 65°C por 5 min y 72°C por 30s; y una extensión final a 72°C por 2 min usando una concentración de etiqueta inversa de 1000 nM, y una concentración de 16,6 u nM para cada uno de los cebadores directos específicos de objetivo de 9600.

**[0183]** Luego se amplificó una alícuota de los productos STA mediante PCR estándar durante 10 ciclos con 1 uM de cebadores directos y de código de barras inversos específicos de etiqueta para generar bibliotecas de secuenciación con código de barras. Una alícuota de cada biblioteca se mezcló con las bibliotecas de diferentes códigos de barras y se purificó usando una columna de centrifugación.

**[0184]** De esta manera, 9.600 cebadores se utilizaron en las reacciones de un solo pozo; los cebadores fueron diseñados para apuntar SNPs encontrados en los cromosomas 1, 2, 13, 18, 21, X e Y. Los amplicones fueron secuenciados utilizando un secuenciador ILLUMINA GAIIIX. Por muestra, el secuenciador generó aproximadamente 3,9 millones de lecturas, con 3,7 millones de lecturas asignadas al genoma (94%), y de ellas, 2,9 millones de lecturas (74%) asignadas a SNP específicos con una profundidad promedio de lectura de 344 y una mediana de profundidad de lectura de 255. Se encontró que la fracción fetal para las cuatro muestras era 9,9%, 18,9%, 16,3% y 21,2%.

**[0185]** Muestras relevantes de ADN genómico materno y paternal fueron amplificadas usando un protocolo 9600-plex semi-anidado y secuenciadas. El protocolo semi-anidado es diferente en el sentido de que aplica 9.600 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 PCR del código de barras fueron las mismas que para el protocolo hemi-anidado.

**[0186]** Los datos de secuenciación se analizaron usando los métodos informáticos descritos en este documento y se determinó que cada uno de un conjunto de diez machos no relacionados de un conjunto de referencia no era el padre biológico de cada uno de los fetos gestantes.

### *Experimento 3*

**[0187]** En un experimento, se amplificaron 45 conjuntos de células usando un protocolo semi-anidado de 1,200 plex, secuenciado, y se hicieron determinaciones de ploidía en tres cromosomas. Tenga en cuenta que este experimento está destinado a simular las condiciones de realizar pruebas de paternidad en células fetales individuales obtenidas de sangre materna, o en muestras forenses donde está presente una pequeña cantidad de ADN del niño. Se colocaron 15 células individuales y 30 conjuntos de tres células en 45 tubos de reacción individuales para un total de 45 reacciones donde cada reacción contenía células de una sola línea celular, pero las diferentes reacciones contenían células de diferentes líneas celulares. Las células se prepararon en 5 ul de tampón de lavado y se lisaron añadiendo 5 ul de tampón de lisis ARCTURUS PICOPURE (APPLIED BIOSYSTEMS) e incubando a 56°C durante 20 min, 95°C durante 10 min.

**[0188]** El ADN de las única/tres células se amplificaron con 25 ciclos de STA (95°C durante 10 min para la activación inicial de la polimerasa, luego 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) usando una concentración de cebador 50 nM de 1200 cebadores directos y marcados hacia atrás específicos de la diana.

**[0189]** El protocolo de PCR semi-anidado incluyó tres amplificaciones paralelas de una dilución del primer producto de STA durante 20 ciclos de STA (95°C durante 10 minutos para la activación inicial de la polimerasa, luego 15 ciclos de 95°C durante 30 segundos; 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) usando una concentración de cebador específico de etiqueta inversa de 1000 nM, y una concentración de 60 nM para cada uno de los 400 cebadores directos anidados específicos de la diana. En las tres reacciones paralelas de 400-plex, el total de 1200 dianas amplificadas en la primera STA se amplificaron así.

**[0190]** Una parte alícuota de los productos STA se amplificó luego por PCR estándar durante 15 ciclos con 1 uM de cebadores inversos directos y con código de barras específicos de etiqueta 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 usando una columna giratoria.

**[0191]** De esta manera, se usaron 1.200 cebadores en las reacciones unicelulares; los cebadores se diseñaron para apuntar a los SNP encontrados en los cromosomas 1, 21 y X. Los amplicones se secuenciaron luego usando un secuenciador ILLUMINA GAIIIX. Por muestra, el secuenciador generó aproximadamente 3,9 millones de lecturas, con un mapeo de 500.000 a 800.000 millones de lecturas al genoma (74% a 94% de todas las lecturas por muestra).

[0192] Relevantes muestras de ADN genómico materno y paterno de líneas celulares se analizaron utilizando el mismo grupo de ensayo 1200-plex semianidado con un protocolo similar con menos ciclos y segunda STA 1200-Plex, y se secuenciaron.

- 5 [0193] Los datos de secuenciación se analizaron usando los métodos informáticos descritos en el presente documento y se determinó que cada uno de un conjunto de diez machos no relacionados de un conjunto de referencia no era el padre biológico del individuo diana para cada una de las 45 células.

*ADN de niños de embarazos previos en sangre materna*

- 10 [0194] Una dificultad para las pruebas de paternidad prenatales no invasivas es diferenciar las células fetales del embarazo actual de las células fetales de los embarazos anteriores. Algunos creen que la materia genética de embarazos anteriores desaparecerá después de un tiempo, pero no se ha demostrado evidencia concluyente. Es posible determinar el ADN fetal presente en la sangre materna de origen paterno (es decir, el ADN que el feto heredó del padre) usando el método PARENTAL SUPPORT™ (PS) y el conocimiento del genoma paterno. Este método puede utilizar información genética parental por fases. Es posible eliminar el genotipo parental de la información genotípica sin fase utilizando datos genéticos de los abuelos (como datos genéticos medidos de un espermatozoides del abuelo), o datos genéticos de otros niños nacidos, o una muestra de un aborto espontáneo. También se podría eliminar la información genética sin fases mediante una fase basada en HapMap o un haplotipado de células paternas. El haplotipado exitoso se ha demostrado al detener las células en la fase de la mitosis cuando los cromosomas son haces apretados y al usar microfluidos para colocar cromosomas separados en pozos separados. Es posible utilizar los datos haplotípicos paternos 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 está 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 sufriera una trisomía. Además, es posible determinar si el ADN fetal no es del padre actual, en cuyo caso se podría usar otros métodos como la prueba triple para predecir anomalías genéticas.

