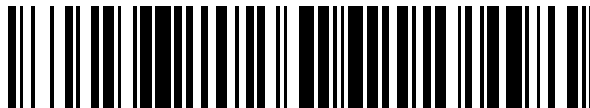


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 640 776**

51 Int. Cl.:

G06F 19/18 (2011.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.09.2010 PCT/US2010/050824**

87 Fecha y número de publicación internacional: **07.04.2011 WO11041485**

96 Fecha de presentación y número de la solicitud europea: **30.09.2010 E 10821214 (3)**

97 Fecha y número de publicación de la concesión europea: **09.08.2017 EP 2473638**

54 Título: **Métodos para denominar de forma no invasiva ploidía prenatal**

30 Prioridad:

30.09.2009 US 277876 P

12.02.2010 US 337931 P

18.05.2010 US 395850 P

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

06.11.2017

73 Titular/es:

NATERA, INC. (100.0%)

201 Industrial Road, Suite 410

San Carlos, CA 94070, US

72 Inventor/es:

RABINOWITZ, MATTHEW;

RYAN, ALLISON;

GEMELOS, GEORGE;

BANJEVIC, MILENA y

DEMKO, ZACHARY

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 640 776 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

Métodos para denominar de forma no invasiva ploidía prenatal

5 Solicitudes relacionadas

La presente solicitud reivindica el beneficio de las siguientes solicitudes de patente provisional de Estados Unidos Serie N° 61/277,876, registrada el 30 de septiembre de 2009; y la serie No. 61/337,931, registrada el 12 de febrero de 2010; la serie No. 61/395,850, registrada el 18 de mayo de 2010.

10

Antecedentes

Normalmente, un ser humano tiene dos conjuntos de 23 cromosomas en cada célula somática, proviniendo cada copia de cada uno de los progenitores. La aneuploidía, estado en el que una célula tiene el número erróneo de cromosomas, es la responsable de un significativo porcentaje de niños nacidos con afecciones genéticas. La detección de anomalías cromosómicas puede servir para identificar a individuos, incluyendo fetos y embriones, con afecciones tales como síndrome de Down, síndrome de Edwards, síndrome de Klinefelters y síndrome de Turner, entre otros. Dado que las anomalías cromosómicas son indeseables generalmente, la detección de dicha anomalía cromosómica en un feto puede servir como base para decidir interrumpir un embarazo.

15

20

El diagnóstico prenatal puede poner en alerta a los médicos y a los progenitores sobre las anomalías en el feto en desarrollo. Algunos métodos de los que se dispone hoy en día, tales como amniocentesis y muestreo de vellosidades coriónicas (MVC), sirven para diagnosticar defectos genéticos con una gran precisión; sin embargo, conllevan el riesgo de aborto espontáneo. Existen otros métodos con los que se puede estimar indirectamente un riesgo de ciertos defectos genéticos de forma no invasiva, por ejemplo, a partir de los niveles hormonales en la sangre materna y/o a partir de los datos de ecografía, sin embargo, su exactitud es mucho más baja. En el documento US2008/0070792A1 se divulga el uso de genotipificación de SNP altamente paralela para diagnóstico fetal. Se ha descubierto recientemente que el ADN fetal desprovisto de células y las células fetales intactas pueden entrar en la circulación de la sangre materna. Esto proporciona la oportunidad de medir directamente información genética sobre el feto, específicamente, el estado de aneuploidía del feto, de una manera no invasiva, por ejemplo, de sangre extraída de la madre.

25

30

Sumario

35

La invención queda definida con las reivindicaciones adjuntas. En el presente documento se divulgan métodos no invasivos para determinar la ploidía prenatal. En una realización de la presente divulgación, se divulgan métodos para determinar el estado de ploidía de un individuo diana, en el que el material genético medido de la diana está contaminado con material genético de la madre, valiéndose de la información conocida de los datos genéticos maternos. Esto se contrapone a los métodos que sirven para determinar el estado de ploidía de un individuo diana a partir de datos genéticos con ruido debido a las mediciones insuficientes; La contaminación en estos datos es aleatoria. Esto también se contrapone a los métodos que sirven para determinar el estado de ploidía de un individuo diana a partir de los datos genéticos que resultan difíciles de interpretar por la contaminación del ADN de individuos no relacionados; la contaminación de los datos es genéticamente aleatoria. En una realización, los métodos descritos en el presente documento sirven para determinar el estado de ploidía de un individuo diana cuando la dificultad de interpretación se debe a la contaminación del ADN de uno de los progenitores; la contaminación de estos datos es idéntica al menos en la mitad de los datos diana y, por lo tanto, es difícil de corregir. Para conseguir este fin, en una realización, en uno de los métodos de la presente invención se utiliza la información conocida del genotipo materno contaminante para crear un modelo de las mediciones genéticas esperadas dada una mezcla del material genético materno y diana, en el que no se conocen los datos genéticos diana a priori. Esta etapa no es necesaria cuando la incertidumbre de los datos genéticos se debe a ruido aleatorio.

40

45

50

55

De acuerdo con los aspectos que se ilustran en el presente documento, se proporciona un método que permite determinar el estado de ploidía de un individuo diana utilizando material genético del individuo diana cuando el material genético del individuo diana está contaminado por otro material genético. En una realización, el individuo diana es un feto, y los datos genéticos del individuo diana comprenden ADN que flota libremente encontrado en la sangre materna, y el material genético contaminante comprende ADN materno que flota libremente también encontrado en la sangre materna. En una realización, el individuo diana es un feto, y los datos genéticos del individuo diana comprenden ADN encontrado en células fetales encontradas en la sangre materna y el material genético contaminante comprende ADN encontrado en células maternas encontradas también en la sangre materna. En una realización, el individuo es un feto y la determinación del estado de ploidía se realiza dentro del contexto de un diagnóstico prenatal no invasivo y se adopta una decisión clínica basada en la determinación del estado de ploidía. En una realización, se utilizan los datos genéticos de uno o ambos progenitores del individuo diana en la determinación del estado de ploidía de la diana. En una realización, los cromosomas de interés incluyen los cromosomas 13, 18, 21, X e Y. En una realización, la determinación se transforma en un informe que puede ser enviado al facultativo sanitario correspondiente. En una realización, la serie de etapas que se han señalado tienen como resultado la transformación de la materia genética de una madre embarazada y el padre en una decisión de

60

65

actuación que tiene como resultado la interrupción o continuación del embarazo. En una realización, la determinación del estado de ploidía se utiliza para adoptar una decisión clínica. En una realización, la decisión clínica puede ser la interrupción del embarazo en los casos en los que se encuentra que el feto tiene una anomalía genética.

5 Si bien la divulgación centra su punto de mira en los datos genéticos de sujetos humanos y, más específicamente, en fetos en desarrollo, así como los individuos emparentados, debe señalarse que los métodos divulgados se aplican a los datos genéticos de toda una gama de organismos y toda una gama de contextos. Las técnicas descritas para obtener la determinación de ploidía son sobre todo pertinentes en el contexto de un diagnóstico prenatal en conjunción con amniocentesis, biopsia de las vellosidades coriónicas, muestreo de tejido fetal y diagnóstico prenatal no invasivo, en el que se aísla una pequeña cantidad de material genético fetal de la sangre materna, por ejemplo, cribado serológico prenatal, la triple prueba y la prueba cuádruple. El uso de este método puede facilitar diagnósticos centrándose en enfermedades hereditarias, predicciones del número de copias de cromosomas, la mayor probabilidad de defectos u anomalías, así como la realización de predicciones de susceptibilidad a varios fenotipos patológicos y no patológicos para individuos que potencien las decisiones clínicas y estilos de vida.

20 En una realización de la presente divulgación, se pueden usar los datos genómicos fetales o embrionarios, con o sin el uso de datos genéticos de individuos emparentados, para detectar si la célula es aneuploide, es decir, en los casos en los que está presente el número erróneo de uno o más cromosomas autosómicos en un individuo, y/o si está presente en el individuo el número erróneo de cromosomas sexuales. Los datos genéticos se pueden utilizar también para detectar disomía uniparental, una afección en la que están presentes dos de un cromosoma determinado, los cuales provienen ambos de uno de los progenitores. Esto se lleva a cabo creando un conjunto de hipótesis sobre los posibles estados del ADN y determinado cuál de las hipótesis tiene la mayor probabilidad de ser cierta dados los datos medidos.

30 En una realización de la presente divulgación, la pequeña cantidad de material genético de un feto que se puede mezclar con el material genético materno se puede transformar por amplificación en una gran cantidad de material genético que codifica datos genéticos similares o idénticos. Los datos genéticos contenidos molecularmente en la cantidad de material genético grande pueden transformarse en datos genéticos en bruto en forma de señales digitales, opcionalmente, almacenadas en la memoria del ordenador, a través de un método de genotipificación. Los datos genéticos en bruto se pueden transformar a través del método PARENTAL SUPPORT™ en denominaciones del número de copias para una o una serie de cromosomas, también almacenados opcionalmente en la memoria del ordenador. La denominación del número de copias se puede transformar en un informe para el médico, que puede actuar entonces según la información del informe.

40 En una realización de la presente divulgación, las mediciones directas del material genético, ampliado o sin ampliar, presente en una pluralidad de loci, puede utilizarse para detectar monosomía, disomía uniparental, trisomía pareada, trisomía no pareada, tetrasomía y otros estados de aneuploidía. Una realización de la presente divulgación aprovecha la ventaja del hecho de que, en ciertas condiciones, el nivel promedio de amplificación y la salida de la señal de la medición es invariable a través de los cromosomas y por tanto, la cantidad promedio de material genético medido en un conjunto de loci vecinos será proporcional al número de cromosomas homólogos presentes y el estado de ploidía puede denominarse de un modo estadísticamente significativo. En otra realización, diferentes alelos tienen perfiles de amplificación característicos estadísticamente diferentes dado un contexto parental determinado y un determinado estado de ploidía; estas diferencias características pueden utilizarse para determinar el estado de ploidía del cromosoma.

50 En una realización de la presente invención, los datos genéticos calculados, sincronizados, reconstruidos y/o determinados del individuo diana y/o de uno o más individuos emparentados pueden utilizarse como datos de entrada para denominar la ploidía según un aspecto de la presente invención.

55 En una realización, un método para determinar un número de copias de un cromosoma de interés en un individuo diana, utilizando mediciones genotípicas obtenidas sobre un material genético del individuo diana, en el que el material genético del individuo diana se mezcla con el material genético de la madre del individuo diana comprende la obtención de datos genotípicos de un conjunto de SNP de los progenitores del individuo diana; la obtención de 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; la creación, en un ordenador, de un conjunto de hipótesis de estado de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada hipótesis dadas las mediciones genéticas de la muestra mixta y los datos genéticos de los progenitores del individuo diana; y la utilización de las probabilidades determinadas de cada hipótesis para determinar el número de copias del cromosoma de interés más probable en el individuo diana.

65 En una realización, el individuo diana es un feto. En una realización, la determinación del número de copias se utiliza para adoptar una decisión clínica. En una realización, el individuo diana es un feto y la decisión clínica es la interrupción del embarazo cuando se encuentra que el feto tiene una anomalía genética, o no interrumpir el embarazo cuando no se encuentra que el feto tenga una anomalía genética. En una realización, el conjunto de SNP

comprende una pluralidad de SNP del cromosoma de interés y una pluralidad de SNP de al menos un cromosoma que se espera que sea disómico en el individuo diana.

5 En una realización, la etapa de determinar en el ordenador la probabilidad de cada una de las hipótesis comprende el uso de los datos genotípicos de los progenitores para determinar los contextos parentales para cada uno de los SNP; la agrupación de las mediciones genotípicas de la muestra mixta en los contextos parentales; el uso de las mediciones genotípicas agrupadas de al menos un cromosoma que se espera que sea disómico para determinar una respuesta de plataforma; el uso de las mediciones genotípicas agrupadas de al menos un cromosoma que se espera que sea disómico para determinar la relación entre el ADN fetal y el materno en la muestra mixta, el uso de la respuesta de plataforma determinada y la relación determinada para predecir una distribución esperada de las mediciones de SNP para cada conjunto de SNP en cada contexto parental según cada una de las hipótesis; y el cálculo de las probabilidades de que cada una de las hipótesis sea cierta dada la respuesta de la plataforma y dada la relación y dadas las mediciones genotípicas agrupadas de la muestra mixta y dadas las distribuciones esperadas previstas para cada contexto parental para cada hipótesis.

15 En una realización, el cromosoma de interés se selecciona del grupo que consiste en cromosoma 13, cromosoma 18, cromosoma 21, el cromosoma X, el cromosoma Y y combinaciones de los mismos. En una realización, el método se utiliza para determinar el número de copias de una serie de cromosomas en el individuo diana, en el que el número se selecciona del grupo que consiste en uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte, veintiuno, veintidós y veintitrés.

20 En una realización, la muestra mixta es sangre materna, plasma materno u otra sustancia tomada de una madre embarazada. En una realización, el material genético del individuo diana es ADN que flota libremente encontrado en la sangre o el suero materno. En una realización, el material genético del individuo diana es ADN nuclear encontrado en una o más células del individuo diana. En una realización, se computa la fiabilidad de la determinación del número de copias de cromosoma. En una realización, se determina la relación entre el ADN fetal y el materno en la muestra mixta para cromosomas individuales.

30 En una realización, la etapa de obtención de datos genotípicos y/ la etapa de obtención de mediciones genotípicas se realiza midiendo el material genético aplicando técnicas seleccionadas del grupo que consiste en sondas padlock, sondas circularizables, micromatrices de genotipificación, ensayos de genotipificación de SNP, micromatrices basadas en chip, micromatrices basadas en perlas, otras micromatrices de SNP, otros métodos de genotipificación, secuenciación de ADN de Sanger, pirosecuenciación, secuenciación de alto rendimiento, secuenciación de terminador colorante reversible, secuenciación por ligación, secuenciación por hibridación, otros métodos de secuenciación de ADN, otras plataformas de genotipificación de alto rendimiento, hibridación *in situ* fluorescente (FISH), hibridación genómica comparada (CGH), CGH de matriz y múltiples combinaciones de los mismos. En una realización, la etapa de medición del material genético se realiza en un material genético que se amplifica antes de medirlo utilizando una técnica que se selecciona del grupo que consiste en reacción en cadena de la polimerasa (PCR), PCR mediada por ligando, PCR de cebador oligonucleótido degenerativa, amplificación por desplazamiento múltiple (MDA), PCR específica de alelo, técnicas de amplificación específica de alelo, amplificación de puente, sondas padlock, sondas circularizables y combinaciones de ellos.

45 En una realización, la etapa de determinación del número de copias del cromosoma de interés se lleva a cabo con el fin de rastrear un estado cromosómico, seleccionándose el estado cromosómico del grupo que consiste en nulisomía, monosomía, disomía, disomía uniparental, euploidía, trisomía, trisomía apareada, trisomía no apareada, trisomía materna, trisomía paterna, tetrasomía, tetrasomía apareada, tetrasomía no apareada, otras aneuploidías, translocación desequilibrada, translocación equilibrada, recombinación, delección, inserción, mosaicismo y combinaciones de los mismos.

50 En una realización, el método se utiliza con el fin de una prueba de paternidad.

55 En una realización, un método para determinar un número de copias de un cromosoma de interés en un individuo diana, utilizando mediciones genotípicas obtenidas sobre un material genético del individuo diana, en el que el material genético del individuo diana se mezcla con el material genético de la madre del individuo diana comprende la obtención de datos genotípicos para un conjunto de SNP de la madre del individuo diana; la obtención de mediciones genotípicas para el conjunto de SNP en una muestra mixta que comprende ADN del individuo diana y también de ADN de la madre del individuo diana; la creación en un ordenador, de un conjunto de hipótesis de estados de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las mediciones genéticas de la muestra mixta y los datos genéticos de los progenitores del individuo diana; y uso de las probabilidades determinadas de cada hipótesis para determinar el número de copias del cromosoma de interés más probable en el individuo diana.

65 Las personas especializadas en la técnica reconocerán, dado el beneficio de la presente divulgación, que se pueden implementar varios de los aspectos y realizaciones de la presente divulgación en combinación o por separado.

Breve descripción de los dibujos

Las realizaciones divulgadas en el presente documento se explicarán mejor haciendo referencia a los dibujos adjuntos, en los que se hace referencia a las estructuras iguales con números iguales a través de diferentes vistas. Los dibujos presentados no son a escala necesariamente, sino que se pone el énfasis por lo general en ilustrar los principios de las realizaciones descritas en el presente documento.

La Figura 1 presenta un ajuste de modelo tanto para los canales X (gráfico de la izquierda) como Y (gráfico de la derecha) en una muestra con un 40 por ciento de ADN del individuo diana.

La Figura 2 muestra los componentes del vector de medición en comparación con las predicciones según tres hipótesis. Este gráfico es del cromosoma 16 de una muestra constituida por 40 por ciento de ADN diana y 60 por ciento de ADN de la madre. La hipótesis verdadera es H110.

La Figura 3 presenta los componentes del vector de medición en comparación con las predicciones según tres hipótesis. La Figura 3 es del cromosoma 21 de la misma muestra que la Figura 2 y la hipótesis correcta es H210.

Si bien los dibujos identificados divulgan realizaciones del presente documento, también se contemplan otras realizaciones, tal como se señala en la explicación. La presente divulgación presenta realizaciones ilustrativas de forma representativa no exhaustiva.