- 25 [0195] Puede haber otras fuentes de material genético fetal disponibles a través de métodos distintos a la extracción de sangre. En el caso del material genético fetal disponible en la sangre materna, hay dos categorías principales: (1) células fetales enteras, por ejemplo, glóbulos rojos fetales nucleados o eritroblastos, y (2) ADN fetal flotante libre. En el caso de las células fetales enteras, existe alguna evidencia de que las células fetales pueden persistir en la sangre materna durante un período prolongado de tiempo, de modo que es posible aislar una célula de una mujer embarazada que contiene el ADN de un niño o el feto de un embarazo previo. También hay evidencia de que el ADN fetal flotante se elimina del sistema en cuestión de semanas. Un desafío es cómo determinar la identidad del individuo cuyo material genético está contenido en la célula, a saber, asegurar que el material genético medido no sea de un feto de un embarazo anterior. En la presente divulgación, el conocimiento del material genético materno se puede utilizar para garantizar que el material genético en cuestión no sea material genético materno. Existen varios métodos para lograr este objetivo, incluidos los métodos basados en informática, como PARENTAL SUPPORT™, como se describe en este documento o cualquiera de las patentes a las que se hace referencia en este documento.

- 40 [0196] En la presente descripción, la sangre extraída de la madre embarazada puede ser separada en una fracción que comprende ADN fetal de libre flotación, y una fracción que comprende glóbulos rojos nucleados. El ADN flotante libre puede enriquecerse opcionalmente y puede medirse la información genotípica del ADN. A partir de la información genotípica medida del ADN de flotación libre, el conocimiento del genotipo materno puede usarse para determinar aspectos del genotipo fetal. Estos aspectos pueden referirse al estado de ploidía, y/o un conjunto de identidades de alelos. Luego, las células nucleadas individuales que son presumiblemente o posiblemente de origen fetal pueden ser genotipadas usando métodos descritos en otra parte de este documento, y otras patentes de referencia, especialmente las mencionadas en este documento. El conocimiento del genoma materno permitiría determinar si alguna célula sanguínea dada es genéticamente materna. Y los aspectos del genotipo fetal que se determinaron como se describió anteriormente permitirían determinar si la única célula sanguínea se deriva genéticamente del feto que se está gestando actualmente. En esencia, 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 actualmente, o (c) genéticamente de un feto de un embarazo anterior.

60

65

## REIVINDICACIONES

1. Un método *ex vivo* para establecer si un supuesto padre es el padre biológico de un feto que se está gestando en una madre embarazada, el método comprende:

realizar mediciones genotípicas de alelos de polimorfismo de un solo nucleótido (SNP), en una pluralidad de loci polimórficos, en material genético obtenido del supuesto padre, material genético obtenido de la madre embarazada y en una muestra mixta de ADN procedente de una muestra de sangre obtenida de la madre embarazada, donde la muestra mixta de ADN comprende ADN fetal y ADN materno;

en donde dichas mediciones genotípicas son mediciones de salida de una plataforma de genotipado que comprende la identidad de la pluralidad de nucleótidos en la pluralidad de loci polimórficos;

determinar, en una computadora, la probabilidad de que el supuesto padre sea el padre biológico del feto que se gesta en la madre embarazada obteniendo frecuencias poblacionales de alelos para cada locus en la pluralidad de loci polimórficos;

crear una partición del rango de posibles fracciones de ADN fetal en la muestra mixta de ADN, en donde la partición de posibles fracciones fetales varía de 2% de fracción fetal a 30% de fracción fetal, determinándose la fracción fetal realizando mediciones genotípicas en una pluralidad de loci polimórficos de la muestra mixta de ADN; obtener datos genotípicos en la pluralidad de loci polimórficos del supuesto padre;

calcular una probabilidad de que el presunto padre sea el padre biológico del feto utilizando las mediciones genotípicas realizadas en el material genético del presunto padre y la muestra mixta de ADN; para cada una de las posibles fracciones fetales en la partición utilizando una estimación de máxima verosimilitud o una técnica *a posteriori* máxima; y

determinar la probabilidad de que el presunto padre sea el padre biológico del feto combinando las probabilidades calculadas de que el presunto padre es el padre biológico del feto para cada una de las posibles fracciones fetales en la partición;

calcular una probabilidad de que el presunto padre no sea el padre biológico del feto dadas las mediciones genotípicas realizadas a partir del material genético de la madre, el presunto padre y la muestra mixta de ADN, y las frecuencias de población de alelos obtenidas dadas; para cada una de las posibles fracciones fetales en la partición; y

determinar la probabilidad de que el presunto padre no sea el padre biológico del feto combinando las probabilidades calculadas de que el presunto padre no es el padre biológico del feto para cada una de las posibles fracciones fetales en la partición; y

establecer si el presunto padre es el padre biológico del feto utilizando una única prueba de rechazo de hipótesis, una estimación de máxima verosimilitud o una técnica *a posteriori* máxima dada la probabilidad determinada de que el presunto padre sea el padre biológico del feto.

2. El método de la reivindicación 1, en donde la muestra mixta de ADN comprende: (i) ADN que era de ADN flotante libre en una fracción de plasma de la muestra de sangre de la madre embarazada, (ii) sangre entera materna o (iii) una fracción de sangre materna que contiene células nucleadas.

3. El método de la reivindicación 1, en donde la determinación comprende:

calcular una estadística de prueba para el supuesto padre y el feto, en donde la estadística de prueba indica un grado de similitud genética entre el presunto padre y el feto, y en donde la estadística de prueba se basa en las mediciones genotípicas hechas del material genético del supuesto padre, las mediciones genotípicas hechas de la muestra mixta de ADN y las mediciones genotípicas obtenidas del material genético de la madre embarazada;

calcular una distribución de una estadística de prueba para una pluralidad de individuos que no están genéticamente relacionados con el feto, donde cada estadística de prueba calculada indica un grado de similitud genética entre un individuo no relacionado de la pluralidad de individuos que no están relacionados con el feto y el feto, en donde la estadística de prueba se basa en mediciones genotípicas hechas de ADN del individuo no relacionado, las mediciones genotípicas hechas de la muestra mixta de ADN y las mediciones genotípicas obtenidas del material genético de la madre embarazada;

en donde la estadística de prueba es una función de los valores porcentuales de las distribuciones de las mediciones de SNP para contextos parentales seleccionados que demuestra la diferencia en los valores, en donde cada contexto parental es un conjunto de alelos donde el genotipo de la madre y el genotipo del supuesto padre del el feto es el mismo para cada alelo;

calcular una probabilidad de que la estadística de prueba calculada para el supuesto padre y el feto sea parte de la distribución de la estadística de prueba calculada para la pluralidad de individuos no relacionados y el feto; y

determinar la probabilidad de que el presunto padre sea el padre biológico del feto utilizando la probabilidad de que la estadística de prueba calculada para el presunto padre sea parte de la distribución de la estadística de prueba calculada para la pluralidad de individuos no relacionados y el feto.

4. El método de la reivindicación 3, en donde establecer si un presunto padre es el padre biológico del feto comprende además:

establecer que el presunto padre es el padre biológico del feto al rechazar una hipótesis de que el presunto padre no está relacionado con el feto si la probabilidad de que el presunto padre sea el padre biológico del feto está por encima de un umbral superior; o

establecer que el presunto padre no es el padre biológico del feto al no rechazar una hipótesis de que el presunto padre no está relacionado con el feto si la probabilidad de que el presunto padre sea el padre biológico del feto está por debajo de un umbral inferior; o

no establecer si un supuesto padre es el padre biológico del feto si la probabilidad se encuentra entre el umbral inferior y el umbral superior, o si la probabilidad no se determinó con la confianza suficientemente alta.

**5.** El método de la reivindicación 1, en donde calcular la probabilidad de que el supuesto padre sea el padre biológico del feto y calcular la probabilidad de que el supuesto padre no sea el padre biológico del feto comprende además:

calcular, para cada uno de la pluralidad de loci polimórficos, una probabilidad de datos de secuencia observados en un locus particular usando un modelo de respuesta de plataforma, una o una pluralidad de fracciones en la posible partición de fracciones fetales, una pluralidad de proporciones de alelos para la madre, una pluralidad de proporciones de alelos para el supuesto padre, y una pluralidad de proporciones alélicas para el feto;

calcular una probabilidad de que el presunto padre sea el padre biológico combinando la probabilidad de los datos de secuencia observados en cada locus polimórfico sobre todas las fracciones fetales en la partición, sobre las proporciones alelo madre en el conjunto de loci polimórficos, sobre las proporciones alelo alegadas padre en el conjunto de loci polimórficos, y sobre las raciones de alelos fetales en el conjunto de loci polimórficos;

calcular una probabilidad de que el supuesto padre no sea el padre biológico combinando la probabilidad de los datos de secuencia observados en cada locus polimórfico sobre todas las fracciones fetales en la partición, sobre proporciones del alelo materno en el conjunto de loci polimórficos, sobre las frecuencias de población para el conjunto de loci polimórficos, y sobre las proporciones de alelos fetales en el conjunto de loci polimórficos;

calcular una probabilidad de que el presunto padre sea el padre biológico basándose en la probabilidad de que el presunto padre sea el padre biológico; y

calcular una probabilidad de que el presunto padre no sea el padre biológico basándose en la probabilidad de que el presunto padre no sea el padre biológico.

**6.** El método de la reivindicación 1, en donde establecer si un supuesto padre es el padre biológico de un feto comprende además:

establecer que el supuesto padre es el padre biológico si la probabilidad calculada de que el supuesto padre es el padre biológico del feto es significativamente mayor que la probabilidad calculada de que el presunto padre no sea el padre biológico; o

establecer que el presunto padre no es el padre biológico del feto si la probabilidad calculada de que el presunto padre es el padre biológico es significativamente mayor que la probabilidad calculada de que el presunto padre no es el padre biológico.

**7.** El método de la reivindicación 1, en donde se calcula una confianza para la determinación establecida de si el supuesto padre es el padre biológico del feto.

**8.** El método de la reivindicación 1, en donde la obtención de mediciones genotípicas del material genético de la madre embarazada comprende realizar mediciones genotípicas en una muestra de material genético de la madre embarazada que consiste en material genético materno.

**9.** El método de la reivindicación 1, en donde la obtención de mediciones genotípicas del material genético de la madre embarazada comprende:

inferir qué mediciones genotípicas de las mediciones genotípicas realizadas en la muestra mixta de ADN son probablemente atribuibles al material genético de la madre embarazada usando frecuencias de una población de alelos y el uso de las mediciones genotípicas que se infiere que son atribuibles al material genético de la madre como las mediciones genotípicas obtenidas.