Descripción detallada

La invención se define en las reivindicaciones adjuntas. En una realización de la presente divulgación, el estado de ploidía puede determinarse para uno, algunos o todos los cromosomas en el individuo. En una realización de la invención, el material genético del individuo diana se utiliza para hacer la determinación de la ploidía y en los casos en los que el material genético del individuo diana está contaminado con el material genético de la madre del individuo diana. En una realización de la divulgación, los datos genéticos de uno o ambos progenitores del individuo diana, incluyendo opcionalmente datos genéticos de otros parientes del individuo diana se utiliza la determinación de la ploidía.

La denominación del número de copias es el concepto que se refiere a la determinación del número y la identidad de cromosomas en un individuo, ya sea por célula o por volumen. En una realización de la divulgación, puede utilizarse la cantidad de material genético contenida en una sola célula, un pequeño grupo de células, una muestra de ADN como representante para el número de cromosomas en el individuo diana. La presente divulgación permite la determinación de aneuploidía del material genético contenido en una pequeña muestra de células, o una pequeña muestra de ADN, siempre y cuando el genoma de al menos uno o ambos progenitores esté disponible. Algunos aspectos de la presente divulgación emplean el concepto de contexto parental, describiendo el contexto parental para un SNP dado, el posible conjunto de alelos que puede heredar un niño de los progenitores. Para cada conjunto de SNP que pertenecen a un contexto parental dado, se espera una distribución estadística específica de las mediciones de SNP y que la distribución varíe dependiendo del contexto parental y del estado de ploidía del segmento de cromosoma en el que se encuentra el SNP. Al analizar las distribuciones reales de los SNP en diferentes contextos parentales y comparándolos con la distribución esperada de dichos SNP para diferentes hipótesis de estados de ploidía es posible calcular qué estado de ploidía es más probable que sea correcto. Esto puede ser particularmente útil en el caso de diagnósticos prenatales, en los que está disponible una cantidad limitada de ADN y en los casos en los que la determinación del estado de ploidía de una diana, como por ejemplo un feto, tiene un alto impacto clínico.

Varias técnicas informáticas cuyo uso puede ser apropiado con conjunción con la presente divulgación se describen en las tres referencias enumeradas a continuación. Publicación No. 2007/0184467, publicada el 9 de agosto de 2007, publicación estadounidense No. 2008/0243398, publicada el 2 de octubre de 2008 y publicación PCT No. WO/2010/017214, publicada el 11 de febrero de 2010. En el presente documento se hace referencia a dichas referencias como Rabinowitz 2006, 2008 y 2009, respectivamente, y se puede hacer referencia a los métodos descritos en dichas referencias, junto con los métodos descritos en la presente divulgación, de forma colectiva como PARENTAL SUPPORT™.

Las mediciones de ADN, ya se obtengan a través de técnicas de secuenciación, matrices de genotipificación o cualquier otra técnica, contienen un grado de error. Hay varios factores que pueden afectar la fiabilidad relativa en una en una medición de ADN dada entre los que se incluyen el método de amplificación, la tecnología utilizada para medir el ADN, el protocolo utilizado, la cantidad de ADN utilizada, la integridad del ADN utilizado, el operador y la frescura de los reactivos, por nombrar solo algunos. Una forma de aumentar la precisión de las mediciones consiste en utilizar técnicas con soporte informático para deducir el estado genético correcto del ADN en la diana sobre la base de la información conocida del estado genético de individuos emparentados. Dado que es de esperar que los individuos emparentados compartan cierto aspecto de su estado genético, cuando los datos genéticos de una pluralidad de individuos emparentados se considera en conjunto, es posible identificar errores probables y omisiones en las mediciones y aumentar la precisión de la información conocida de los estados genéticos de todos los individuos emparentados. Asimismo, se puede computar la fiabilidad para cada denominación realizada.

Para los fines de la presente divulgación, un soporte legible por ordenador es un soporte que almacena los datos del ordenador en una forma legible por máquina. A modo de ejemplo y sin limitación, un soporte legible por ordenador puede comprender un soporte de almacenamiento en ordenador así como soportes, métodos o señales de comunicación. El soporte de almacenamiento en ordenador se denomina también memoria de ordenador, incluye soportes volátiles y no volátiles, extraíbles y no extraíbles implementados en cualquier método o tecnología para almacenar información, como por ejemplo instrucciones legibles por ordenador, estructuras de datos, módulos de programa y otros datos. Entre los soportes de almacenamiento en ordenador se incluyen, sin limitarse a ellos RAM, ROM, EPROM, EEPROM, memoria instantánea u otras tecnologías de memoria en estado sólido; CD-ROM, DVD, u otros medios de almacenamiento ópticos; casetes, cintas, discos y otros dispositivos de almacenamiento magnéticos; o cualquier otro medio que se pueda utilizar para almacenar tangiblemente la información deseada y al que se pueda acceder a través del ordenador.

Se pueden realizar varias implementaciones de los sistemas y técnicas descritos en el presente documento en un circuito electrónico digital, circuito integrado, circuitos ASIC especialmente diseñados (circuitos integrados de aplicación específica), hardware informático, firmware, software y/o combinaciones de los mismos. Estas diversas implementaciones pueden incluir uno o más programas informáticos que se pueden ejecutar y/o interpretar sobre un sistema programable que incluya al menos un procesador programable, que puede tener un fin especial o general, acoplado para recibir datos e instrucciones y que sirva también para transmitir datos e instrucciones a un sistema de almacenamiento, al menos un dispositivo de entrada y al menos un dispositivo de salida. Dichos programas informáticos (también conocidos como programas, software, aplicaciones de software o código) pueden incluir instrucciones de máquina para un procesador programable y se pueden implementar en cualquier forma de lenguaje de programación incluyendo lenguajes de programación orientados a un objeto y/o un procedimiento de alto nivel y/o lenguajes de montaje/máquina. Un programa informático se puede desplegar de cualquier forma, incluyendo un programa único o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Se puede desplegar un programa informático para ejecutarlo o interpretarlo en un ordenador o en varios ordenadores en un emplazamiento, o se puede distribuir a través de varios emplazamientos e interconectarse mediante una red de comunicación.

Definiciones

SNP (Polimorfismo de un único nucleótido) se refiere a un único nucleótido que puede diferir entre los genomas de dos miembros de la misma especie. El uso del término no deberá implicar ningún límite en cuanto a la frecuencia en la que aparece cada variante. El término SNP puede incluir otras variaciones alélicas que pueden aparecer en una serie de nucleótidos.

Denominar un SNP se refiere al acto de adoptar una decisión sobre el estado verdadero de un par de bases en particular, teniendo en cuenta la evidencia directa e indirecta.

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.

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

Denominar un alelo se refiere al acto de determinar el estado genético de un locus de ADN en particular. Esto puede implicar la denominación de un SNP, una pluralidad de SNP o determinar si está presente o no una inserción o delección en ese locus, o determinar el número de inserciones que puede estar presente en ese locus, o determinar si está presente o no otra variante genética, como por ejemplo repeticiones en tándem simples (STR) o cuántas están presentes en ese locus.

Locus se refiere a una localización específica del gen o la secuencia de ADN en un cromosoma.

Denominación de ploidía, también “denominación del número de copias de cromosoma”, “denominación del número de copias”, “determinación del estado de ploidía,” o determinación del número de copias, es el acto de determinar la cantidad y posiblemente también la identidad cromosómica de uno o más cromosomas presentes en una célula.

Denominar una hipótesis, se refiere a determinar que hipótesis tiene mayor probabilidad de ser verdadera. El acto de denominar puede consistir en el punto en el que se adopta la decisión sobre qué hipótesis se destacará como denominación.

Fiabilidad se refiere a la probabilidad estadística de que el SNP, alelo, conjunto de alelos denominados o número de copias de cromosomas o copias de segmentos de cromosoma determinado represente correctamente el estado genético real del individuo.

Aneuploidía se refiere al estado en el que está presente un número erróneo de cromosomas en una célula. En el caso de una célula humana somática, puede referirse al caso en el que una célula no contiene 22 pares de cromosomas autosómicos y un par de cromosomas del sexo. En el caso de un gameto humano, puede referirse al

caso en el que la célula no contiene uno de los 23 cromosomas. Al hacer referencia a un solo cromosoma autosómico, puede referirse al caso en el que están presentes más o menos de dos cromosomas homólogos. Al hacer referencia al cromosoma del sexo, puede referirse al caso en el que están presentes más o menos de dos del cromosoma X o el cromosoma Y, o exactamente dos cromosomas Y.

5 *Estado de ploidía* es la cantidad e identidad cromosómica de uno o más cromosomas en una célula. Puede referirse al número total y la identidad de cada cromosoma encontrado normalmente en cada célula de un individuo dado. Puede referirse al número e identidad de cromosoma(s) para un número de cromosomas en particular para un individuo determinado.

10 *Identidad cromosómica* se refiere al número de cromosomas referente. Los humanos normalmente tienen 22 tipos de cromosomas autosómicos numerados y dos tipos de cromosomas del sexo. Puede referirse también el origen parental del cromosoma. También puede referirse a un cromosoma específico heredado del progenitor, es decir, el cromosoma que ha heredado el/la progenitor/a de su padre, o el cromosoma que ha heredado el/la progenitor/a de su madre. También puede referirse a otros rasgos de identificación de un cromosoma. La identidad de un cromosoma puede referirse a la identidad real de un cromosoma en particular o las identidades del cromosoma de un número de cromosomas en particular, en cada célula de un individuo en particular. Por ejemplo, la identidad cromosómica podría ser: "cromosoma 21" o puede referirse a un cromosoma 21 en particular con un estado genético en particular, es decir, por ejemplo, "heredado de la madre y homólogo pero no idéntico a otros dos cromosomas 21 encontrados en un individuo de sexo femenino en particular con síndrome de Down.

20 *Número de cromosoma* se refiere al número cardinal asignado comúnmente a un cromosoma determinado, teniendo el ser humano 22 pares de cromosomas autosómicos y un par de cromosomas del sexo, para un total de 23. El número de cromosoma puede ser un número de 1 a 23 y en el caso del cromosoma 23 se puede hacer referencia a él como X o Y. Puede hacer referencia a una clase de cromosomas, por ejemplo, se puede encontrar que un niño con síndrome de Down tiene tres cromosomas 21.

25 *El estado del material genético* o simplemente "estado genético" se refiere a la identidad real de un conjunto de SNP en el ADN. Puede referirse a los haplotipos sincronizados del material genético y puede referirse a la secuencia de ADN, incluyendo inserciones, deleciones, repeticiones y mutaciones en un individuo. Puede referirse también al estado de ploidía real de uno o más cromosomas, segmentos cromosómicos o conjunto de segmentos cromosómicos en un individuo.

30 *Anomalía genética* se refiere a un estado genético que está muy correlacionado con una anomalía fenotípica. Aneuploidía es un ejemplo de anomalía genética. Un estado genético que tiene como resultado la muerte del feto o un bebé es una anomalía fenotípica.

35 *Mediciones genotípicas* (o "mediciones genéticas") es un tipo de datos genotípicos, como por ejemplo datos numéricos, digitales, representaciones pictóricas o figurativas de datos genotípicos que se obtienen utilizando técnicas de genotipificación para establecer ciertas secuencias de pares de base y/o identidades, cantidades u otras características del material genético, principalmente, ADN. Las mediciones genéticas pueden contener errores u omisiones.

40 *Datos genéticos* se refiere a los datos que describen un estado genético. Los datos genéticos pueden adoptar la forma de mediciones genéticas, se pueden codificar de forma analógica o digital, se pueden codificar con un ordenador o pueden adoptar la forma de una secuencia genética molecular.

45 *Medir material genético* se refiere al acto de transformar datos genéticos a partir de una manifestación física, por ejemplo, una secuencia de pares de base específica, en una representación figurativa de los datos genéticos, como por ejemplo la representación de la secuencia genética almacenada digitalmente en un ordenador.

50 *Mosaicismo* se refiere a un conjunto de células en un embrión, u otro individuo, que son heterogéneas en lo que se refiere a su estado de ploidía.

55 *Cromosomas homólogos* son cromosomas que contienen el mismo conjunto de genes que se pueden aparear normalmente durante la meiosis.

60 *Cromosomas idénticos* son cromosomas que contienen el mismo conjunto de genes y para cada gen pueden tener el mismo conjunto de alelos que son idénticos, o casi idénticos.

65 *Pérdida de alelo* o "ADO" se refiere a la situación en la que no se detecta uno de los pares de base en un conjunto de pares de base de cromosomas homólogos en un alelo determinado. Se puede hacer referencia a ADO como LDO.

Pérdida de locus o "LDO" se refiere a la situación en la que no se detectan ambos pares de base en un conjunto de pares de base de cromosomas homólogos en un alelo determinado.

Homocigótico se refiere a tener alelos o SNP similares en los correspondientes loci cromosómicos.

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

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

Segmento de un cromosoma se refiere a una sección de un cromosoma que puede oscilar en tamaño desde un par de base hasta un cromosoma entero.

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

Copias se refiere al número de copias de un segmento de cromosomas y puede referirse a copias idénticas, o puede referirse a copias homólogas no idénticas de un segmento de cromosoma, en el que diferentes copias del segmento de cromosoma contienen un conjunto de loci sustancialmente similar y cuando uno o más de los alelos son diferentes. Adviértase que, en algunos casos de aneuploidía, como en el error de copia M2, es posible tener algunas copias del segmento de cromosoma dado que son idénticas, así como algunas copias del mismo segmento de cromosoma que no son idénticas.

Haplotipo es una combinación de alelos en múltiples loci que se transmiten juntos en el mismo cromosoma. Haplotipo puede referirse a tan solo dos loci o a un cromosoma entero dependiendo del número de eventos de recombinación que tengan lugar entre un conjunto dado de loci. Haplotipo puede referirse también a un conjunto de polimorfismos de un solo nucleótido (SNP) en un solo cromátido que están estadísticamente asociados. Un haplotipo puede referirse también a una "cadena", haciendo referencia al hecho de que los haplotipos están físicamente conectados en una cadena de ADN.

Datos haplotípicos denominado también "datos sincronizados" o "datos genéticos ordenados" se refiere a los datos de un solo cromosoma de un genoma diploide o poliploide, es decir, la copia de un cromosoma materno o paterno segregado en un genoma diploide.

Sincronización se refiere al acto de determinar los datos genéticos haplotípicos de un individuo dados los datos genéticos diploides (o poliploidía) desordenados. Puede referirse al acto de determinar cuál de los dos genes en un alelo, para un conjunto de alelos encontrados en un cromosoma, está asociado con cada uno de los dos cromosomas homólogos en un individuo.

Datos sincronizados se refiere a los datos genéticos en los que se ha determinado el haplotipo.

Datos genéticos "dentro", "de", "en", "de" "sobre" un individuo, (también "*datos genotípicos*") se refiere a los datos que describen los aspectos del genoma de un individuo. Puede referirse a uno o a un conjunto de loci, secuencias parciales o enteras, cromosomas parciales o enteros o el genoma entero.

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

Hipótesis del número de copias, también "hipótesis del estado de ploidía" se refiere a la hipótesis que concierne a cuántas copias de un cromosoma en particular hay en un individuo por célula. Puede referirse a una hipótesis que concierne a la identidad de cada uno de los cromosomas, incluyendo el progenitor de origen de cada cromosoma, y cuál de los dos cromosomas del progenitor está presente en el individuo. Puede referirse también a una hipótesis que concierne a cuál cromosoma o segmentos de cromosoma, si los hay, de cada uno de los individuos emparentados corresponde genéticamente a un cromosoma dado de un individuo.

Individuo diana se refiere al individuo cuyo estado genético está siendo determinado. En un contexto, solamente está disponible una cantidad limitada de ADN del individuo diana. En un contexto, el individuo diana es un feto. En algunas realizaciones, puede haber más de un individuo diana. En algunas realizaciones, cada niño, embrión, feto o esperma proveniente de un par de progenitores puede considerarse individuo diana.

Individuo emparentado se refiere a cualquier individuo que está genéticamente relacionado y que por tanto comparte bloques de haplotipo con el individuo diana. En un contexto, el individuo emparentado puede ser un progenitor genético del individuo diana, o cualquier material genético derivado de uno de los progenitores, como esperma, un cuerpo polar, un embrión, un feto o un niño. Puede referirse también a un hermano o a un abuelo.

Progenitor se refiere a la madre o padre genéticos de un individuo. Un individuo tendrá típicamente exactamente dos progenitores, una madre y un padre. Un progenitor puede considerarse como un individuo.

Contexto parental, (también "contexto"), se refiere al estado genético de un SNP dado, en cada uno de los cromosomas homólogos correspondientes para cada uno de los dos progenitores de la diana.

Aislamiento se refiere a una separación física del material genético diana desde otro material genético o material biológico contaminante. Puede referirse también a un aislamiento parcial en el que se separa la diana de aislamiento de parte o la mayoría, pero no todo, del material contaminante. Por ejemplo, el aislamiento de ADN fetal puede referirse al aislamiento de una fracción de ADN fetal que está preferentemente enriquecida en ADN fetal en comparación con la muestra original.