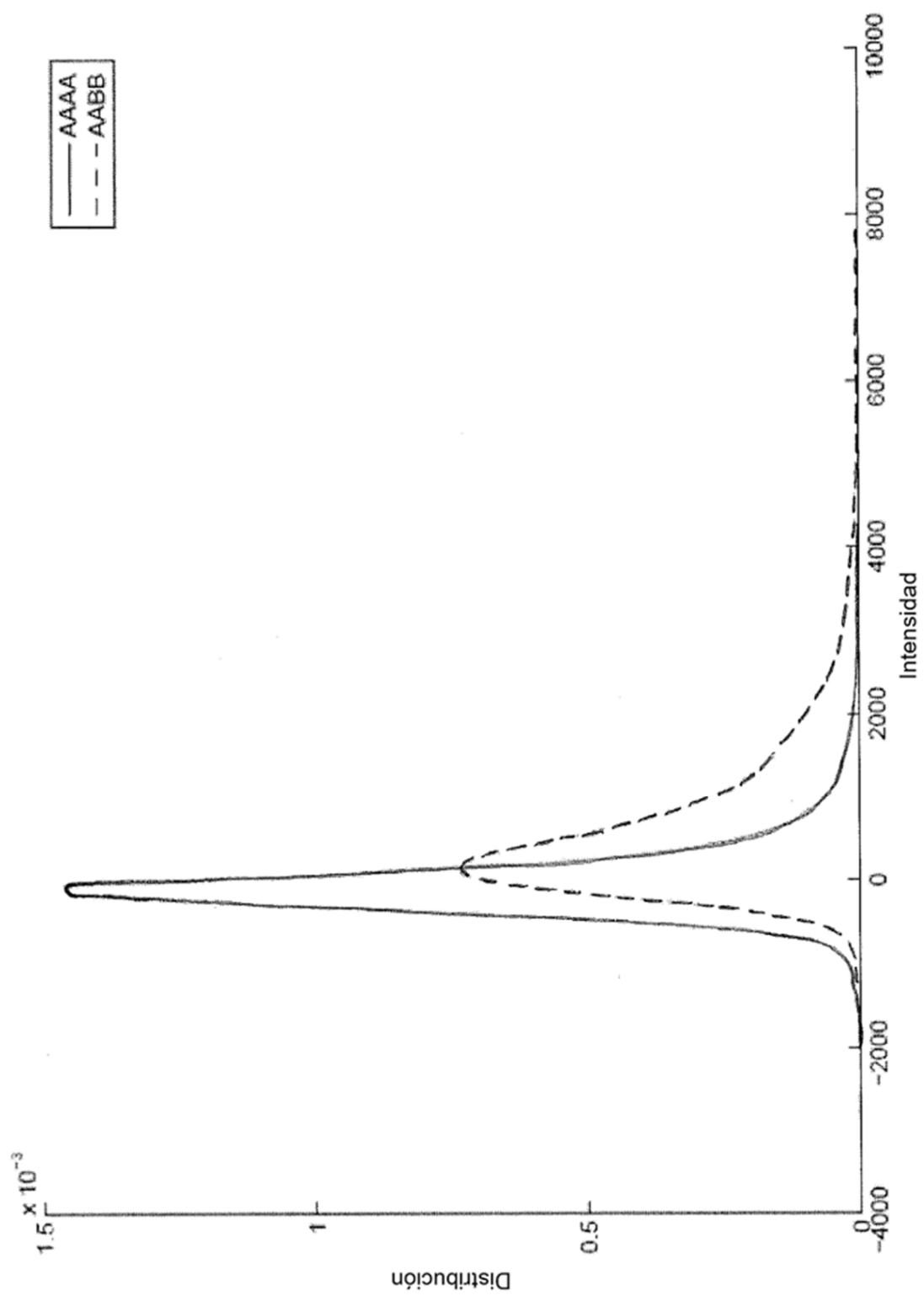


Figura 1

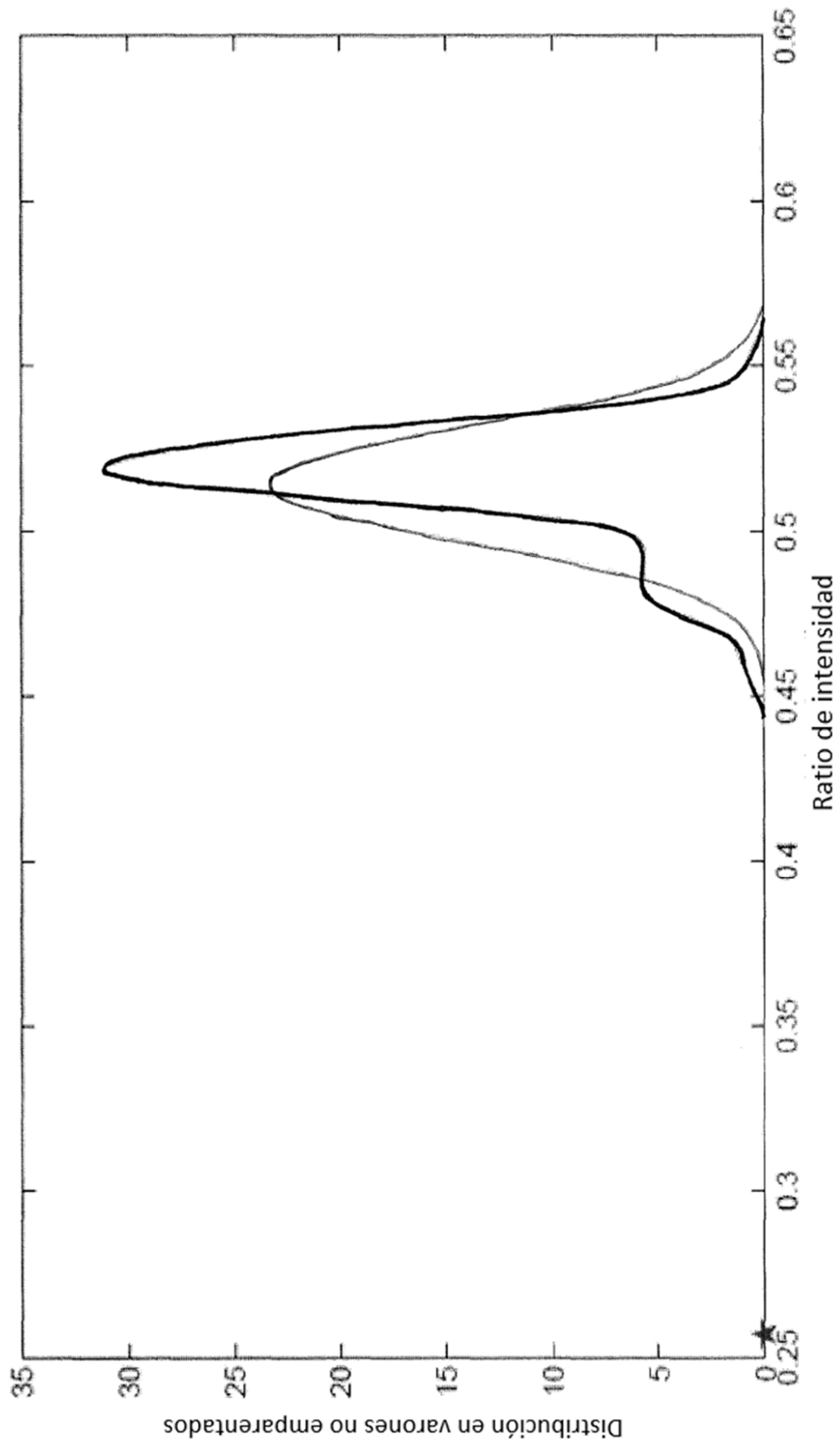


Figura 2

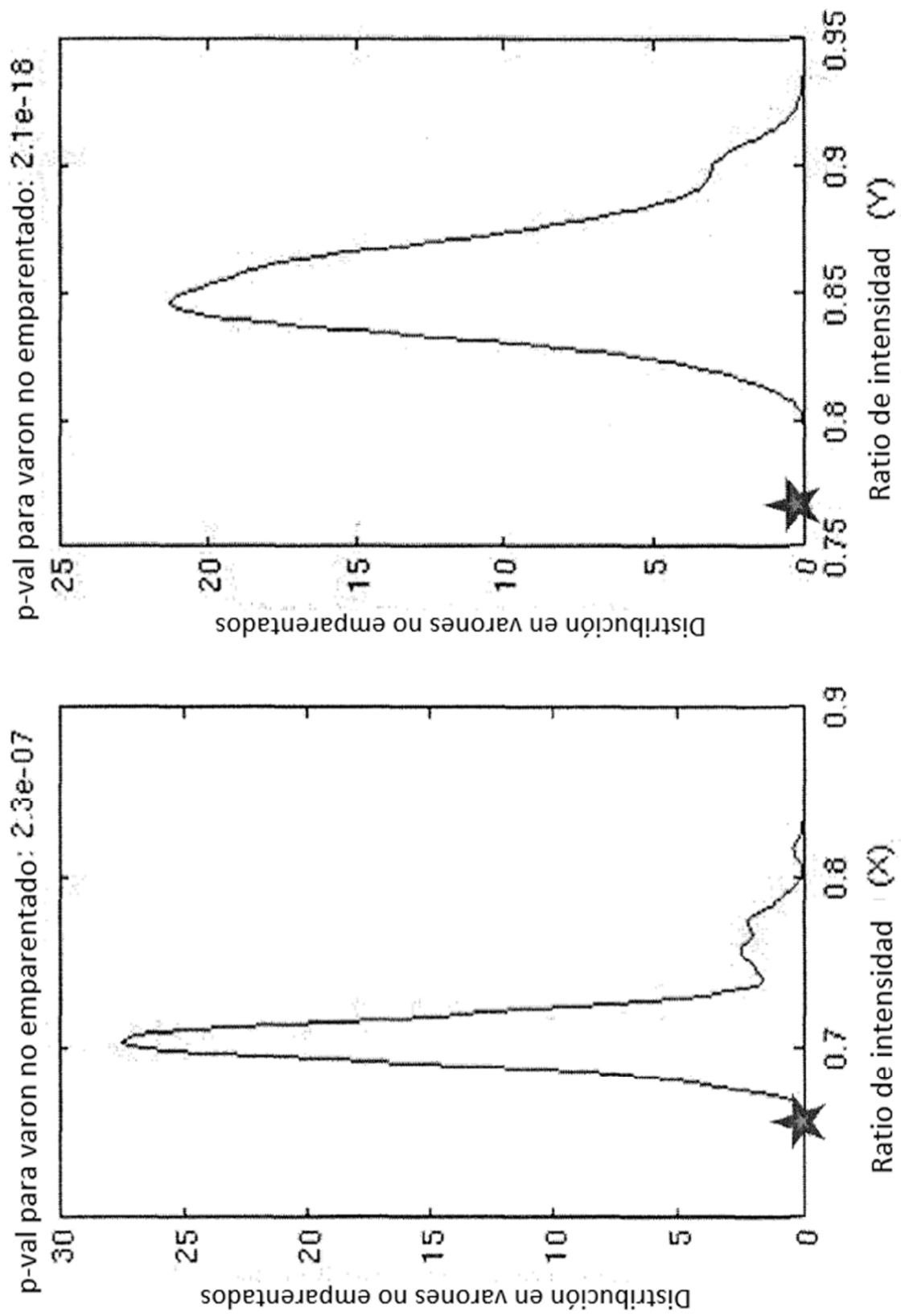


Figura 3

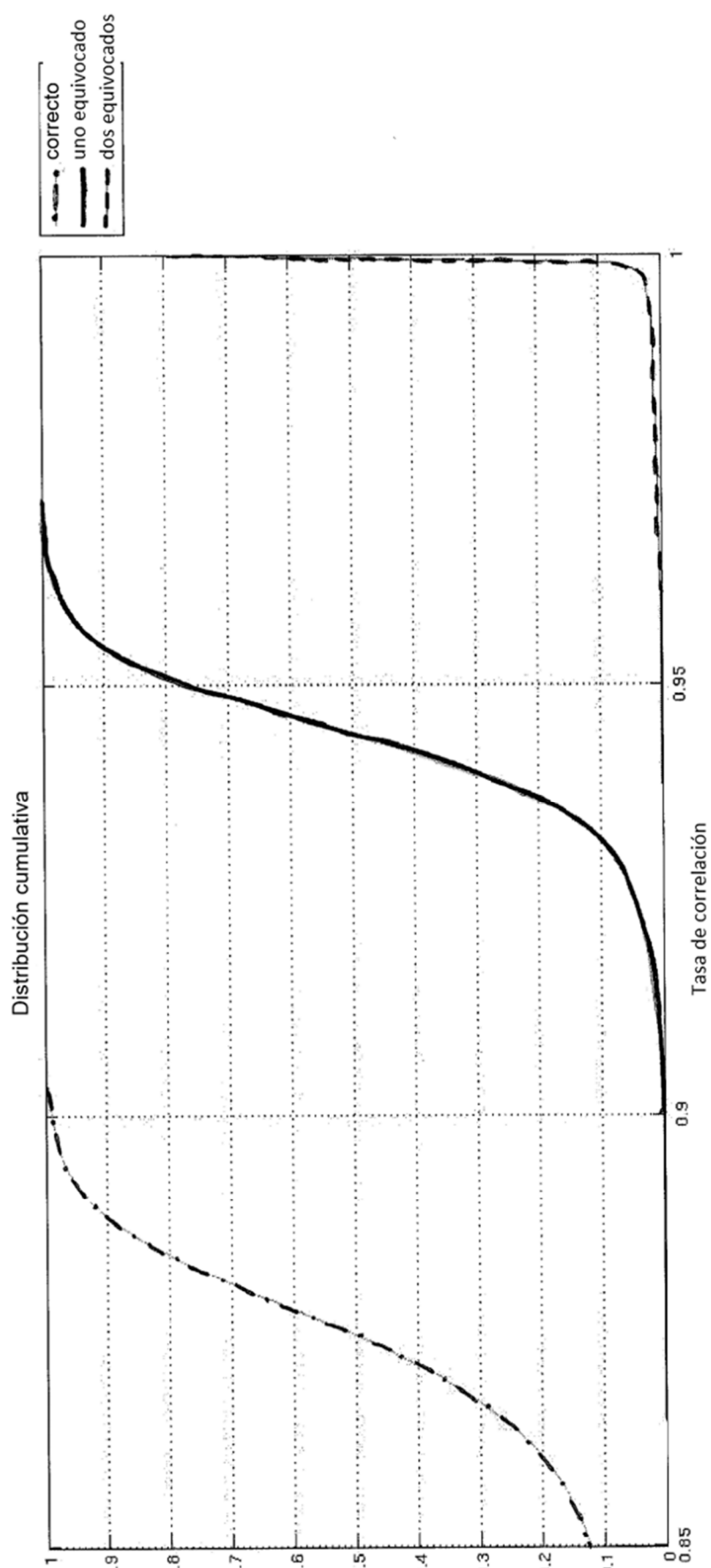