Decisión clínica se refiere a cualquier decisión para emprender o no emprender una acción, que tiene un resultado que afecta a la salud o supervivencia de un individuo. En el contexto de un diagnóstico prenatal, una decisión clínica puede referirse a una decisión de abortar o no abortar un feto. Una decisión clínica puede referirse a una decisión de llevar a cabo posteriores pruebas o a emprender acciones mitigantes.

Respuesta de plataforma se refiere a la caracterización matemática de características de entrada/salida de una plataforma de medición genética y puede utilizarse como medida de las diferencias de medición estadísticamente predecibles.

Método con soporte informático 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 por interferencia estadística del estado más probable, en lugar de medir físicamente el estado directamente. En una realización de la presente divulgación, la técnica de soporte informático puede ser una de las descritas en la presente patente. En una realización de la presente divulgación puede ser PARENTAL SUPPORT™.

Intensidad de canal se refiere a la concentración del fluorescente u otra señal asociada con un alelo, par de base u otro marcador genético dado que se produce a partir de un método utilizado para medir datos genéticos. Puede referirse a un conjunto de datos salida de un dispositivo para medir ADN. En una realización, puede referirse al conjunto de datos de salida de una matriz de genotipificación.

Contexto parental

El contexto parental se refiere al estado genético de un SNP dado, en cada uno de los dos cromosomas correspondientes para cada uno de los progenitores de la diana. Adviértase 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 progenitores. El contexto parental para un SNP dado puede consistir en cuatro pares de base, dos paternos y dos maternos; pueden ser iguales o diferentes entre sí. En la presente divulgación, se puede escribir como " $m_1m_2|f_1f_2$ ", en la que m_1 y m_2 son el estado genético de un SNP dado en los dos cromosomas maternos y f_1 y f_2 son el estado genético del SNP dado en los dos cromosomas paternos. En algunas realizaciones, el contexto parental puede escribirse como " $f_1f_2|m_1m_2$ ". Adviértase que los subíndices "1" y "2" se refieren al genotipo, en el alelo dado, del primero y segundo cromosoma; ha de advertirse asimismo que la selección de qué cromosoma se etiqueta "1" y cuál se marca "2" es arbitraria.

Debe advertirse que en la presente divulgación, se suele utilizar A y B para representar se forma genérica las entidades de los pares de base; A o B podrían representar perfectamente igual C (citosina), G (guanina), A (adenina) o T (timina). Por ejemplo, si en un alelo dado, el genotipo de la madre era T en un cromosoma y G en el cromosoma homólogo, y el genotipo del padre en el alelo es G en ambos cromosomas homólogos, se puede decir que el alelo del individuo diana tiene el contexto parental AB|BB; en algunos contextos, puede ser igualmente correcto decir que el alelo del individuo diana tiene el contexto parental AB|AA o BA|AA. Adviértase, en teoría, pueden aparecer en un alelo dado cualquiera de los cuatro alelos posibles y, por tanto, es posible por ejemplo que la madre tenga un genotipo AT y el padre tenga un genotipo GC en un alelo dado. Sin embargo, los datos empíricos indican que en la mayoría de los casos se observan solamente dos de los cuatro pares de base posibles en un alelo dado. En la presente divulgación, la explicación da por supuesto que solo se observaran dos posibles pares de base en un alelo dado, aunque podrían modificarse las realizaciones divulgadas en el presente documento para tener en cuenta los casos en los que no se sostiene esta suposición.

Un "contexto parental" puede referirse a un conjunto o subconjunto de SNP diana que tienen el mismo contexto parental. Por ejemplo, si se trata de medir 1000 alelos en un cromosoma dado en un individuo diana, entonces el contexto AA|BB podría referirse al conjunto de todos los alelos del grupo de 1000 alelos en los que el genotipo de la madre de la diana era homocigótico en el SNP, y el genotipo del padre en la diana es homocigótico, pero en los que el genotipo materno y el genotipo paterno son diferentes en ese locus. Si los datos parentales no están sincronizados y por tanto $AB = BA$, entonces hay nueve contextos parentales posibles: AA|AA, AA|AB, AA|BB, AB|AA, AB|AB, AB|BB, BB|AA, BB|AB, y BB|BB. Si los datos parentales están sincronizados y, por tanto, $AB \neq BA$, entonces hay dieciséis contextos parentales 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. Asimismo es posible que los datos genéticos de uno de los progenitores esté sincronizado, mientras que los datos genéticos del otro progenitor no estén sincronizados, en cuyo caso habría doce contextos parentales. Cada alelo de SNP en un cromosoma, excluyendo algunos SNP en los cromosomas del sexo, tiene uno de estos contextos parentales. Adviértase que algunos de estos contextos pueden comportarse igual que otros contextos y es posible reunir dichos contextos; esto podría ser funcionalmente

equivalente al uso de número total de contextos. Alternativamente, se podría optar por ignorar ciertos contextos para los fines de un análisis.

Una vez que se han determinado los contextos parentales, los SNP de cada contexto parental se pueden agrupar, de modo que las mediciones de SNP de la muestra genética diana puedan tratarse estadísticamente, como un grupo y compararse con el comportamiento esperado para varias hipótesis. Agrupamiento de SNP por contexto se refiere simplemente crear subconjuntos de SNP que se diferencian según el contexto parental, pudiéndose tratar cada subconjunto en bloque. El agrupamiento de SNP es beneficioso ya que el comportamiento en bloque esperado de un conjunto de SNP depende del contexto parental.

El concepto de contextos parentales puede ser útil en el contexto de la determinación de ploidía de copia. Cuando se genotifican, es de esperar que los SNP dentro de un primer contexto parental se comporten estadísticamente de forma similar cuando se miden para un estado de ploidía dado. En cambio, se puede esperar que algunos conjuntos de SNP de un segundo contexto parental se comporten de forma estadísticamente diferente a la de los del primer contexto parental en ciertas circunstancias, como por ejemplo ciertos estados de ploidía y la diferencia de comportamiento puede ser característica de un estado de ploidía, o conjunto, en particular. Existen muchas técnicas estadísticas que podrían utilizarse para analizar las respuestas medidas en varios loci dentro de varios contextos parentales.

Hipótesis

Una hipótesis puede referirse a un posible estado genético. Puede referirse a un posible estado de ploidía. Un conjunto de hipótesis se refiere a un conjunto de posibles estados genéticos. En algunas realizaciones, puede diseñarse un conjunto de hipótesis de modo que una hipótesis de un conjunto se corresponda con el estado genético real de cualquier individuo dado. En algunas realizaciones, se puede diseñar un conjunto de hipótesis, de modo que cada uno de los estados genéticos razonablemente posibles se pueda describir con al menos una hipótesis de un conjunto. En algunas realizaciones de la presente divulgación, un aspecto del método es determinar qué hipótesis se corresponde con el estado genético real del individuo en cuestión.

En una realización de la presente divulgación, una etapa implica la creación de una hipótesis. En algunas realizaciones, puede ser una hipótesis del número de copias. En algunas realizaciones, puede implicar una hipótesis que concierna a qué segmentos de un cromosoma para cada uno de los individuos emparentados se corresponde genéticamente con qué segmentos, si los hay, de otros individuos emparentados. La creación de una hipótesis puede referirse al acto de establecer los límites de los parámetros, de modo que el conjunto entero de posibles estados genéticos que se consideren quede abarcado por dichos parámetros. La creación de una hipótesis puede referirse al acto de establecer los límites de los parámetros, de modo que un conjunto limitado de posibles estados genéticos en consideración quede abarcado por dichos parámetros. La creación de un conjunto de hipótesis puede referirse a la estimación y/o descripción de vínculos estadísticamente esperados de valores medidos que se corresponden con cada una de las hipótesis. La creación de un conjunto de hipótesis puede referirse a una persona capacitada con conocimiento para enumerar dichos estados de ploidía posibles que puedan ser razonablemente probables según las circunstancias. En una realización, puede referirse a la estimación del perfil de mediciones de SNP de un individuo diana tal como se mide sobre una matriz de SNP de alto rendimiento para un conjunto de contextos parentales.

Una "hipótesis del número de copias" también denominado "hipótesis de ploidía" o "hipótesis del estado de ploidía" puede referirse a una hipótesis que concierna a un posible estado de ploidía para un cromosoma dado, o una sección de un cromosoma, en el individuo diana. Puede referirse también al estado de ploidía en más de uno de los cromosomas en el individuo. Un conjunto de hipótesis del número de copias puede referirse a un conjunto de hipótesis, correspondiéndose cada hipótesis a un estado de ploidía posible diferente en un individuo sobre un cromosoma, o puede referirse a una combinación de hipótesis de un solo cromosoma sobre más de un cromosoma, pudiendo variar el número de cromosomas diferentes en los seres humanos de 2 a 23. Un individuo normal contiene uno de los cromosomas de cada uno de los progenitores. No obstante, debido a los errores en la meiosis y la mitosis, es posible que un individuo tenga 0, 1, 2, o más de un cromosoma dado de cada uno de los progenitores. En la práctica, es raro ver más de dos de un cromosoma dado de un progenitor. En la presente divulgación, las realizaciones consideran solamente la hipótesis posible en la que 0, 1 o 2 copias de un cromosoma dado provienen de un progenitor. En algunas realizaciones de la invención, para un cromosoma dado, existen nueve posibles hipótesis: tres posibles hipótesis concernientes a 0, 1 o 2 cromosomas de origen materno, multiplicadas por las tres posibles hipótesis concernientes a 0, 1 o 2 cromosomas de origen paterno. Asignemos que (m,f) se refieren a la hipótesis en la que m es el número de un cromosoma dado heredado de la madre y f es el número de un cromosoma dado heredado del padre. Entonces, las nueve hipótesis son (0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1) y (2,2) y se pueden escribir también H00, H01, H02, H10, H11, H12, H20, H21, H22. Las diferentes hipótesis se corresponden a diferentes estados de ploidía. Por ejemplo, (1,1) se refiere a un cromosoma disómico normal; (2,1) se refiere a trisomía materna y (0,1) se refiere a monosomía. En algunas realizaciones, se puede escribir la hipótesis como (m,f_x,f_y), para tener en cuenta el cromosoma del sexo, refiriéndose f_x a un cromosoma X o un cromosoma autosómico heredado del padre y refiriéndose f_y a un cromosoma Y heredado del padre. Cuando se utiliza esta notación para cromosomas autosómicos, la f_y puede actuar simplemente como marcador de posición. Por tanto, un

embrión euploide que sea H101 en el cromosoma 23 será de sexo masculino y si fuera H110 en el cromosoma 23, sería de sexo femenino. Por ejemplo, H000 representa la hipótesis de nulisomía; H100, H010 y H001 representan la hipótesis de monosomía; H110 y H101 representan la hipótesis de disomía normal; H200, H020, H002 y H011 representan la hipótesis de disomía uniparental; y H210, H120 y H111 representan la hipótesis de trisomía; y H220, H211 y H202 representan algunas de las hipótesis de tetrasomía posibles.

En algunas realizaciones, el caso de trisomía, en el que dos cromosomas son heredados de uno de los progenitores y un cromosoma es heredado del otro progenitor se pueden diferenciar dos casos más: uno en el que los dos cromosomas son idénticos (error de copia apareada) y uno en el que los dos cromosomas son homólogos pero no idénticos (error de copia sin aparear).

En algunas realizaciones, en las que los datos parentales están sincronizados y por tanto cada alelo puede estar especificado como parte de cualquiera de los dos haplotipos, hay dieciséis hipótesis posibles. En una realización, en la que solamente está sincronizado un progenitor, puede haber doce hipótesis. Es posible utilizar otros conjuntos de hipótesis. En una realización, se pueden descartar algunas hipótesis consideradas como improbables.

En algunas realizaciones de la presente divulgación, la hipótesis de ploidía puede referirse a una hipótesis concerniente a qué cromosoma de otros individuos emparentados se corresponde con un cromosoma encontrado en el genoma del individuo diana. En algunas realizaciones, el método emplea la información conocida de que es posible esperar que los individuos emparentados compartan bloques de haplotipos y utilizando los datos genéticos medidos de individuos emparentados, además de la información sobre qué bloques de haplotipos se aparean entre el individuo diana y el individuo emparentado, es posible deducir los datos genéticos correctos para un individuo diana con una fiabilidad más alta que con al uso de tan solo las mediciones genéticas del individuo diana. Así pues, en algunas realizaciones, la hipótesis de la ploidía puede concernir no solamente al número de cromosomas, sino también a qué cromosomas en los individuos emparentados son idénticos o casi idénticos a uno o más cromosomas en el individuo diana.

Una vez definido el conjunto de hipótesis, cuando los algoritmos operan en los datos genéticos de entrada, pueden producir una probabilidad estadística determinada para cada una de las hipótesis consideradas. Las probabilidades de las diversas hipótesis se pueden determinar calculando matemáticamente cada una de las diversas hipótesis, el valor de la probabilidad, tal como se señala según una o más de las técnicas, algoritmos y/o métodos expertos descritos a lo largo de la presente divulgación, divulgaciones relacionados y/o abarcadas en la técnica PARENTAL SUPPORT™, utilizando los datos genéticos pertinentes como entrada. El cálculo puede producir un valor exacto, puede dar una estimación, puede incluir un término de error, puede incluir fiabilidad y puede representar una probabilidad estadística.

Una vez calculadas las probabilidades de las diferentes hipótesis, tal como se determina a través de una pluralidad de técnicas, se pueden combinar. Esto puede entrañar para cada una de las hipótesis multiplicar las probabilidades según se determinan con cada técnica. El producto de las probabilidades de las hipótesis se puede normalizar. Advértase que una hipótesis de ploidía se refiere a un posible estado de ploidía para un cromosoma.

En algunas realizaciones, si la probabilidad para una hipótesis dada es superior a la probabilidad para cualquiera de las demás hipótesis, entonces se puede determinar la hipótesis como la más probable. En algunas realizaciones, la hipótesis se puede determinar como la más probable y el estado de ploidía u otros estados genéticos se pueden denominar si la probabilidad normalizada está por encima de un umbral. En una realización, esto puede significar que el número e identidad de los cromosomas que están asociados con esa hipótesis se puedan denominar como el estado de ploidía. En una realización, esto puede significar que la identidad de los alelos que están asociados con esa hipótesis se pueda denominar como el estado alélico y/o el estado genético. En algunas realizaciones, el umbral puede estar comprendido entre aproximadamente 50 % y aproximadamente 80 %. En algunas realizaciones, el umbral puede estar comprendido entre aproximadamente 80 % y aproximadamente 90 %. En algunas realizaciones, el umbral puede estar comprendido entre aproximadamente 90 % y aproximadamente 95 %. En algunas realizaciones, el umbral puede estar comprendido entre aproximadamente 95 % y aproximadamente 99 %. En algunas realizaciones, el umbral puede estar comprendido entre aproximadamente 99 % y aproximadamente 99,9 %. En algunas realizaciones, el umbral puede estar por encima de aproximadamente 99,9 %.

Soporte parental

Algunas realizaciones de la divulgación se pueden utilizar en combinación con el método PARENTAL SUPPORT™ (PS), describiéndose otras realizaciones del mismo en las tres solicitudes de patente: Rabinowitz 2006, 2008 y 2009. En algunas realizaciones, los métodos divulgados en el presente documento se pueden considerar como parte del método PARENTAL SUPPORT™. En algunas realizaciones, el método PARENTAL SUPPORT™ es una colección de métodos que se pueden emplear para determinar los datos genéticos, con una alta precisión, de una célula, o un reducido número de células, específicamente para determinar alelos relacionados con enfermedad, otros alelos de interés y/o el estado de ploidía de uno o más cromosomas en la(s) célula(s). PARENTAL SUPPORT™ puede referirse a cualquiera de estos métodos. PARENTAL SUPPORT™ es un ejemplo de método con soporte informático.

El método PARENTAL SUPPORT™ utiliza datos genéticos parentales conocidos, es decir, datos genéticos haplotípicos y/o diploides de la madre y/o del padre, junto con la información conocida del mecanismo de meiosis y la medición imperfecta del ADN diana y posiblemente de uno o más individuos emparentados para reconstruir *in silico* en un ordenador el genotipo en una pluralidad de alelos, y/o el estado de ploidía de un embrión o de cualquier célula(s) diana y el ADN diana en la localización de loci claves con un alto grado de fiabilidad. El método PARENTAL SUPPORT™ puede reconstruir no solamente polimorfismos de un solo nucleótido que han sido medidos de forma insuficiente, sino también inserciones y deleciones y SNP o regiones enteras de ADN que no se han medido en absoluto. Además, el método PARENTAL SUPPORT™ puede medir tanto loci ligados a enfermedad múltiples como rastrear aneuploidía, desde una sola célula o desde la misma cantidad de ADN reducida. En algunas realizaciones, el método PARENTAL SUPPORT™ puede utilizarse para caracterizar una o más células de embriones biopsados durante un ciclo de FIV para determinar el estado genético de una o más células. En algunas realizaciones, el método PARENTAL SUPPORT™ puede utilizarse para determinar el estado de ploidía de un feto a partir de ADN fetal que flota libremente y/o células fetales que se pueden encontrar en la sangre materna, o de alguna otra fuente.