Figura 4

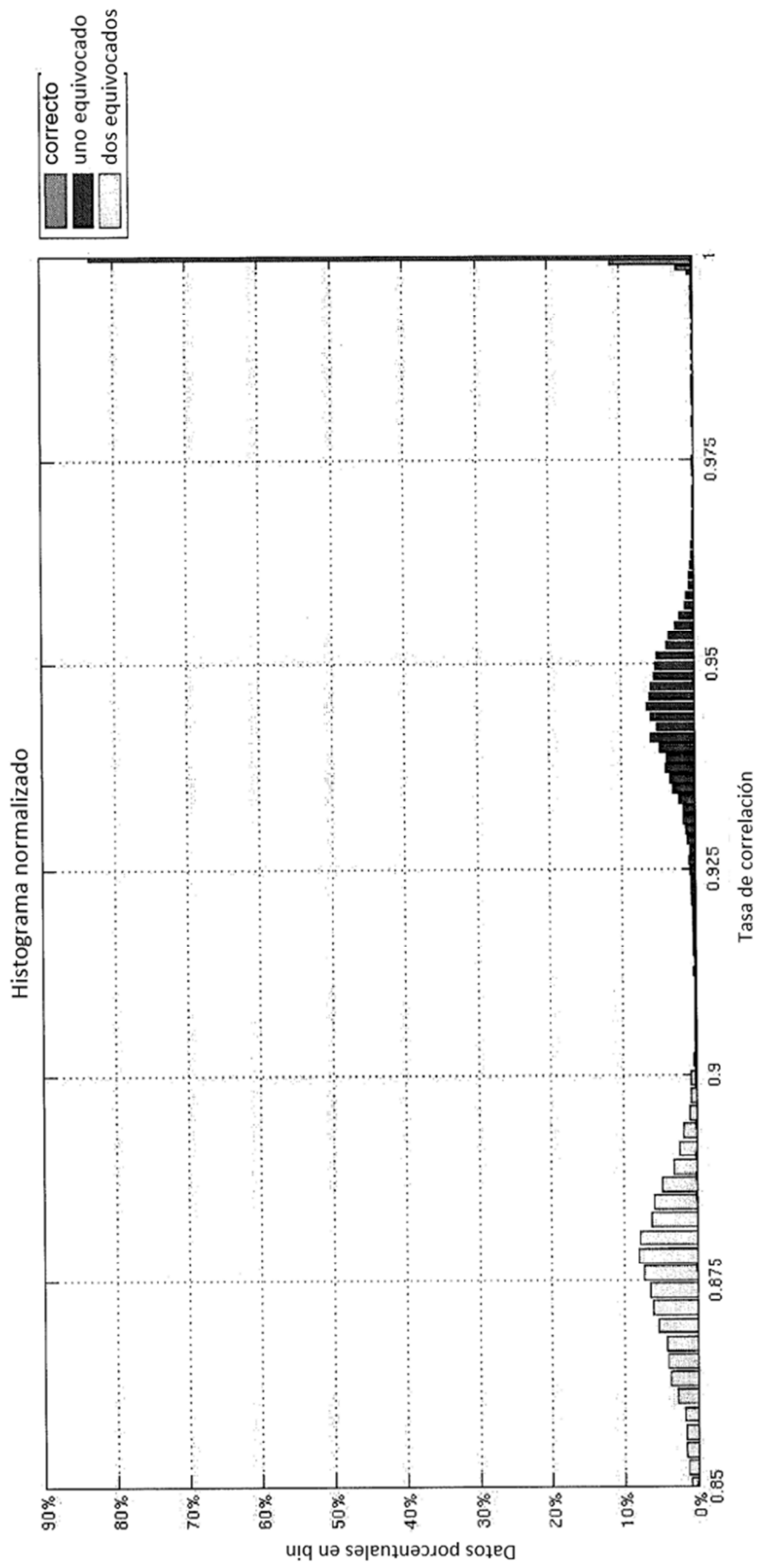


Figura 5

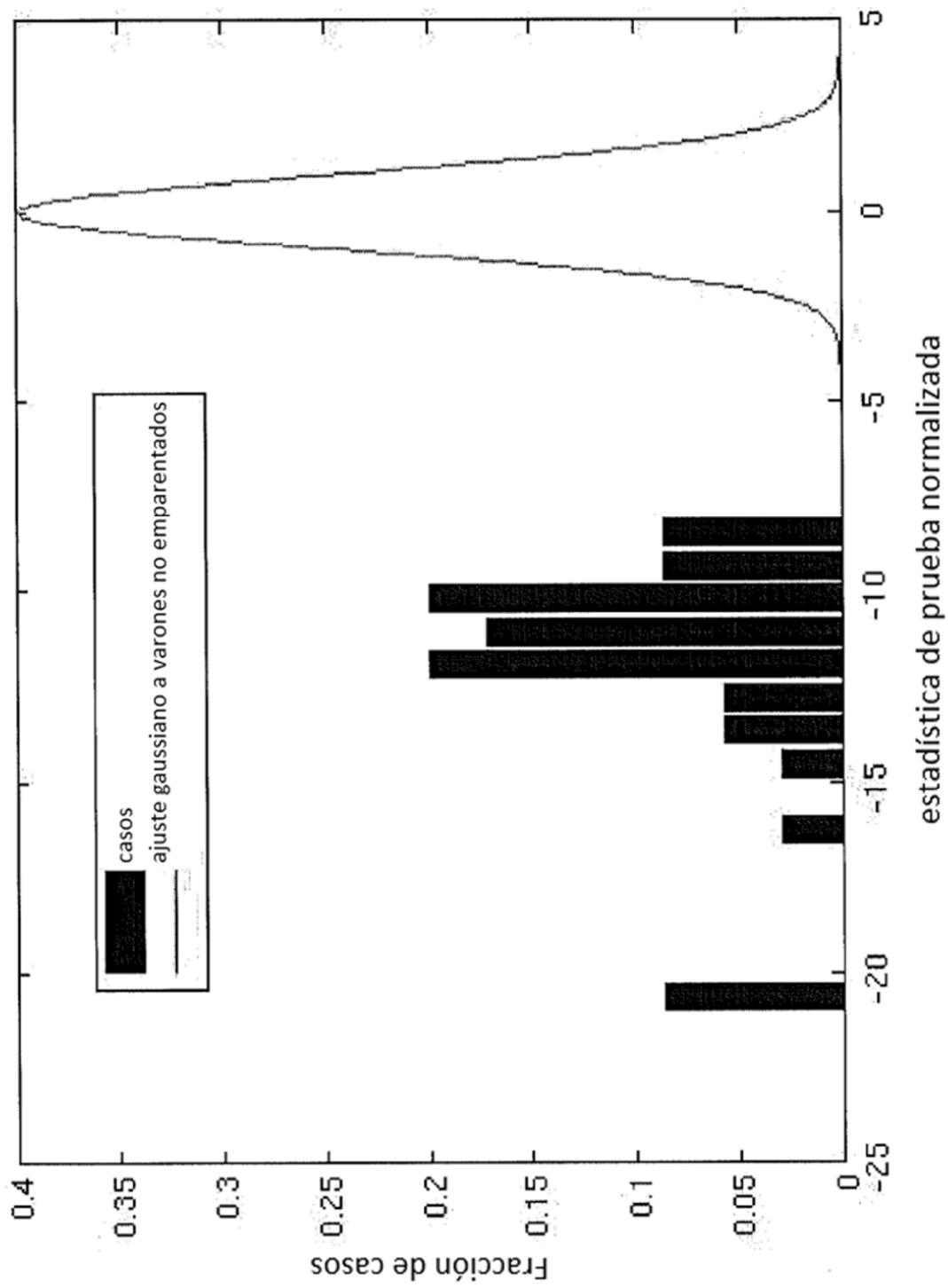


Figura 6

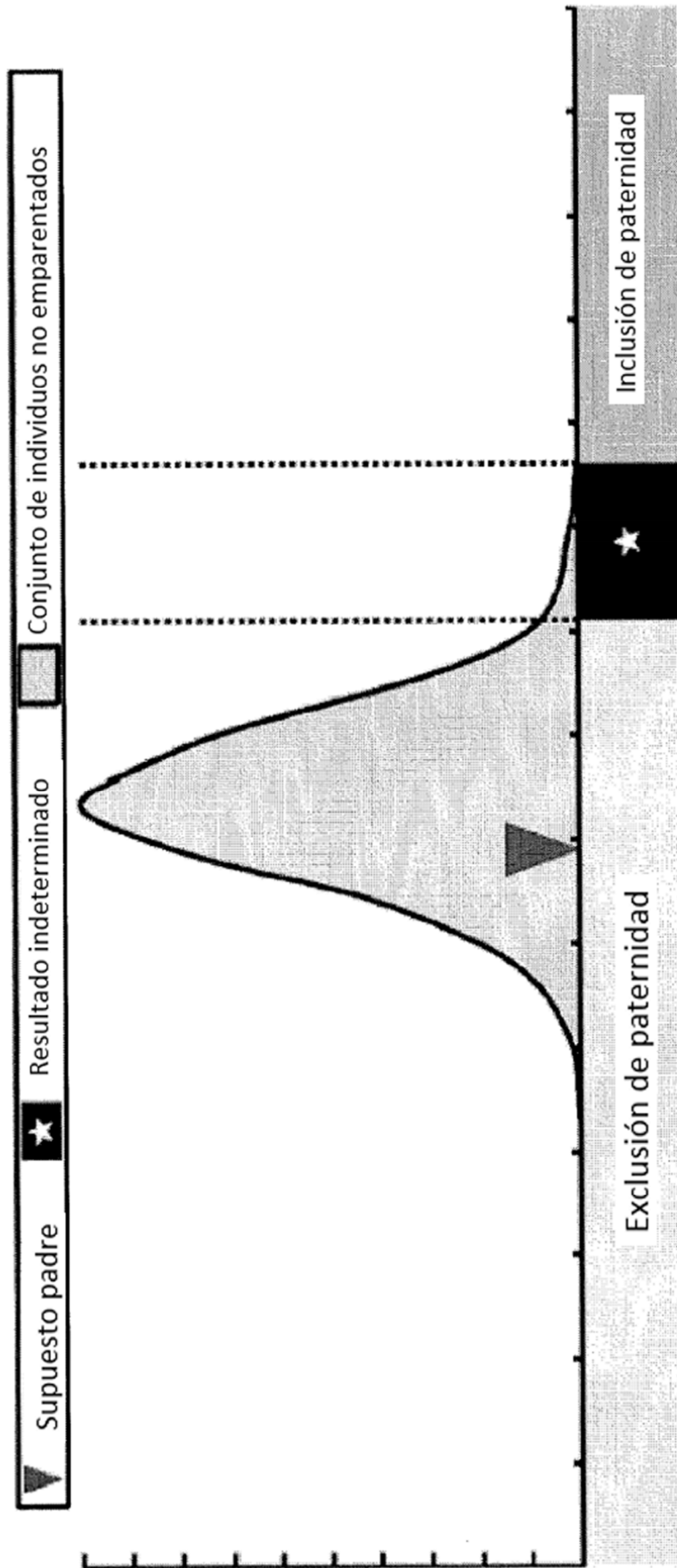


Figura 7