En una realización, el método PARENTAL SUPPORT™ permite la limpieza de datos genéticos con ruido. Esto se puede realizar deduciendo los alelos genéticos correctos en el genoma diana (embrión o feto) utilizando el genotipo de individuos emparentados (progenitores) como referencia. PARENTAL SUPPORT™ puede ser particularmente pertinente cuando solamente se dispone de una reducida cantidad de material genético (p.ej. PGD o NIPGD) y cuando las mediciones directas de los genotipos conllevan inherentemente ruidos debido a la cantidad limitada de material genético. El método PARENTAL SUPPORT™ es capaz de reconstruir con alta precisión secuencias de alelos diploides ordenadas en el embrión, junto con el número de copias de segmentos de cromosomas, incluso aunque las mediciones de diploide sin ordenar se caractericen por un alto índice de pérdida de alelos, ausencias, desviaciones de amplificación variable y otros errores. El método puede emplear tanto un modelo genético subyacente como un modelo subyacente de error de la medición. El modelo genético puede determinar tanto las probabilidades de alelo en cada SNP como las probabilidades cruzadas entre SNP. Las probabilidades de alelo se pueden modelizar en cada SNP sobre la base de los datos obtenidos de los progenitores y las probabilidades de cruce del modelo entre los SNP sobre la base de los datos obtenidos de la base de datos HapMap, según el desarrollo del proyecto internacional HapMap. Dado el modelo genético subyacente y el modelo de error de la medición apropiados, se puede aplicar una estimación máxima *a posteriori* (MAP) con modificaciones para la eficiencia de la computación, para estimar los valores de alelo ordenados correctos en cada SNP en el embrión.

Un aspecto de la tecnología PARENTAL SUPPORT™ es un algoritmo de denominación del número de copias que, en algunas realizaciones, utiliza contextos de genotipo parental. Para denominar el número de copias de cromosomas, el algoritmo puede utilizar el fenómeno de pérdida de locus (LDO) combinado con las distribuciones de los genotipos embrionarios esperados. Durante la amplificación de todo el genoma, se produce LDO necesariamente. El índice de LDO está en consonancia con el número de copias del material genético del que se deriva, es decir, con un menor número de copias de cromosoma se produce una mayor LDO y al contrario. Así pues, de ello se infiere que los loci con ciertos contextos de genotipos parentales se comportan de un modo característico en el embrión, en relación con la probabilidad de contribuciones alélicas al embrión. Por ejemplo, si ambos progenitores tienen estados BB homocigóticos, entonces el embrión nunca deberá tener estados AB o AA. En este caso, es de esperar que las mediciones sobre el canal de detección A tengan una distribución determinada por el ruido de la contaminación residual y diversas señales de interferencia, pero no genotipos válidos. En cambio, si ambos progenitores tienen estados AA homocigóticos, entonces el embrión nunca deberá tener estados AB o BB y se espera que las mediciones en el canal A tengan la máxima intensidad posible dado el índice de LDO en una reacción de amplificación de todo el genoma en particular. Cuando el estado del número de copias subyacente en el embrión difiere de disomía, los loci correspondientes a contextos parentales concretos se comportan de un modo predecible sobre la base del contenido alélico adicional al que contribuye uno de los progenitores o que está ausente. Esto permite determinar el estado de ploidía en cada cromosoma, o segmento de cromosoma. Los detalles de las realizaciones de este método se describen a lo largo de la presente divulgación.

Respuesta de plataforma

Existen muchos métodos que se pueden utilizar para medir datos genéticos. Ninguno de los métodos conocidos actualmente dentro de la técnica tiene la capacidad de medir los datos genéticos con un 100 % de precisión; sino que más bien, siempre hay errores y/o desviaciones estadísticas de los datos. Se puede esperar que el método de medición introduzca ciertas desviaciones estadísticamente predecibles en la medición. Se puede esperar que ciertos conjuntos de ADN, amplificados a través de ciertos métodos y medidos con ciertas técnicas tengan como resultado mediciones que son cuantitativa y cualitativamente diferentes a otros conjuntos de ADN, amplificados a través de otros métodos y/o medidos con diferentes técnicas. En algunos casos, estos errores se pueden deber al método de medición. En algunos casos, dicho error puede deberse al estado del ADN. En algunos casos, dicha desviación puede deberse a la tendencia de algunos tipos de ADN a responder de forma diferente para un método de medición genética dado. En algunos casos, las mediciones pueden diferir de maneras que se correlacionan con el número de células usado. En algunos casos, las mediciones pueden diferir sobre la base de la técnica de medición, por ejemplo, qué técnica de secuenciación o técnica de genotipificación de matriz se utiliza. En algunos casos, diferentes cromosomas se pueden amplificar en diferentes grados. En algunos casos, ciertos alelos pueden tener

mayor o menor probabilidad de amplificarse. En algunos casos, el error, la desviación o la respuesta diferencial puede deberse a una combinación de factores. En muchos de estos casos, o en todos ellos, la previsibilidad estadística de estas diferencias de medición, denominado "respuesta de plataforma" puede utilizarse para corregir estos factores y puede tener como resultado datos para los cuales se aumenta al máximo la precisión y en los que cada medida está asociada con una fiabilidad apropiada.

La respuesta de plataforma puede describirse como una caracterización matemática de las características de entrada/salida de una plataforma de medición genética, como TAQMAN, la AFFYMETRIX GENECHIP o la ILLUMINA INFINIUM BEADARRAY. La respuesta de plataforma puede ser específica para una plataforma en particular, un modelo de máquina de genotipificación en particular, una máquina de genotipificación en particular o incluso un científico en particular que utilice una máquina de genotipificación en particular. La información de entrada en el canal es el material genético amplificado con cualquier material genético hibridado, marcado fluorescentemente. La información de salida del canal podría consistir en denominaciones de alelo (cualitativas) o mediciones numéricas en bruto (cuantitativa), dependiendo del contexto. Por ejemplo, en el caso de que se reduzca la información de salida numérica en bruto de la plataforma para denominaciones de genotipo cualitativas, la respuesta de la plataforma puede consistir en una matriz de transición de error que describa la probabilidad condicional de observar una denominación de genotipo de salida en particular dada una entrada de genotipo verdadera en particular. En una realización, en la que se deja como mediciones numéricas en bruto la información de salida de la plataforma, la respuesta de la plataforma puede ser una función de la densidad de probabilidad condicional que describa la probabilidad de la información de salida numérica dada una entrada de genotipo verdadera en particular.

En algunas realizaciones de la presente divulgación, puede utilizarse la información conocida de la respuesta de plataforma para corregir la desviación estadísticamente. En algunas realizaciones de la presente divulgación, la información conocida de la respuesta de plataforma puede utilizarse para aumentar la precisión de los datos genéticos. Esto puede realizarse llevando a cabo una operación estadística sobre los datos que actúe de manera opuesta a la tendencia a la desviación del proceso de medición. Esto puede implicar asociar la fiabilidad apropiada a un dato dado, de manera que cuando se combina con otros datos, la hipótesis que se encuentra como la más apropiada sea realmente la que con mayor probabilidad corresponda al estado genético real del individuo en cuestión.

En algunas realizaciones de la presente divulgación, puede utilizarse un método estadístico para eliminar la desviación de los datos debida a la tendencia de ciertos alelos maternos o paternos a amplificarse de manera desproporcionada con respecto a otros alelos. En algunas realizaciones de la presente divulgación, puede utilizarse un método estadístico para eliminar la desviación en los datos debida a la tendencia de ciertas sondas de amplificar ciertos SNP de manera desproporcionada con respecto a otros SNP.

Imaginemos los dos espacios dimensionales en los que el eje de coordenadas x es la intensidad de canal x y el eje de coordenadas y es la intensidad de canal y. En este espacio, es posible esperar que la media de los contextos caiga en la línea definida por las medias de los contextos BB|BB y AA|AA. En algunos casos, puede observarse que las medias de contextos promedio no caen en esta línea, sino que están desviados estadísticamente; esto se puede denominar "desviación fuera de línea". En algunas realizaciones de la presente divulgación, puede utilizarse un método estadístico para corregir la desviación fuera de línea en los datos.

En algunos casos, se pueden producir puntos extendidos en las medias de los contextos por translocación. Si se produce translocación, es posible esperar anomalías en los criterios de valoración del cromosoma solamente. Por lo tanto, si se descompone el cromosoma en segmentos y se traza un gráfico de la media de contextos de cada segmento, entonces puede esperarse que los segmentos que queden situados en una translocación respondan como una verdadera trisomía o monosomía, mientras que el resto de los segmentos parezcan disómicos. En algunas realizaciones de la presente divulgación, puede utilizarse un método estadístico para determinar si la translocación ha tenido lugar sobre un cromosoma dado observando las medias de los contextos de diferentes segmentos del cromosoma.

Determinación de ploidía cuando el material genético materno del individuo diana está contaminado

En el método de la invención, es posible determinar el estado de ploidía de un feto de manera no invasiva midiendo el ADN fetal contenido en la sangre materna. Adviértase que esto puede complicarse considerablemente por el hecho de que la cantidad de ADN fetal disponible en la sangre materna pueda ser reducido. La cantidad de ADN que flota libremente fetal que se encuentra en el suero es normalmente menos de 50 % y frecuentemente menos de 20 % y el ADN que flota libremente materno con contaminación residual hace que las mediciones en el ADN fetal tengan mucho ruido y sean difíciles de interpretar. El número de células fetales en la sangre materna es frecuentemente menos de 1 célula de cada 100.000 y puede ser de hasta 1 célula por millón, o incluso más baja. Este método supera las dificultades descritas, así como otras dificultades conocidas en la técnica. Este método puede aplicarse en los casos en los que la cantidad de ADN diana está en cualquier proporción con respecto al ADN no diana; por ejemplo, el ADN diana podría constituir cualquier intervalo entre 0,01 % y 99,99 % del ADN presente.

La primera etapa del método consiste en hacer mediciones genómicas de la madre y opcionalmente del padre, de manera que se sepan los datos genéticos diploides en un gran número de alelos de uno o de ambos progenitores. El número de alelos puede oscilar entre 100 y 100.000.000. En una realización, el número de alelos oscila entre 500 y 100,000 por cromosoma dirigido. En la realización de la invención, el número de alelos oscila entre 1,000 y 20.000 por cromosoma dirigido. En la realización de la invención, los alelos son SNP conocidos como polimórficos en la población humana. Una vez conocidos los genotipos parentales en un conjunto de SNP, se puede subdividir los SNP en una serie de conjuntos de SNP, correspondiendo cada conjunto al conjunto de SNP en un contexto parental en particular.

A continuación es posible determinar la respuesta de plataforma del sistema utilizando las mediciones genéticas de ciertos contextos. También es posible determinar la relación entre el ADN diana y el ADN materno en la muestra utilizando las mediciones genéticas de ciertos contextos. Asimismo, es posible determinar el ADO observado dada las mediciones genéticas observadas.

La siguiente etapa consiste en crear una serie de hipótesis, una para cada estado de ploidía de interés hipotético en un cromosoma de interés y determinar la distribución estadística esperada de las mediciones genotípicas para ese niño hipotético, dados los índices ADO esperados y dada la respuesta de plataforma esperada. Por ejemplo, en el cromosoma 21, se pueden prever varios genotipos del niño hipotéticos, por ejemplo, uno para un niño que sea disómico en el cromosoma 21 (H110) y uno de un niño que tenga trisomía materna en el cromosoma 21 (H210). Adviértase que para cromosomas autosómicos, $(H\alpha\beta\gamma)$ representa la hipótesis en la que están presentes copias α de un cromosoma proveniente de la madre, están presentes copias β de un cromosoma proveniente del padre y γ es un marcador de posición establecido en 0; en el caso del cromosoma del sexo, $(H\alpha\beta\gamma)$ representa la hipótesis en la que están presentes copias α en un cromosoma proveniente de la madre, P indica el número de cromosomas X provenientes del padre que están presentes y Y indica el número de cromosomas Y provenientes del padre que están presentes.

Adviértase que los genotipos hipotéticos no son necesariamente genotipos SNP-por-SNP, sino que más bien es de esperar que sean distribuciones estadísticas de SNP dentro de un contexto parental dado. Por ejemplo, imaginemos que examinamos solamente el contexto parental AA|AB, que significa el conjunto de SNP del individuo diana en el que la madre es homocigótica y el padre es heterocigoto. Es de esperar que el niño H110 tenga la misma posibilidad de un SNP AA o AB dentro del contexto parental y, por tanto, sería de esperar observar aproximadamente una relación A:B 3:1 para los SNP que están en el contexto parental AA|AB contexto parental. Es de esperar que el niño H210 tenga igual posibilidad de ser AAA o AAB dentro de ese contexto parental y por tanto, podría esperarse observar aproximadamente una relación A:B 5:1 para los SNP que están en el contexto parental AA|AB. Al observar las intensidades de canal medidas para los diversos nucleótidos puede ser posible determinar qué estado genético es el más probable para ese cromosoma: disomía o trisomía.

A continuación se describen ciertos aspectos de una realización de la invención con términos matemáticos más sólidos. Esta sección explica cómo se pueden tomar las mediciones genéticas parentales y las mediciones genéticas de la muestra mixta del material genético fetal y materno en forma de información de salida de la plataforma de genotipificación y transformar dichas mediciones en una denominación del número de copias.

Definiciones variables:

y = intensidad medida promedio de SNP en un contexto dado en un cromosoma en particular, en un canal en particular
 x = el número de copias de alelo estadísticamente esperado presentes por locus, para el canal que se está midiendo, para los SNP en el contexto
 Δ = la fracción de ADN fetal en la muestra
 n = la fracción de SNP que son A para un genotipo dado
 v = un término para representar el ruido observado que es una variable aleatoria con una distribución desconocida.

Es posible afirmar que $x \sim (1-\Delta)n_{madre} + \Delta n_{feto}$ y también que $y = f(x) + v$, es decir, la distribución de las mediciones de un conjunto de SNP dentro de un contexto parental dado será en función del número de alelos esperados en la muestra y la respuesta de plataforma, además de un factor ruido.

En una realización de la invención, puede darse por supuesto que $f(x)$ es un polinomio de segundo orden, es decir $f(x) \sim f_1x^2 + f_2x + f_3$. En otra realización, puede darse por supuesto que $f(x)$ es un polinomio de primer orden, es decir, $f(x) \sim f_1x + f_2$. En otra realización, puede darse por supuesto que $f(x)$ es una ecuación exponencial u otra ecuación algebraica, o alguna combinación de ellas. Supongamos que v es una distribución de Gauss con una media 0 y una desviación típica = V

Debe entenderse que podría suponerse que $f(x)$ es cualquier número de funciones, como un polinomio de primer orden, un polinomio de tercer orden, cualquier otro polinomio, cualquier relación exponencial, algebraica u otra entre

x e y. Debe entenderse asimismo que v podría ser cualquier número de distribuciones, incluyendo una distribución de Gauss, una distribución de Rayleigh, una distribución de Pearson o una distribución de Bernoulli.

En este punto, x se conoce en los términos de Δ y n y f_1, f_2, f_3, Δ y la distribución de v, parametrizado por V, es desconocida. Para una muestra dada, se realiza una medición genotípica, y, de la muestra para una serie de SNP para cada contexto, para cada canal, sobre una serie de cromosomas, incluyendo el (los) cromosoma(s) de interés, cuyo estado de ploidía ha de ser determinado, así como al menos un cromosoma que se puede esperar que sea disómico. A continuación, se combina cada conjunto de y en un vector. Adviértase que el conjunto de cromosomas cuyo estado de ploidía ha de ser determinado y el al menos un cromosoma que se puede esperar que sea disómico se pueden solapar.

Por ejemplo, en el ser humano, existe un conjunto de cromosomas que puede tener como resultado un nacimiento con vida incluso siendo aneuploides, más comúnmente cromosoma 13, 18, 21, X e Y. Asimismo, se conoce niños vivos que nacen con aneuploidía en los cromosomas 4, 5, 7, 8, 9, 11, 15, 16, 22. Adviértase que otros estados de aneuploidía, como translocaciones y disomía uniparental, en cualquier cromosoma pueden dar lugar a niños nacidos con anomalías cromosómicas. Puede utilizarse uno de los cromosomas que se encuentra de forma infrecuente como aneuploide en fetos en gestación con latido, por ejemplo 1, 2 o 3, como cromosoma diploide de referencia. Alternativamente, puede usarse uno de los cromosomas a los que va dirigida la prueba de aneuploidía como referencia, ya que es poco probable que exista más de una anomalía cromosómica importante en total en el feto en gestación. En una realización de la invención, los cromosomas a los que se dirige la detección de aneuploidía incluyen 13, 18, 21, X y Y.

Dado y, o y_m medidos, para el cromosoma que se espera que sea disómico y dado el número esperado de A medidas en la muestra, $x_{(H110)}$, es posible entonces encontrar una estimación de máxima probabilidad para f_1, f_2, f_3, v y Δ . La estimación de máxima probabilidad se puede realizar utilizando un método de gradiente descendiente no lineal. Una vez estimados f_1, f_2, f_3, v y Δ , se pueden hacer distribuciones en cuanto al valor previsto de y, y_p , para las diversas hipótesis de estado de ploidía, por ejemplo $y_{\text{previsto}(H110)}$ e $y_{\text{previsto}(H210)}$.

El y_m observado se puede comparar con las distribuciones para y_p y puede determinarse la probabilidad de cada hipótesis, que es la probabilidad de observar y_m de acuerdo con el modelo previsto. La hipótesis con la máxima probabilidad corresponde al estado de ploidía del feto que es más probable. Se puede calcular una fiabilidad en la denominación de ploidía a partir de las diferentes probabilidades de las diversas hipótesis.

Para un cromosoma en particular, se da por supuesto que se han calculado las probabilidades $p(y_m|H110)$, $p(y_m|H210)$ y $p(y_m|H120)$. Asimismo, se da por supuesto que la probabilidad a priori de cada hipótesis es conocida a partir de un estudio de población estadístico. Por ejemplo, $p(H110)$ es la probabilidad global de disomía en este cromosoma para la población de interés. Si $p(y_m|H110)$ es la probabilidad máxima, entonces la fiabilidad de disomía se calcula aplicando la regla Bayes como fiabilidad = $p(y_m|H110) p(H110) / (p(y_m|H110) p(H110) + p(y_m|H210) p(H210) + p(y_m|H120) p(H120))$.

Visión general del método

En una realización, la presente divulgación presenta un método según el cual se puede determinar el estado de ploidía de un feto en gestación, en uno o más cromosomas, de manera no invasiva, utilizando la información genética determinada por el ADN fetal encontrado en la sangre materna y los datos genéticos de la madre y el padre. El ADN fetal se puede purificar, purificar parcialmente o no purificar; las mediciones genéticas se pueden realizar sobre el ADN proveniente de más de un individuo. Los métodos de tipo informático pueden deducir la información genética del individuo diana, como por ejemplo el estado de ploidía, a partir de las mediciones genotípicas en bloque en un conjunto de alelos. El conjunto de alelos puede contener varios subconjuntos de alelos, pudiéndose corresponder uno o más de los subconjuntos a alelos que se encuentran en el individuo diana pero que no se encuentran en los individuos que no son diana y uno o más conjuntos distintos pueden corresponder a los alelos que se encuentran en el individuo no diana y no se encuentran en el individuo diana. El conjunto de alelos puede contener también subconjuntos de alelos, encontrándose el alelo en la diana y la no diana en relaciones que difieren según lo esperado. El método puede implicar la comparación de las relaciones de intensidades de salida medidas para diversos subconjuntos de alelos con las relaciones esperadas dados diversos posibles estados de ploidía. La respuesta de plataforma se puede determinar y se puede incorporar en el método una corrección de las desviaciones del sistema. La determinación de la ploidía puede realizarse con una fiabilidad computada. La determinación de la ploidía puede ir ligada a una acción clínica. Dicha acción clínica puede ser interrumpir o no interrumpir un embarazo. Una realización de la invención implica el caso en el que el individuo diana es un feto y el individuo no diana es la madre biológica del feto.

En una explicación básica, el método funciona del siguiente modo. Una versión simple de la idea consiste en tratar de cuantificar la cantidad de ADN fetal en los SNP, teniendo el feto un alelo que no tiene la madre. En primer lugar, se miden los datos genotípicos de los progenitores aplicando un método que produce datos para un conjunto de SNP. A continuación, se clasifican los SNP en contextos parentales. Los SNP encontrados en los contextos en los que la madre es heterocigótica, AB, se consideran menos informativos, ya que el ADN contaminante en la sangre

materna tendrá una gran cantidad de ambos alelos. Los SNP encontrados en contextos en los que la madre y el padre tienen el mismo conjunto de alelos se consideran asimismo menos informativos, ya que la contaminación residual y la señal fetal coinciden. El método simple se centra en los contextos en los que el padre tiene un alelo que la madre no tiene, por ejemplo: AA|AB y AA|BB (y BB|AB y BB|BB, aunque son los mismos por simetría.) En el caso del contexto AA|BB, se espera que el feto sea AB y por tanto deberá aparecer el alelo B en el ADN fetal. En el caso del contexto AA|AB, se espera que el feto sea AA la mitad de las veces y AB la mitad de las veces, lo que significa que el alelo B debería aparecer en el ADN fetal la mitad de las veces.

Una vez seleccionados los contextos apropiados y agrupados los SNP por contextos parentales, por ejemplo, los contextos AA de la madre, se identifican los SNP apropiados en los que se ha medido el SNP B, lo que indica que el feto es AB, junto con las cantidades de ADN medidas para cada uno de dichos SNP. En este momento, se comparan las intensidades de las mediciones de los SNP para un cromosoma supuestamente disómico con las intensidades de las mediciones de los SNP para el cromosoma de interés y se ajustan apropiadamente para la respuesta de plataforma. Si las intensidades de los SNP para cada uno de los dos cromosomas son aproximadamente iguales, entonces el cromosoma de interés se considera disómico y si las intensidades en el cromosoma es aproximadamente un 50 % más que la intensidad del cromosoma supuestamente disómico, entonces, el cromosoma de interés se considera trisómico paterno.

Adviértase que esto es una explicación básica de una versión simple del método. En una realización, se pueden utilizar algunos contextos, o todos ellos, incluyendo aquellos de mayor o menor capacidad de información. En una realización, se pueden utilizar algunos o todos los SNP. Para los contextos y SNP que son más informativos, por ejemplo, los SNP en el contexto AA|BB, las mediciones pueden tener un mayor peso en el cálculo global. Para los contextos y SNP que son menos informativos, por ejemplo, los SNP en el contexto AA|AA, las mediciones pueden tener un menor peso en el cálculo global. La explicación que se ha expuesto se centra en la medición del número de cromosomas paternos. Puede utilizarse un método similar para determinar el número de cromosomas maternos, realizando los ajustes apropiados. Por ejemplo, las relaciones de intensidades de SNP esperadas para las hipótesis de disomía y trisomía serán diferentes, ya que los datos genotípicos maternos con contaminación residual y los datos genotípicos fetales serán similares o idénticos. Por ejemplo, en el caso en el que la muestra mixta contiene 20 % de ADN fetal y 80 % de ADN materno, al examinar un contexto AA|BB en cuanto a la disomía, se podría esperar una relación de 90:10 para las mediciones A:B (80 % A más 20 % 1:1 A:B), para trisomía materna se podría esperar una relación más cercana a 93,3:6,7 (80 % A más 20 % 2:1 A:B) y para trisomía paterna, se podría esperar una relación más cercana a 86,7:13,3 (80 % A más 20 % 1:2 A:B).

Adviértase que este método puede utilizarse igualmente bien con más o menos información genotípica de los progenitores. Por ejemplo, si se desconoce el genotipo del padre, el método puede considerarse como más informativos todos los contextos en los que la madre es homocigótica (AA) y se puede calcular la posibilidad de que el feto tenga SNP B de forma aproximada a partir de las heterocigosidades de la población. Al mismo tiempo, si el genotipo del padre está sincronizado, es decir, se conocen los haplotipos, se puede aumentar las precisiones del número de copias, ya que habrá correlaciones más fuertes entre los contextos esperados. Por ejemplo, imaginemos tres SNP correlacionados en un cromosoma en el que los contextos son AA|AB, AA|BA, AA|AB (el padre está sincronizado.) Si se detecta el alelo B en la sangre materna para el primer SNP, existe una probabilidad mucho más alta de detectar un B para el tercer alelo, en contraposición con el segundo alelo, ya que un feto euploide hereda solamente un haplotipo de cada padre. Al mismo tiempo, si está sincronizado el genotipo de la madre, se aumentan de forma similar las precisiones, ya que habrá correlaciones más esperadas entre las contribuciones fetales esperadas con las intensidades de SNP relativas.

Al utilizar cada uno de los contextos parentales y cromosomas conocidos como euploides, es posible estimar a través de un conjunto de ecuaciones simultáneas, la cantidad de ADN en la sangre materna de la madre y la cantidad de ADN de la sangre materna del feto. Estas ecuaciones simultáneas son posibles gracias a la información conocida de los alelos presentes en la madre y, opcionalmente, el padre. En una realización, se usan los datos genéticos tanto de la madre como del padre. En particular, los alelos presentes en el padre y que no están presentes en la madre proporcionan una medición directa del ADN fetal. Entonces, es posible examinar los cromosomas de interés en particular, como por ejemplo el cromosoma 21 y observar si las mediciones en este cromosoma en cada contexto parental coinciden o no con una hipótesis en particular, como por ejemplo H_{mp} en la que m representa el número de cromosomas maternos y p representa el número de cromosomas paterno, p.ej., representando H_{11} euploide, o representando H_{21} y H_{12} trisomía materna y paterna, respectivamente.

En algunas realizaciones de la divulgación, se puede emplear un método con la información conocida del genotipo materno y sin la información conocida del genotipo paterno. En este caso, es posible deducir contextos del padre examinando los datos de SNP para las mediciones en la muestra mixta que no se pueden explicar con los datos de la madre. Se podría identificar los SNP en los que la madre es homocigótica (AA) y examinar después los datos de SNP de la muestra mixta para los alelos B. Para los SNP es posible deducir que el padre era AB o BB y el feto es AB. Igualmente, para los SNP en los que la madre es AA y no se midió B en la muestra mixta, es posible deducir que el feto es AA con cierta probabilidad, estando correlacionada la probabilidad con los índices de ADO y LDO. Es posible utilizar los datos parentales con cierto grado de incertidumbre asociada a las mediciones. Los métodos

descritos en el presente documento se pueden adaptar para determinar el estado de ploidía del feto dadas cantidades mayores y/o menores de información genética de los progenitores.

Algunos supuestos

5 Adviértase que estos supuestos no tienen por qué ser verdaderos para que este método funcione según lo pretendido, sino que representan el caso idealizado para el que está diseñada esta derivación.

- 10 – * La cantidad de material genético esperado en la sangre materna de la madre es constante en todos los loci.
- * La cantidad de material genético esperada presente en la sangre materna del feto es constante a través de todos los loci suponiendo que los cromosomas sean euploides.
- * Los cromosomas que no son viables (excluyendo 13, 18, 21, X, Y) son todos euploides en el feto. En una realización, solamente algunos de los cromosomas no viables en el feto necesitan ser euploides.

15 *Problema de formulación general:*

Se puede escribir $y_{ijk} = g_{ijk}(x_{ijk}) + v_{ijk}$ en la que x_{ijk} es la cantidad de ADN en el alelo $k = 1$ o 2 (1 representa el alelo A y 2 representa el alelo B), $j = 1...23$ representan el número de cromosomas y $i = 1...N$ representa el número de locus en el cromosoma, g_{ijk} es la respuesta de plataforma para un locus y alelo en particular ijk y v_{ijk} es ruido independiente de la medición para ese locus y alelo. La cantidad de material genético es dada por $x_{ijk} = am_{ijk} + \Delta c_{ijk}$ en la que a es el factor de amplificación (o el efecto neto de filtración difusión, amplificación etc.) del material genético presente en cada uno de los cromosomas maternos, m_{ijk} (0,1 o 2) es el número de copias del alelo en particular en los cromosomas maternos, Δ es el factor de amplificación del material genético presente en cada uno de los cromosomas del niño y c_{ijk} es el número de copias (0,1,2 o 3) del alelo en particular en los cromosomas del niño. Adviértase que para la primera explicación simplificada, se da por supuesto que a y Δ son independientes del locus y el alelo, es decir independientes de i, j y k . Por tanto, se puede afirmar que:

$$y_{ijk} = g_{ijk}(am_{ijk} + \Delta c_{ijk}) + v_{ijk}$$

30 *Enfoque en el que se utiliza un modelo afín que es uniforme a través de todos los loci:*

Es posible modelizar g con un modelo afín y dar por supuesto para mayor simplicidad que el modelo es el mismo para cada uno de los locus y alelos, si bien será evidente tras la lectura de la presente divulgación cómo modificar el enfoque cuando el modelo afín depende de i, j, k . Supongamos que el modelo de respuesta de plataforma es

$$35 \quad g_{ijk}(x_{ijk}) = b + am_{ijk} + \Delta c_{ijk}$$

en la que los factores de aplicación a y Δ se emplean sin pérdida de generalidad y se añade un intercepto b del eje y que define el nivel de ruido cuando no hay material genético. El objetivo es estimar a y Δ . También es posible estimar b independientemente, pero en esta sección, se da por supuesto que el nivel de ruido es constante de forma aproximada a través de los loci y solamente se utiliza el conjunto de ecuaciones basadas en contextos parentales para estimar a y Δ . La medición en cada locus viene dada por:

$$y_{ijk} = b + am_{ijk} + \Delta c_{ijk} + v_{ijk}$$

45 Suponiendo que el ruido v_{ijk} es independiente e idénticamente distribuido (i.i.d.) para cada una de las mediciones dentro de un contexto parental en particular, T , se pueden sumar las señales dentro de ese contexto parental. Los contextos parentales se representan en lo que se refiere a los alelos A y B, en los que los dos primeros alelos representan a la madre y los dos segundos alelos representan al padre: $T \square \{AA|BB, BB|AA, AB|AB, AA|AA, BB|BB, AA|AB, AB|AA, AB|BB, BB|AB\}$. Para cada contexto T , hay un conjunto de loci i, j conformándose el ADN parental para ese contexto representado por $i, j \square T$. Por tanto:

$$y_{T,k} = \frac{1}{N_T} \sum_{i,j \in T} y_{i,j,k} = b + am_{k,T} + \Delta c_{k,T} + v_{k,T}$$

55 en la que $m_{k,T}$, $c_{k,T}$ y $v_{k,T}$ representan las medias de los valores correspondientes a lo largo de todos los loci que conforman el contexto parental T , o a lo largo de todos los $i, j \in T$. La media o los valores esperados $c_{k,T}$ dependerán del estado de ploidía del niño. En la tabla a continuación, se describe la media o los valores esperados $m_{k,T}$ y $c_{k,T}$ para $k = 1$ (alelo A) o 2 (alelo B) y todos los contextos parentales T . Los valores esperados se calculan dando por supuestas diferentes hipótesis sobre el niño, por ejemplo: euploidía y trisomía materna. Las hipótesis se representan con la notación H_{mf} , en la que m se refiere al número de cromosomas de la madre y f se refiere al número de

cromosomas del padre, p.ej., H₁₁ es euploide, H₂₁ es trisomía materna. Advértase que hay simetría entre algunos estados conmutando A y B, pero se incluyen todos los estados para mayor claridad:

Contexto	AA/BB	BB/AA	AB/AB	AA/AA	BB/BB	AA/AB	AB/AA	AB/BB	BB/AB
m _{A,T}	2	0	1	2	0	2	1	1	0
m _{B,T}	0	2	1	0	2	0	1	1	2
CA,T H ₁₁	1	1	1	2	0	1,5	1,5	0,5	0,5
CB,T H ₁₁	1	1	1	0	2	0,5	0,5	1,5	1,5
CA,T H ₂₁	2	1	1,5	3	0	2,5	2	1	0,5
CB,T H ₂₁	1	2	1,5	0	3	0,5	1	2	2,5

5 Se describe aquí un conjunto de ecuaciones en las que se describen todos los valores esperados $y_{T,k}$, que se pueden fundir en forma de matriz, del siguiente modo:

$$Y = B + A_H P + v$$

10 en la que

$$Y = \begin{bmatrix} Y_{AA|BB,1} & Y_{BB|AA,1} & Y_{AB|BB,1} & Y_{AA|AA,1} & Y_{BB|BB,1} & Y_{AA|AB,1} & Y_{AB|AA,1} & Y_{AB|BB,1} & Y_{BB|AB,1} \\ Y_{AA|BB,2} & Y_{BB|AA,2} & Y_{AB|AB,2} & Y_{AA|AA,2} & Y_{BB|BB,2} & Y_{AA|AB,2} & Y_{AB|AA,2} & Y_{AB|BB,2} & Y_{BB|AB,2} \end{bmatrix}^T$$

$$P = \begin{bmatrix} a \\ \Delta \end{bmatrix}$$

15

es la matriz de parámetros para su estimación

$$B = b\mathbf{1}$$

20

en la que $\mathbf{1}$ es la matriz 18x1 de unos

$$v = \begin{bmatrix} v_{A,AA|BB} & \dots & v_{B,BB|AB} \end{bmatrix}^T$$

25

es la matriz 18x1 en lo que se refiere a los ruidos

y A_H es la matriz de encapsulación de datos en la tabla, en la que los valores para cada hipótesis H sobre el estado de ploidía del niño son diferentes. A continuación, se exponen ejemplos de la matriz A_H para la hipótesis de ploidía H₁₁ y H₂₁

$$A_{H_{21}} = \begin{bmatrix} 2.0 & 1.0 \\ 0 & 1.0 \\ 1.0 & 1.0 \\ 2.0 & 2.0 \\ 0 & 0 \\ 2.0 & 1.5 \\ 1.0 & 1.5 \\ 1.0 & 0.5 \\ 0 & 0.5 \\ 0 & 1.0 \\ 2.0 & 1.0 \\ 1.0 & 1.0 \\ 0 & 0 \\ 2.0 & 2.0 \\ 0 & 0.5 \\ 1.0 & 0.5 \\ 1.0 & 1.5 \\ 2.0 & 1.5 \end{bmatrix} \quad A_{H_{11}} = \begin{bmatrix} 2.0 & 2.0 \\ 0 & 1.0 \\ 1.0 & 1.5 \\ 2.0 & 3.0 \\ 0 & 0 \\ 2.0 & 2.5 \\ 1.0 & 2.0 \\ 1.0 & 1.0 \\ 0 & 0.5 \\ 0 & 1.0 \\ 2.0 & 2.0 \\ 1.0 & 1.5 \\ 0 & 0 \\ 2.0 & 3.0 \\ 0 & 0.5 \\ 1.0 & 1.0 \\ 1.0 & 2.0 \\ 2.0 & 2.5 \end{bmatrix}$$

30

Para estimar a Δ , o la matriz P, se agregan los datos a través de todos los cromosomas que se pueden suponer como euploides en la muestra del niño. Esto podría incluir algunos o todos los cromosomas $j = 1 \dots 23$ que han sido medidos, excepto aquellos que son inciertos y por tanto sometidos a prueba. En una realización, los cromosomas inciertos incluyen $j = 13; 18, 21, X$ e Y. En una realización, es posible aplicar una prueba de concordancia para los resultados en los cromosomas del individuo para detectar aneuploidía de mosaico en los cromosomas no viables.