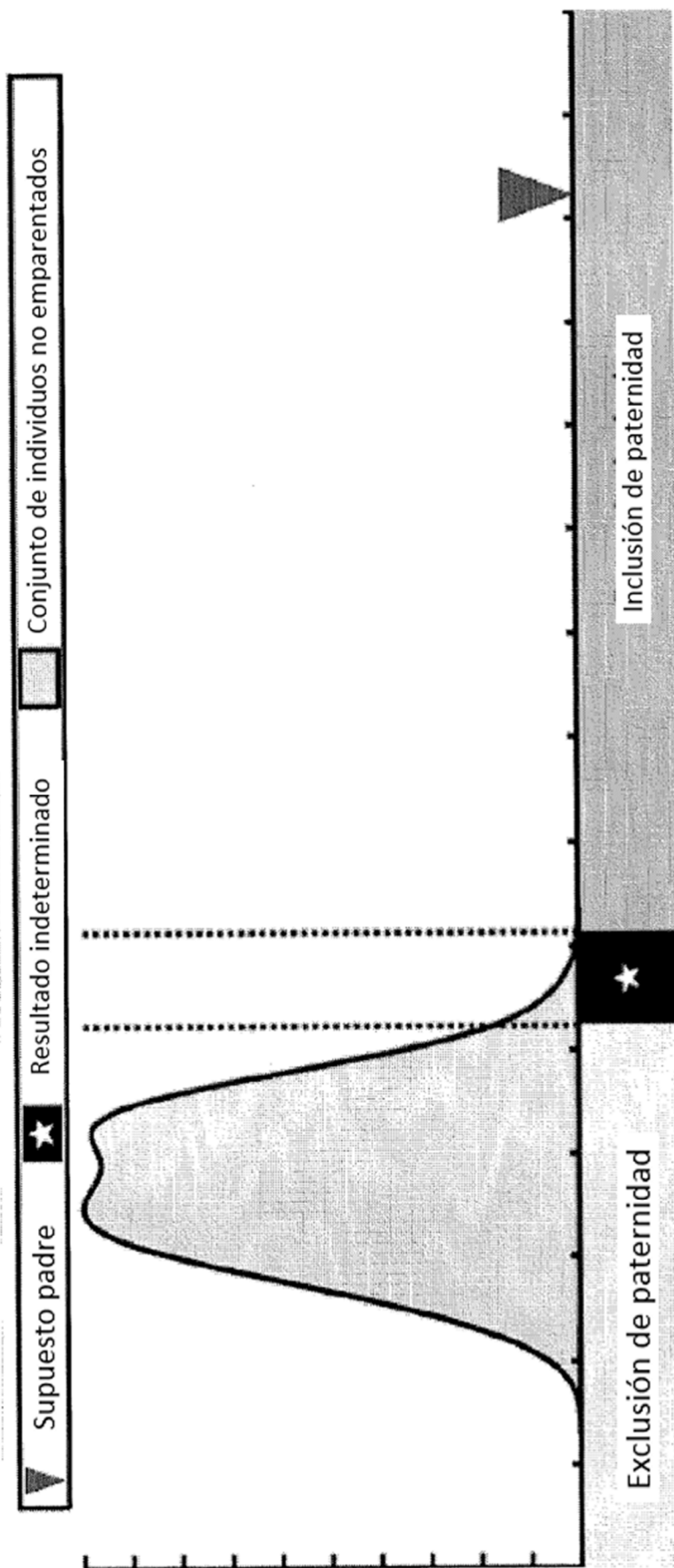


Figura 8



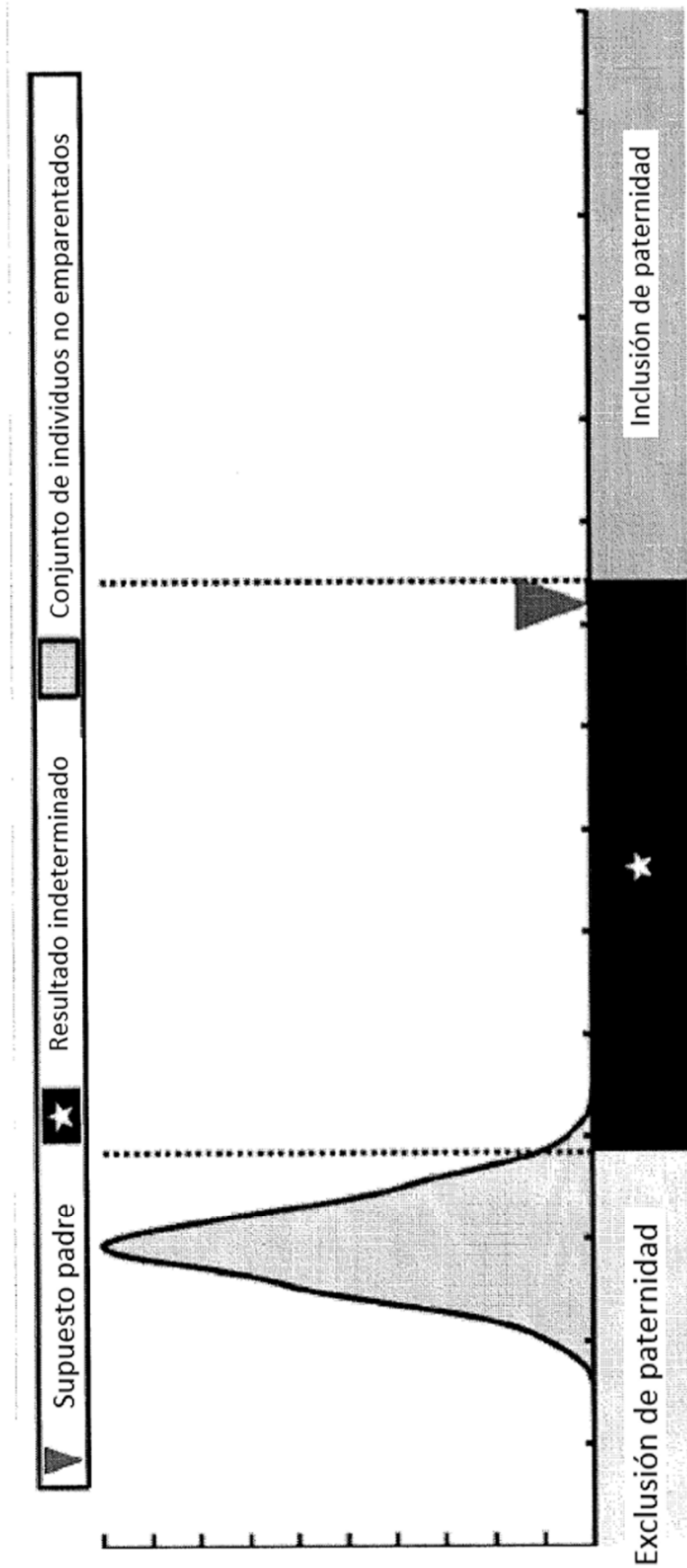


Figura 9