35

Para mayor claridad de la notación, se define Y' como Y medido en todos los cromosomas euploides e Y'' como Y medido sobre un cromosoma de ensayo en particular, como por ejemplo cromosoma 21, que puede ser aneuploide. Ha de aplicarse la matriz $A_{H_{21}}$ a los datos euploides para estimar los parámetros:

$$\hat{P} = \underset{P}{\operatorname{argmin}} \|Y' - B - A_{H_{21}} P\|_2 = (A_{H_{21}}^T A_{H_{21}})^{-1} A_{H_{21}}^T \tilde{Y}$$

en la que $\tilde{Y} = Y' - B$ es decir, los datos medidos eliminando la desviación. La solución de cuadrados mínimos anterior es solamente la solución de máxima probabilidad si cada uno de los términos de la matriz de ruido v tiene una varianza similar. En algunos casos, no es el caso, sobre todo simplemente porque el número de loci N_T utilizados para computar la medición media para cada contexto T puede ser diferente para cada contexto. Como en el caso anterior, N_T se refiere al número de loci utilizado en los cromosomas conocidos como euploides y C' representa la matriz de covarianza para la media de las mediciones en los cromosomas conocidos como euploides. Existen muchos enfoques para estimar la covarianza C' de la matriz de ruido v , que se pueden suponer como distribuida como $v \sim N(0, C')$. Dada la matriz de covarianza, la estimación más probable de P es

$$\hat{P} = \underset{P}{\operatorname{argmin}} \|C'^{-1/2}(Y' - B - A_{H_{21}} P)\|_2 = (A_{H_{21}}^T C'^{-1} A_{H_{21}})^{-1} A_{H_{21}}^T C'^{-1} \tilde{Y}$$

Un enfoque simple para estimar la matriz de covarianza es dar por supuesto que todos los términos de v son independientes (es decir términos que no están fuera de la diagonal) e invocar el Teorema de Límite Central para que la varianza de cada término de v escale $1/N_T$ y entonces encuentra la matriz 18×18

$$C' = \begin{bmatrix} 1/N'_{AA|BB} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1/N'_{BB|AB} \end{bmatrix}$$

Una vez estimado P' , se usan estos parámetros para determinar la hipótesis más probable en el cromosoma sometido a estudio, como por ejemplo el cromosoma 21. Es decir, se pueden seleccionar las siguientes hipótesis:

$$H^* = \underset{H}{\operatorname{argmin}} \|C''^{-1/2}(Y'' - B - A_H \hat{P})\|_2$$

Una vez encontrada H^* es posible estimar el grado de fiabilidad de la determinación de H^* . Supongamos por ejemplo que hay dos hipótesis en consideración: H_{11} (euploide) y H_{21} (trisomía materna). Asignemos que $H^* = H_{11}$. Las mediciones de distancia que corresponden a cada una de las hipótesis se pueden computar del siguiente modo:

$$d_{11} = \|C''^{-1/2}(Y'' - B - A_{H_{11}} \hat{P})\|_2$$

$$d_{21} = \|C''^{-1/2}(Y'' - B - A_{H_{21}} \hat{P})\|_2$$

Se puede demostrar que el cuadrado de estas mediciones de la distancia está distribuido aproximadamente como variable aleatoria del chi cuadrado con 18 grados de libertad. Asignemos que χ_{18} representa la correspondiente función de densidad de probabilidad de dicha variable, entonces se puede encontrar la relación en las probabilidades p_H de cada una de las hipótesis de acuerdo con:

$$\frac{p_{H_{11}}}{p_{H_{21}}} = \frac{\chi_{18}(d_{11}^2)}{\chi_{18}(d_{21}^2)}$$

Las probabilidades de cada hipótesis pueden calcularse sumando la ecuación $p_{H_{11}} + p_{H_{21}} = 1$. La fiabilidad de que el cromosoma sea de hecho euploide es dada por $p_{H_{11}}$.

En algunas realizaciones, es posible modificar el enfoque anterior para diferentes desviaciones b en cada uno de los canales que representan los alelos A y B. La matriz de desviación B se redefine del siguiente modo:

$$B = \begin{bmatrix} b_A \vec{1} \\ b_B \vec{1} \end{bmatrix}$$

en la que $\vec{1}$ es una matriz 9×1 de unos. Tal como se ha explicado anteriormente, puede darse por supuesto cada uno de los parámetros b_A y b_B sobre la base de mediciones a priori, o se pueden incluir en la matriz P y estimar activamente (es decir, hay un rango suficiente en las ecuaciones a lo largo de todos los contextos para hacerlo).

En una realización, en la formulación general en la que $y_{ijk} = g_{ijk}(am_{ijk} + \Delta c_{ijk}) + v_{ijk}$, es posible medir o calibrar directamente la función g_{ijk} para cada locus y alelos, de manera que se puede invertir la función (que es monótonica para la inmensa mayoría de las plataformas de genotipificación). A continuación, es posible utilizar inversión de la función para refundir las mediciones en lo que se refiere a la cantidad de material genético para que el sistema de ecuaciones sea lineal, es decir, $y'_{ijk} = g_{ijk}^{-1}(y_{ijk}) = am_{ijk} + \Delta c_{ijk} + v'_{ijk}$. Este enfoque es particularmente favorable cuando g_{ijk} es una función afín, de manera que la inversión no produce amplificación o desviación del ruido en v'_{ijk} .

En algunas realizaciones, el término ruido modificado $v'_{ijk} = g_{ijk}^{-1}(v_{ijk})$ se puede amplificar o desviar por inversión de la función. Otra realización que puede ser más óptima desde la perspectiva del ruido es linealizar las mediciones en torno al punto de operación, es decir:

$$y_{ijk} = g_{ijk}(am_{ijk} + \Delta c_{ijk}) + v_{ijk}$$

se puede refundir como:

$$y_{ijk} \approx g_{ijk}(am_{ijk}) + g'_{ijk}(am_{ijk})\Delta c_{ijk} + v_{ijk}$$

en el caso en el que la fracción de ADN que flota libremente en la sangre materna del niño sea reducida, Δ « a y la expansión es una aproximación razonable. Alternativamente, para la respuesta de plataforma, como por ejemplo la de ILLUMINA BEADARRAY, que se aumenta monótonicamente y para la cual la segunda derivada es normalmente negativa, es posible mejorar la estimación de linealización de acuerdo con $y_{ijk} \approx g_{ijk}(am_{ijk}) + 0,5 (g'_{ijk}(am_{ijk}) + g''_{ijk}(am_{ijk} + \Delta c_{ijk}))\Delta c_{ijk} + v_{ijk}$. El conjunto de ecuaciones resultante se pueden resolver iterativamente para a y Δ aplicando un método como por ejemplo optimización de Newton-Raphson.

En algunas realizaciones, es posible medir en la cantidad total de ADN en el cromosoma de ensayo (madre más feto) y compararlo con la cantidad de ADN de todos los demás cromosomas, sobre la base del supuesto de que la cantidad de ADN deberá ser constante a través de todos los cromosomas. Para estimar los vínculos de fiabilidad de una forma significativa, es posible examinar una desviación típica a través de otras señales de cromosoma que debería ser euploide para estimar la varianza de señal y generar un vínculo de fiabilidad. Para calibrar las desviaciones de amplificación entre diferentes cromosomas, es posible encontrar una función de regresión que enlace cada nivel de señal medio del cromosoma al nivel de señal medio de todos los demás cromosomas, combinar la señal de todos los cromosomas ponderando sobre la base de la varianza del ajuste de regresión y observar si el cromosoma de ensayo de interés está dentro del intervalo aceptable tal como se define mediante los demás cromosomas.

En algunas realizaciones, puede utilizarse este método en combinación con otros métodos descritos anteriormente por Gene Security Network, sobre todo los métodos que forman parte de PARENTAL SUPPORT™ y que se mencionan a lo largo de la presente divulgación, de manera que se pueda sincronizar a los progenitores para saber qué contiene cada uno de los cromosomas materno y paterno individuales. Al considerar la oportunidad relativa de cada uno de los alelos en loci heterocigóticos, es posible determinar qué haplotipo de la madre está presente en el niño. A continuación, es posible comparar el nivel de señal del haplotipo materno medible con el haplotipo paterno que está presente (sin ruido de contaminación residual de la madre) y observar cuándo esa relación 1:1 no se satisface por la aneuploidía que causa un desequilibrio entre los alelos materno y paterno.

No se pretende que esta lista de posibles variaciones del método sea exhaustiva. Se pueden emplear otras variaciones. Adviértase que en la presente divulgación, para los fines del cálculo, se han podido plantear también ciertos supuestos sobre los parámetros, características de los datos, variables, etc. En estos casos, es posible plantear otros supuestos sin por ello cambiar la esencia de la divulgación.

Modelización

En una realización, se pueden producir los datos en bruto con una micromatriz que mide la respuesta de cada uno de los alelos posibles en una selección de SNP. En una realización, la micromatriz puede ser una micromatriz de SNP ILLUMINA o una micromatriz de SNP AFFYMETRIX. En otras realizaciones, se pueden emplear también otras fuentes de datos, como por ejemplo un número suficientemente grande de sondas TAQMAN o una matriz no basada en SNP. Los datos genéticos en bruto pueden ser de otras fuentes también, como por ejemplo secuenciación de ADN.

En esta realización, se espera normalmente que un SNP sea uno de dos nucleótidos. Por ejemplo, se puede esperar que sea G o C y puede medirse para la respuesta de G o C; alternativamente, en un SNP que pudiera tener A o T, se puede medir para la respuesta A y T. Dado que solamente son posibles dos alelos en cada SNP, las mediciones pueden agregarse sin atender así el SNP es A/T o C/G. Por el contrario, la presente divulgación puede referirse a respuestas en los canales x e y, y alelos genéricos A o B. Por lo tanto, los posibles genotipos en este ejemplo son AA, AB y BB para todos los SNP. Existen otras formas de agrupamiento de denominaciones de alelo que no afectarán a la esencia de la divulgación.

Las mediciones se pueden agregar inicialmente sobre SNP del mismo contexto parental sobre la base de genotipos parentales sin ordenar. Cada contexto se puede definir según el número de alelos A y el número de alelos B en cada progenitor: $[a_m \ b_m \ a_f \ b_f]$ en la que $a_m + b_m = 2$ y $a_f + b_f = 2$. Por ejemplo, todos los SNP en los que el genotipo de la madre es AA y el genotipo del padre es BB son miembros del contexto AA|BB. La combinación de 3 genotipos posibles sobre 2 progenitores significa que las mediciones de un solo cromosoma podrían consistir en 18 medias de contexto, 9 de cada canal. Consideremos una hipótesis del número de copias para el niño en la forma (n_m, n_f) en la que n_m es el número de copias de la madre y n_f es el número de copias del padre del cromosoma. Asignemos que el número esperado de A (promediado sobre los SNP) en el niño es k_x y que el número esperado de B es k_y (para un contexto en particular, condicionado sobre la hipótesis). El número esperado de alelos depende del contexto y de la hipótesis.

$$\begin{aligned} k_x &= 0.5a_m n_m + 0.5a_f n_f \\ k_y &= 0.5b_m n_m + 0.5b_f n_f \end{aligned} \quad (1)$$

La cantidad de ADN medido en un SNP dependerá del número de alelos presentes en ese SNP en los cromosomas materno y fetal y las concentraciones globales de ADN presentes en la muestra de la madre y el feto. El factor α refleja la concentración global de ADN en la muestra y la relación entre la madre y el niño es de 1 a δ .

Para SNP en contextos en los que los progenitores son homocigóticos, se conocen los genotipos de un niño disómico. Por ejemplo, si uno de los genotipos del progenitor es AA y el otro es BB, el genotipo del niño debe ser AB. En cambio, los SNP en los que uno de los progenitores es heterocigoto tendrán un genotipo del niño desconocido. Si se considera el contexto AB|BB, en el que el niño puede heredar A o B de la madre, la suposición más general es que el niño heredará el A y el B con igual probabilidad y, por tanto, aproximadamente la mitad de los genotipos del niño en este contexto serán AB y la mitad será BB. Se pueden plantear otros supuestos en lo que se refiere a la probabilidad de que un niño herede un alelo dado de un progenitor dado. Si bien el genotipo de cada SNP del niño no es conocido, se conocen de este modo los valores medios de k_x y k_y para SNP en cada contexto, y así, las ecuaciones indicadas a continuación se refieren a estas medias.

En el ejemplo en el que el contexto parental es AB|BB, el número medio de A en los SNP del niño es 0,5 y el número medio de B es 1,5. Las cantidades x_x y x_y se refieren a la cantidad media de ADN presente en los SNP en un contexto particular, en el que x_x es el ADN que se mida en el canal x (alelo) y x_y es el ADN que se mida en el canal y (alelo B).

$$\begin{aligned} x_x &= \alpha (m_x + \delta k_x) \\ x_y &= \alpha (m_y + \delta k_y) \end{aligned} \quad (2)$$

La cantidad de ADN se puede medir a través de las respuestas de plataforma sobre los canales x e y. Los NP en el mismo contexto se pueden agregar para producir mediciones y_x, y_y que son las respuestas de la media de los contextos en los canales x e y. Supongamos que los SNP son i.i.d.

Gracias a un exhaustivo análisis (para el algoritmo medio del cromosoma completo, como parte de PARENTAL SUPPORT™) se han encontrado diferencias sistemáticas en la amplificación entre cromosomas. Asignemos que y_c e y_1 son las medias del mismo contexto y la misma muestra de cromosoma c y cromosoma 1 respectivamente. El valor esperado de y_c/y_1 se define como β_c y se puede calcular a partir de un extenso conjunto de datos de entrenamiento. Los datos de entrenamiento consisten en cientos de blastómeros que han sido analizados según un protocolo de laboratorio consistente. Las ponderaciones de cromosoma β dependen del tipo de micromatriz (ya que diferentes matrices miden diferentes SNP) y el tipo de tampón de lisis utilizado, pero pueden ser consistentes entre las muestras por lo demás. Por lo tanto, el número esperado de A o B puede ponderarse mediante β atendiendo a este efecto, con el resultado de un cromosoma-número de alelos ponderado \hat{m} o \hat{k} .

$$\hat{m}_{xc} = m_{xc} \beta_{xc}$$

$$\hat{m}_{yc} = m_{yc} \beta_{yc}$$

$$\hat{k}_{xc} = k_{xc} \beta_{xc}$$

$$\hat{k}_{yc} = k_{yc} \beta_{yc}$$

Al atender a la variación de cromosomas utilizando un número de alelos ponderado, se puede considerar el modelo de respuesta de plataforma $f_x(x_x), f_y(x_y)$ consistente a través de los cromosomas. No obstante, se puede observar que varía la desviación b según el cromosoma y el canal y el ruido de la medición v variará en cada medición. La desviación de un cromosoma y canal en particular es la media del ruido-solo contexto y por tanto es una cantidad conocida (medida directamente). El ruido – solo contextos son AA|AA para el canal y, y BB|BB para el canal x, ya

que en estos casos, el número de alelos medidos esperado es cero. Por lo tanto, la medición da un valor basal para la respuesta de plataforma en ausencia de la señal que mide. Puede darse por supuesto que la covarianza del ruido escalar asociado a cada medición de la media de los contextos es proporcional a $1/n$ siendo n el número de SNP incluido. Esto se corresponde con el supuesto de SNP i.i.d. dentro de cada contexto. Los componentes de ruido pueden suponerse como independientes y distribuirse normalmente.

$$y_x = f_x(x_x) + b_x + v_x$$

$$y_y = f_y(x_y) + b_y + v_y$$

Respuesta de plataforma cuadrática

En una realización, puede utilizarse un modelo de respuesta de plataforma (relación afín entre la cantidad de ADN y la señal medida). En otra realización, puede utilizarse una respuesta de plataforma cuadrática $f(x) = f_1x^2 + f_2x$. En algunas realizaciones, se puede usar una respuesta de plataforma cuadrática en la que f_1 y f_2 son específicas para cada muestra y el canal de medición y x es la cantidad de ADN. Se pueden emplear otros modelos de respuesta de plataforma, incluyendo relaciones de algoritmo de orden superior o relaciones exponenciales. La sustitución de (2) para la cantidad de ADN tiene como resultado el siguiente modelo para las respuestas del canal x e y en el cromosoma c desde el contexto i .

$$\begin{aligned} y_{xci} &= f_{1x}\alpha^2\hat{m}_{xci}^2 + f_{1x}\alpha^2\delta^2\hat{k}_{xci}^2 + 2f_{1x}\alpha^2\delta\hat{m}_{xci}\hat{k}_{xci} + f_{2x}\alpha\hat{m}_{xci} + f_{2x}\alpha\delta\hat{k}_{xci} + b_{xc} + v_{xci} \\ y_{yci} &= f_{1y}\alpha^2\hat{m}_{yci}^2 + f_{1y}\alpha^2\delta^2\hat{k}_{yci}^2 + 2f_{1y}\alpha^2\delta\hat{m}_{yci}\hat{k}_{yci} + f_{2y}\alpha\hat{m}_{yci} + f_{2y}\alpha\delta\hat{k}_{yci} + b_{yc} + v_{yci} \end{aligned} \quad (3)$$

Sin perder la generalidad, la concentración α de ADN y las respuestas de plataforma f_{1x} , f_{2x} , f_{1y} , f_{2y} pueden combinarse para formar el conjunto de 5 parámetros para la muestra. Adviértase que cuando el modelo incluye términos de la forma $p_{1x}\delta^2$ $p_{1x}\delta$ y $p_{2x}\delta$ etc., la estimación del parámetro no se puede resolver exactamente utilizando métodos lineales.

$$\begin{aligned} y_{xci} &= p_{1x}\hat{m}_{xci}^2 + p_{1x}\delta^2\hat{k}_{xci}^2 + 2p_{1x}\delta\hat{m}_{xci}\hat{k}_{xci} + p_{2x}\hat{m}_{xci} + p_{2x}\delta\hat{k}_{xci} + b_{xc} + v_{xci} \\ &= g_{xci}(p) + b_{xc} + v_{xci} \\ y_{yci} &= p_{1y}\hat{m}_{yci}^2 + p_{1y}\delta^2\hat{k}_{yci}^2 + 2p_{1y}\delta\hat{m}_{yci}\hat{k}_{yci} + f_{2y}\hat{m}_{yci} + f_{2y}\delta\hat{k}_{yci} + b_{yc} + v_{yci} \\ &= g_{yci}(p) + b_{yc} + v_{yci} \end{aligned} \quad (4)$$

$$p = \begin{bmatrix} p_{1x} \\ p_{2x} \\ p_{1y} \\ p_{2y} \\ \delta \end{bmatrix} = \begin{bmatrix} f_{1x}\alpha^2 \\ f_{2x}\alpha^2 \\ f_{1y}\alpha^2 \\ f_{2y}\alpha^2 \\ \delta \end{bmatrix} \quad (5)$$

En esta descripción, puede darse por supuesto que este conjunto de parámetros p es común a todos los cromosomas y contextos de genotipo parental para una sola muestra y, así, el modelo para un solo cromosoma c y el contexto i se pueden escribir en la siguiente forma condensada sobre la base de la función g de la respuesta de plataforma no lineal.

$$\begin{bmatrix} y_{xci} \\ y_{yci} \end{bmatrix} = \begin{bmatrix} g_{xci}(p) \\ g_{yci}(p) \end{bmatrix} + \begin{bmatrix} b_{xc} \\ b_{yc} \end{bmatrix} + \begin{bmatrix} v_{xci} \\ v_{yci} \end{bmatrix}$$

$$y_{ci} = g_{ci}(p) + b_c + v_{ci}$$

El conjunto de N mediciones de una muestra se puede combinar para formar una ecuación de vector en p .

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} g_1(p) \\ \vdots \\ g_N(p) \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

$$y = g(p) + b + v$$

En otras realizaciones, los parámetros pueden ser diferentes para diferentes cromosomas o para diferentes muestras.

Respuesta de plataforma cuadrática linealizada

5

Consideremos la linealización de la respuesta de plataforma cuadrática en $x = x_0$:

$$f(x) \approx f_1 x^2_0 + f_2 x_0 + (2f_1 x_0 + f_2)(x - x_0)$$

10 La sustitución de la contribución de la madre $\alpha \hat{m}$ para la cantidad de ADN nominal x_0 tiene como resultado el siguiente modelo:

$$y_{xci} = f_{1x} \alpha^2 \hat{m}^2_{xci} + 2f_{1x} \alpha^2 \delta \hat{m}_{xci} \hat{k}_{xci} + f_{2x} \alpha \hat{m}_{xci} + f_{2x} \alpha \delta \hat{k}_{xci}$$

15

$$y_{y ci} = f_{1y} \alpha^2 \hat{m}^2_{y ci} + 2f_{1y} \alpha^2 \delta \hat{m}_{y ci} \hat{k}_{y ci} + f_{2y} \alpha \hat{m}_{y ci} + f_{2y} \alpha \delta \hat{k}_{y ci}$$

Aunque la respuesta de plataforma está linealizada, el modelo sigue siendo no lineal en el conjunto de parámetros de modelo desconocidos definidos en (5). En una realización, se puede implementar un método de estimación lineal construyendo un conjunto de parámetros aumentado que elimina los términos no lineales añadiendo grados extra de libertad. Este conjunto de parámetros aumentado tiene 8 grados de libertad. En otra realización, es posible intentar este tipo de solución lineal para todo el modelo cuadrático. Se muestran cuatro parámetros para el canal X y se definen de forma similar los parámetros para el canal Y.

20

$$q_x = \begin{bmatrix} q_{1x} \\ q_{2x} \\ q_{3x} \\ q_{4x} \end{bmatrix} = \begin{bmatrix} f_{1x} \alpha^2 \\ f_{2x} \alpha \\ f_{1x} \alpha^2 \delta \\ f_{2x} \alpha \delta \end{bmatrix} \quad (6)$$

25

Utilizando este conjunto de parámetros, se puede escribir en forma de matriz el modelo linealizado para un cromosoma c y contexto i.

30

$$A_{xci} = [\hat{m}^2_{xci} \quad \hat{m}_{xci} \quad 2\hat{m}_{xci} \hat{k}_{xci} \quad \hat{k}_{xci}]$$

$$A_{y ci} = [\hat{m}^2_{y ci} \quad \hat{m}_{y ci} \quad 2\hat{m}_{y ci} \hat{k}_{y ci} \quad \hat{k}_{y ci}]$$

$$A_{ci} = \begin{bmatrix} A_{xci} & 0 \\ 0 & A_{y ci} \end{bmatrix}$$

35

$$\begin{bmatrix} y_{xci} \\ y_{y ci} \end{bmatrix} = A_{ci} \begin{bmatrix} q_x \\ q_y \end{bmatrix} + \begin{bmatrix} b_{xci} \\ b_{y ci} \end{bmatrix} + \begin{bmatrix} v_{xci} \\ v_{y ci} \end{bmatrix}$$

$$y_{ci} = A_{ci} q + b_{ci} + v_{ci}$$

40

Las mediciones de todos los cromosomas, contextos y canales se pueden combinar en una sola ecuación de matriz con parámetros q en R^8 tal como sigue:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

$$y = Aq + b + v \quad (7)$$

45

Recuérdese que y son la media de las mediciones del contexto, A es el conjunto de coeficientes conocidos, q es el conjunto de parámetros que se va a estimar, b es el vector de desviación conocido y v es el ruido de Gauss medio supuesto en cero.

Estimación de parámetros

50

En una realización, la estrategia para la estimación de parámetros es dar por supuesto que un subconjunto de cromosomas del niño es disómico (que tienen una copia de cada progenitor) y usarlos para averiguar los parámetros de modelo para la muestra del niño. Dichos parámetros de modelo de muestra se utilizan luego para clasificar el

resto de los cromosomas, para determinar cuántas copias están presentes de cada uno de los progenitores. Por lo tanto, se pueden calcular las contribuciones de alelo del niño \hat{m}_{xci} , \hat{m}_{yci} a partir de (1) en la etapa de estimación de parámetros bajo el supuesto de que las contribuciones del número de copias de la madre y el padre n_m y n_f son ambos uno. Si D es el número de cromosomas disómicos supuesto, entonces el vector de medición y para la estimación de parámetros tiene el tamaño 18D (de nueve medias de los contextos medidos sobre dos canales).

Modelo de sensor cuadrático linealizado

El modelo cuadrático linealizado (7) lleva a soluciones de cuadrados mínimos directos (LS) o probabilidad máxima (ML) para la mejor estimación de q. La solución de probabilidad máxima depende del número de SNP incorporado en cada medición, dado en la matriz diagonal N. En una realización, se utiliza la solución de probabilidad máxima ya que la capacidad de información de los diferentes componentes de medición varía bastante y la matriz N que determina esta variación es conocida.

$$q^*_{LS} = \operatorname{argmin} \|y - (Aq + b)\|^2 = (A^T A)^{-1} A^T (y - b)$$

$$q^*_{ML} = \operatorname{argmax} P(y; q) = \operatorname{argmin} \|N^{-0.5}(y - Aq - b)\|^2 = (A^T N^{-1} A)^{-1} A^T N^{-1} (y - b)$$

Modelo de sensor cuadrático

Es posible que el modelo de sensor cuadrático no conduzca a soluciones de forma cerrada para la estimación del parámetro p que se ajuste mejor a las mediciones. En otra realización, se puede aplicar un método de optimización de gradiente descendente que mejora iterativamente sobre una conjetura inicial para p con el fin de minimizar una función de coste. La formulación de cuadrados mínimos no lineales para p minimiza la diferencia de cuadrados medios entre los datos medidos y los valores previstos por el modelo.

$$p^* = \operatorname{argmin} \|y - g(p) - b\|^2$$

Las funciones de optimización no lineal comerciales, como FMINCON de MATLAB, usan métodos iterativos para encontrar un mínimo local de una función de coste proporcionada por el usuario por aproximación numérica del gradiente de función.

La estimación del parámetro q* sobre la base del modelo linealizado puede proporcionar una condición inicial conveniente para la optimización no lineal ya que resuelve una aproximación del mismo problema pero se puede calcular de forma cerrada a un reducido coste computacional. La comparación de parámetros linealizados (q) y no lineales (p) a continuación demuestra que el mapeo de p a q no es reversible.

$$q = \begin{bmatrix} q_{1x} \\ q_{2x} \\ q_{3x} \\ q_{4x} \\ q_{1y} \\ q_{2y} \\ q_{3y} \\ q_{4y} \end{bmatrix} = \begin{bmatrix} f_{1x} \alpha^2 \\ f_{2x} \alpha \\ f_{1x} \alpha^2 \delta \\ f_{2x} \alpha \delta \\ f_{1y} \alpha^2 \\ f_{2y} \alpha \\ f_{1y} \alpha^2 \delta \\ f_{2y} \alpha \delta \end{bmatrix}, \quad p = \begin{bmatrix} p_{1x} \\ p_{2x} \\ p_{1y} \\ p_{2y} \\ \delta \end{bmatrix} = \begin{bmatrix} f_{1x} \alpha^2 \\ f_{2x} \alpha^2 \\ f_{1y} \alpha^2 \\ f_{2y} \alpha^2 \\ \delta \end{bmatrix}$$

El mapeo de p a q se escribirá q(p) y es el siguiente.

$$q(p) = [p_{1x} \quad p_{2x} \quad p_{1x} \delta \quad p_{2x} \delta \quad p_{1y} \quad p_{2y} \quad p_{1y} \delta \quad p_{2y} \delta]^T$$

dado q = q*_{MLE}, seleccionar p₀ = argmina ||q - q(p)||₂, que tiene una solución de polinomio de forma cerrada. A continuación, puede utilizarse p₀ como condición inicial para una solución iterativa de p* = arg min ||y-g(p)-b||.

Puede utilizarse una estimación de la distribución del vector de ruido v en el cálculo de las probabilidades de observación. El vector de error de ajuste e = y - g(p*) - b es una muestra de la distribución de v. Recuérdese que el supuesto de SNP i.i.d. implica que las medias de contexto tendrán una varianza proporcional al número incluido de SNP. Por lo tanto, la covarianza V de v tiene la forma γN^{-1} en la que γ es escalar y N es la matriz diagonal que define

el número de SNP medido en cada media de contexto. La matriz N es conocida y γ se estima como la varianza de los componentes de $N^{0.5}e$.

Determinación del número de copias

5 Después de estimar los parámetros de modelo para una muestra en particular sobre la base de un conjunto de cromosomas disómicos conocidos, la tarea es estimar el número de copias para el cromosoma de interés o para el resto de cromosomas. Recuérdese que una hipótesis del número de copias del niño tiene la forma $H_{n_m n_f}$ en la que (n_m, n_f) representa el número de copias a las que contribuye la madre y el padre, respectivamente. En una realización, se pone el punto de mira en la detección de trisomías, en las que uno de los progenitores contribuye con una copia extra ya que estos errores tienen como resultado fetos viables y afecciones como síndrome de Down. La hipótesis del número de copias predice el número esperado de alelos del niño presentes en un SNP con un contexto parental en particular, de acuerdo con (1). Por ejemplo, consideremos el contexto AA|BB en el que la madre tiene el genotipo AA y el padre tiene el genotipo BB. Bajo la hipótesis de disomía H_{11} , el genotipo del niño será AB, pero bajo la hipótesis de trisomía materna H_{21} , los genotipos del niño serán AAB y se puede detectar una señal más alta en el canal x debido a A extra. El número de alelos del niño presentes aparece en la matriz A en el modelo linealizado y en la función $g(p)$ del modelo cuadrático y depende de la hipótesis supuesta de esta manera. Por lo tanto, el supuesto de una hipótesis del número de copias en particular h tiene como resultado en el modelo correspondiente $A^h q$ o $g^h(p)$. Se evaluarán las diversas hipótesis considerando la probabilidad de los datos observados con diferentes modelos.

20 Consideremos el vector de medición $y_c \in \mathbb{R}^{18}$ de un cromosoma c . Recuérdese que y_c contiene las mediciones medias de 18 contextos del cromosoma, en el que cada una es una media de las mediciones de los SNP en el contexto de genotipo parental. La sustitución de una hipótesis en el modelo averiguado tiene como resultado una distribución $p(y_c|h)$ que depende implícitamente de los parámetros del modelo averiguados. Las probabilidades de las diversas hipótesis h se pueden resolver a partir de las probabilidades $\{p(y_c|h)\}$ incorporando distribuciones previas utilizando la regla de Bayes. La clasificación es posible cuando las distribuciones $p(y_c|h_i)$ y $p(y_c|h_j)$ se pueden distinguir para diferentes hipótesis h_i y h_j . Para un solo cromosoma, definimos d_{ij} como la diferencia de cuadrados medios en la salida del modelo que compara la hipótesis h_i y h_j .

30
$$d_{ij} = \frac{1}{18} \sum_{i=1}^{18} (g^i(p^*) - g^j(p^*))^2$$

Se puede esperar una denominación de alta fiabilidad entre la hipótesis h_i y h_j cuando d_{ij} es grande en comparación con la varianza de ruido sensor.

35 La estimación se puede basar en el modelo sensor cuadrático, $y = g(p) + b + v$. Condicionado por un conjunto de parámetros de modelo y una hipótesis, el vector de medición y_c se distribuye normalmente con la media $g^h(p^*) + b$ y la covarianza $V = \gamma N^{-1}$. Al definir el vector de error $e_c^h = y_c - g^h(p^*) - b$ para una hipótesis h y un cromosoma c , es posible observar que e_c^h se distribuye normalmente con una media cero y covarianza V y $e_c^{hT} V^{-1} e_c^h$ tiene la distribución del chi cuadrado con 18 grados de libertad.

40
$$p(y_c|h) = p_{x_{18}} \left((y_c - g^h(p^*) - b) \frac{1}{\gamma} N (y_c - g^h(p^*) - b) \right)$$

Denominación del número de copias con datos genéticos paternos sincronizados

45 En una realización de la divulgación, se pueden emplear los datos del genotipo del padre sincronizados. En esta sección se describe una realización que aprovecha la ventaja de los datos parentales sincronizados. Esta sección describe una extensión de una realización descrita anteriormente; está diseñada para el caso en el que se dispone de los datos genotípicos sincronizados del padre y permite una estimación de parámetros y el ajuste de la hipótesis con más precisión.

50 Cuando los datos del genotipo están sincronizados, entonces se puede distinguir el genotipo AB del genotipo BA. Por tanto, en el genotipo AB, el primer haplotipo tiene un alelo A en un locus dado y el segundo haplotipo tiene un alelo B en el locus, mientras que en el genotipo BA, el primer haplotipo tiene un alelo B en el locus y el segundo haplotipo tiene el aleo en el locus. Cuando el genotipo no está sincronizado, o está desordenado, no se hace distinción entre AB y BA y normalmente se hace referencia a él como AB.

60 La sincronización del genotipo del padre puede realizarse a través de diversos métodos, entre los que se incluyen varios de los que se pueden encontrar en las tres solicitudes de patente Rabinowitz 2006, 2008 y 2009. En la presente sección se da por supuesto que se dispone de los datos genotípicos sincronizados del padre, lo que significa que se conoce en todos los cromosomas el genotipo del padre ordenado en todos los SNP, es decir, es posible distinguir entre el primer y el segundo haplotipo del genotipo del padre. Si los datos genotípicos del padre están sincronizados y por lo tanto $AB \neq BA$ para el padre, y en cambio los datos de la madre no están sincronizados,

es decir $AB=BA$ para la madre, entonces hay doce contextos parentales diferentes posibles: $AA|AA$, $AA|AB$, $AA|BA$, $AA|BB$, $AB|AA$, $AB|AB$, $AB|BA$, $AB|BB$, $BB|AA$, $BB|AB$, $BB|BA$ y $BB|BB$.

5 Las mediciones se pueden agregar inicialmente en los SNP del mismo contexto parental sobre la base de los genotipos del padre sincronizados. Cada contexto se puede definir por el número de alelos A y el número de alelos B de la madre, de la primera cadena del padre y de la segunda cadena del padre: $[a_m \ b_m \ a_{f1} \ a_{f2} \ b_{f1} \ b_{f2}]$ en la que $a_m + b_m = 2$ y $a_{f1} + b_{f1} = 1$, $a_{f2} + b_{f2} = 1$. La combinación de 3 genotipos posibles de la madre y 4 genotipos posibles del padre significa que las mediciones de un cromosoma único consistirán en 24 contextos medios, 12 en cada canal.

10 Consideremos una hipótesis del número de copias para el niño, para un cromosoma en particular, de la forma (n_m, n_f) en la que n_m es el número de copias de la madre y n_f es el número de copias del padre del cromosoma. Para el genotipo paterno sincronizado, esta hipótesis se puede escribir en una forma (n_m, n_{f1}, n_{f2}) en la que n_m es el número de copias de la madre y n_{f1} es el número de copias del padre de primera cadena, n_{f2} es el número de copias del padre de segunda cadena del cromosoma, siendo $n_f = n_{f1} + n_{f2}$. (Adviértase: es una notación diferente a la mencionada a lo largo de la presente divulgación, que está en la forma (m, f_x, f_y) y tiene en cuenta el cromosoma del sexo). Por lo tanto, la hipótesis de disomía normal, anteriormente escrita en la forma $(n_m, n_f) = (1, 1)$ se puede extender en dos sub-hipótesis $(n_m, n_n, n_{f2}) = (1, 1, 0)$ y $(n_m, n_{f1}, n_{f2}) = (1, 0, 1)$, en las que se diferencian dos haplotipos paternos. La trisomía materna, anteriormente escrita en la forma $(2, 1)$, se puede extender en dos sub-hipótesis $(2, 1, 0)$ y $(2, 0, 1)$. La trisomía paterna se puede extender en sub-hipótesis que incluyen trisomías mitóticas paternas $(1, 2, 0)$, $(1, 0, 2)$ y la trisomía meiótica paterna $(1, 1, 1)$.

25 Debido a los posibles cruces entre cadenas paternas, la hipótesis del niño, escrita en forma (n_m, n_{f1}, n_{f2}) no tiene por qué permanecer igual en todo el cromosoma. Por ejemplo, supongamos que un cromosoma tiene una disomía normal con la primera cadena paterna $(1, 1, 0)$, en un conjunto de SNP adyacentes. Si existe un cruce de cadenas paternas en el siguiente SNP, cambia la hipótesis del número de copias del niño a $(1, 0, 1)$, implicando ahora la segunda cadena del padre.

30 Para mantener una hipótesis constante sobre un conjunto de SNP dado para el cálculo, se divide el cromosoma en segmentos N de SNP adyacentes. Es posible dividir los cromosomas en segmentos de varias formas, por ejemplo, para mantener el número e SNP por segmento constante, o para mantener el número de segmentos por cromosoma constante. En el presente documento se da por supuesto que la hipótesis del número de copias es constante en todo el segmento, sin presencia de cruces. En esta explicación se omiten los segmentos ambiguos con posibles cruces paternos, para mayor claridad.

35 Para cada uno de los segmentos, se agrupan las mediciones por contexto parental y se agregan las mediciones de intensidad en cada grupo de SNP. Por lo tanto, en este caso, las mediciones de un único cromosoma consistirán en medias de $24 \cdot N$ contextos, $12 \cdot N$ en cada canal (para cada N segmentos en un cromosoma).

40 Asignemos que el número de A esperado (media en las SNP) en el niño es k_x y el número esperado de Bs es k_y (para un contexto en particular, condicionado según una hipótesis). El número de alelos esperado depende del contexto y la hipótesis. Para cada segmento del cromosoma, para cada contexto parental ordenado:

$$k_x = 0.5a_m n_m + a_{f1} n_{f1} + a_{f2} n_{f2}$$

$$45 \quad k_y = 0.5b_m n_m + b_{f1} n_{f1} + b_{f2} n_{f2}$$

El modelo es similar al modelo para contextos parentales sin ordenar:

$$50 \quad x_x = \alpha (m_x + \delta k_x)$$

$$x_y = \alpha (m_y + \delta k_y)$$

$$y_x = f_x(x_x) + b_x + v_x$$

$$55 \quad y_y = f_y(x_y) + b_y + v_y$$

y es posible utilizar el modelo $f(x) = f_1 x^2 + f_2 x$

60 Se pueden utilizar cromosomas supuestamente disómicos para ajustar los parámetros del modelo (cromosomas "de entrenamiento"), es decir, suponer que $(n_m, n_f) = (1, 1)$. Es posible determinar la sub-hipótesis de disomía exacta (n_m, n_{f1}, n_{f2}) , ya sea $(1, 1, 0)$ o $(1, 0, 1)$ en cada segmento de cada cromosoma de "entrenamiento" examinando las respuestas de intensidad para diferentes contextos parentales ordenados, para cada segmento por separado del siguiente modo:

En primer lugar, se determina el nivel de ruido para la respuesta del canal x examinando la respuesta del canal x para el contexto parental BB|BB y se determina el nivel de ruido para el canal y examinando la respuesta del canal y para el contexto parental AA|AA, (el canal x mide los alelos A y el canal y mide los alelos B). A continuación, si la hipótesis es (1,1,0), se espera que las respuestas del canal y para el contexto parental ordenado AA|AB sean solamente ruido, sin señal y tengan el mismo comportamiento que las respuestas para el contexto AA|AA. De manera similar, se espera que las respuestas de canal x para el contexto parental ordenado BB|BA sean solo ruido, sin señal y que tengan el mismo comportamiento que las respuestas para el contexto BB|BB.

5 Si la hipótesis es (1,1,0), se espera que las respuestas del canal y para el contexto parental ordenado AA|AB sean solamente ruido, sin señal y tengan el mismo comportamiento que las respuestas para el contexto AA|AA. De manera similar, se espera que las respuestas de canal x para el contexto parental ordenado BB|BA sean solo ruido, sin señal y que tengan el mismo comportamiento que las respuestas para el contexto BB|BB.

10 Si la hipótesis es (1,0,1), las respuestas del canal y para el contexto parental ordenado AA|BA deberán ser solo ruido, sin señal y tendrán el mismo comportamiento que las respuestas para el contexto AA|AA. De manera similar, las respuestas del canal x para el contexto parental ordenado BB|AB deberán ser solo ruido sin señal y tendrán el mismo comportamiento que las respuestas para el contexto BB|BB.

15 Si seleccionamos como las sub-hipótesis más probables en este segmento cualquiera de las hipótesis (1,1,0) o (1,0,1), que se ajustan mejor a los datos. Es posible omitir de los posteriores análisis los segmentos en los que la opción es ambigua, es decir, segmentos en los que tiene lugar una probabilidad cruzada.

20 Para entrenar el modelo utilizando cromosomas disómicos, se ajustan los parámetros (α , δ , f_1 , f_2) para este modelo a partir de las observaciones $12 \times 2 \times N \times n_i$, en las que n_i es el número de cromosomas de "entrenamiento" utilizado.

25 En una realización, se pone el punto de mira en la detección de trisomías, en las que uno de los progenitores contribuye con una copia extra. Adviértase que los nacimientos con aneuploidía más viables son como resultado de trisomías. La hipótesis que se ajusta con cromosomas de "ensayo" (los cromosomas de interés) se puede plantear de forma similar a la de los genotipos sin ordenar, a excepción de que cada sub-hipótesis de trisomía (por ejemplo (101) frente a (119)) puede ajustarse por separado para cada segmento y las hipótesis para el estado de ploidía de los segmentos se pueden agregar, centrándose solamente en el estado global de ploidía (considerando ahora que (101) y (110) son iguales, ambos disomía; centrándose por ejemplo en disomía frente a trisomía materna, frente a trisomía paterna) y se puede calcular la estadística para cromosomas enteros.

30 En particular, supongamos que en el segmento i , la probabilidad de una sub-hipótesis en particular en el formato de hipótesis ordenado es $P_i(n_m, n_{f1}, n_{f2})$. En el formato de hipótesis sin ordenar se calcula la probabilidad de la hipótesis de disomía como $P_i(n_m, n_f) = P_i(1,1) = P_i(1,1,0) + P_i(1,0,1)$. Para trisomía materna $P_i(2,1) = P_i(2,1,0) + P_i(2,0,1)$. Para trisomía paterna $P_i(1,2) = p_{mt} * (P_i(1,2,0) + P_i(1,0,2)) + p_{me} * P_i(1,1,1)$, en la que p_{mt} es la probabilidad de trisomía paterna mitótica dado que se ha producido trisomía paterna y p_{me} es la probabilidad de trisomía paterna mitótica dado que se ha producido trisomía paterna, determinado a partir de la bibliografía y la práctica en general. Adviértase que las trisomías mitóticas y meióticas se pueden diferenciar o es posible que no se diferencien.

35 Dada la probabilidad de hipótesis $P_i(n_m, n_f)$ para cada segmento $i=1, \dots, N$, se calcula la probabilidad en todo el cromosoma como $P(n_m, n_f) = \prod_{i=1, \dots, N} P_i(n_m, n_f)$. La denominación de hipótesis para cada cromosoma se realiza seleccionando la hipótesis con mayor probabilidad.

40 Debería ser evidente, dado el beneficio de la presente divulgación, cómo modificar el método en el caso en el que el genotipo materno está sincronizado y el genotipo paterno no está sincronizado. Deberá ser asimismo evidente, dado el beneficio de la presente invención, cómo modificar el método en el caso en el que ambos genotipos paterno y materno están sincronizados.

Sección experimental

45 A continuación, se describe el aspecto experimental de la divulgación. Para demostrar la reducción de práctica de la divulgación, se obtuvo una mezcla de células de varios individuos, en los que el estado de ploidía de los individuos era conocido y se utilizaron los algoritmos antes descritos para determinar el estado de ploidía de uno de los individuos.

50 Se prepararon muestras genómicas de una línea celular de cultivo celular materno (AG16778, CORIELL) y de descendencia (AG16777, CORIELL). Se cultivaron las células en condiciones normales (1x medio RPMI 1640, 15 % de suero bovino fetal (FBS), 0,85 % estreptomicina) y se purificó ADN genómico utilizando un equipo QIAAMP DNA Micro Kit (QIAGEN) de acuerdo con las recomendaciones del fabricante. Se cuantificó ADN purificado utilizando un instrumento NANODROP (THERMO SCIENTIFIC) y se diluyó en las concentraciones apropiadas en tampón 1xTris-EDTA. Se preparó una serie de tres muestras genómicas mixtas (a-c) combinando (a) 59,4 ng ADN AG16777 con 132,6 ng ADN AG16778 (30 % AG16777), (b) 76,8 ng ADN AG 16777 con 115,2 ng ADN AG 16778 (40 % AG16777) y (c) 115,2 ng ADN AG 16777 con 76,8 ng ADN AG16778 (60 % AG 16777). Se diluyeron las tres muestras en H₂O para una concentración total de ADN de 3 ng/ul. Se almacenaron las muestras a - 20 °C y después se analizaron en la plataforma de matriz INFINIUM (ILLUMINA), actuando de acuerdo con las recomendaciones del fabricante.

Este método es apropiado para cualquier ácido nucleico que se pueda utilizar para la plataforma de matriz INFINIUM de ILLUMINA o cualquier otro método de genotipificación basado en el SNP, por ejemplo ADN que flota libremente aislado desde el plasma o amplificaciones (p.ej. amplificación de genoma entero, PCR) del mismo, ADN genómico aislado de otros tipos de células (p.ej., linfocitos) o amplificaciones de los mismos. Se puede emplear cualquier método que genere ADN genómico (p.ej. extracción de ADN, purificación) para la preparación de la muestra.

El ADN genómico utilizado en esta punto se puede mezclar previamente para simular una mezcla de ADN fetal y materno, si bien el método también es aplicable para ADN (o amplificaciones del mismo) como tales (es decir, sin mezclar previamente). Se prepararon tres muestras a partir de estas líneas celulares, con 30, 40 y 60 por ciento de ADN de descendencia (en relación con el de la madre). La línea celular de descendencia tiene trisomía en el cromosoma 21.

La Figura 1 presenta el ajuste de parámetros del modelo para (b), el 40 por ciento de la muestra. En el eje de las x se muestra el número total de alelos en el canal de interés, $m + \delta k$. Estos valores oscilan entre cero y cuatro. Considerando el canal x, no se esperan alelos en el contexto BB|BB, se espera un intervalo de cuatro alelos en el contexto AA|AA, siendo dos del ADN de la madre y dos del ADN del niño. El eje de las x mide la respuesta de plataforma en función del número de alelos. Los círculos son la media de los contextos medidos (9 en cada canal para cada uno de los cromosomas disómicos supuestos) y la línea presenta el valor correspondiente previsto por los parámetros de modelo p^* , para el mismo número de alelos. Advértase que los valores del eje de las x en los dos gráficos son bastante diferentes, lo que demuestra que las respuestas del canal x e y deben modelizarse por separado.

La Figura 2 presenta los 18 componentes de la medición y_{16} del cromosoma 16 en la muestra con 40 por ciento de ADN fetal. Las primeras nueve mediciones son del canal x y las siguientes nueve mediciones son del canal y. Los contextos se ordenan del siguiente modo: AA|AA, AA|AB, AA|BB, AB|AA, AB|AB, AB|BB, BB|AA, BB|AB, BB|BB. Se comparan las 18 mediciones con los valores previstos para las tres hipótesis H11, H12 y H21. Está claro que los datos encajan muy estrechamente con la hipótesis H11 (disomía). Se produjo la denominación correcta con el algoritmo, asignando la probabilidad de 1.0 sobre la base de una distribución previa uniforme.

La Figura 3 presenta el cromosoma 21 que corrobora H21. Se realizó también la denominación correcta asignando la probabilidad 1.0. El conjunto completo de denominaciones de hipótesis y probabilidades asignadas se presenta en la Tabla 1. Las mediciones de la media de contextos para los cromosomas clasificados para las muestras (a), (b) y (c), se presentan en las Tablas 2, 3 y 4, respectivamente. En estas tablas, las columnas se corresponden con los cromosomas y las filas se corresponden con las mediciones de la media de contextos ordenados tal como se describe para la Figura 2 por canal y después por contexto.

Tabla 1: denominaciones de hipótesis de algoritmos y probabilidades asignadas a cada cromosoma clasificado. La hipótesis correcta para cada cromosoma se presenta en la columna de encabezamiento

Muestra	ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H21)	ch22 (H11)
30 % niño	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H21 (1,0)	H11 (1,0)
40 % niño	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H21 (1,0)	H11 (1,0)
60 % niño	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H11 (1,0)	H21 (1,0)	H11 (1,0)

Tabla 2: Medias de los contextos para muestra (a) con 30 por ciento de ADN de niño

ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H21)	ch22 (H11)
13391,0	13396,3	12737,6	12610,0	14139,3	13669,2	13319,9
12720,2	12986,3	12257,6	11849,5	13484,8	12696,4	13033,4
12474,1	12259,0	11153,3	11295,9	13145,1	13000,0	11616,9
10096,4	10076,4	9118,7	10133,4	10396,8	10062,3	10119,0
9231,1	9342,0	8880,2	8523,3	9370,7	9120,1	8874,9
8133,1	7753,6	7552,1	7611,9	8629,2	8504,7	7635,7
4809,1	4907,5	4522,1	4103,8	4723,6	4222,9	4482,7
2778,9	2989,5	2708,7	2926,5	3029,2	2678,2	2813,7
932,0	924,6	915,3	930,7	936,7	955,1	921,0
1530,8	1520,7	1452,7	1514,7	1557,4	1467,0	1507,1
4897,6	4880,0	4428,4	4652,7	5139,2	4370,2	4880,5
7991,4	7858,6	7259,3	7149,2	8376,7	7583,1	7284,4
12680,9	12625,4	12151,6	12257,	13690,0	13237,8	12740,3

ES 2 640 776 T3

ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H21)	ch22 (H11)
14408,8	14339,7	14093,1	13331,4	14892,0	14312,9	14088,5
15857,3	15468,9	14785,9	15413,2	16198,5	15942,9	15458,1
19034,9	19216,9	18527,3	17459,5	19282,5	19296,2	18271,1
19949,6	20399,3	19282,9	19156,4	21229,8	20667,8	20356,4
20759,7	20659,5	19992,0	19859,9	21461,8	21454,6	20586,5

Tabla 3: Medias de los contextos para muestra (b) con 40 por ciento de ADN de niño

ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H21)	ch 22
12550,6	12579,9	11992,0	11803,7	13301,4	13019,7	12486,2
11761,7	12087,9	11409,0	10959,7	12553,3	11961,6	12050,4
11470,2	11190,0	10345,8	10184,9	11997,4	12311,9	10620,9
9730,8	9724,5	8820,5	9744,3	10006,4	9796,9	9635,1
8663,3	8788,1	8336,2	8028,7	8812,4	8652,7	8313,7
7374,2	6921,7	6858,2	6788,7	7869,5	7931,6	6872,3
5240,3	5318,0	4882,5	4355,9	5226,3	4499,0	4848,2
2876,9	3081,3	2786,8	3032,0	3133,0	2721,5	2884,9
747,0	739,0	725,6	742,6	754,5	775,7	735,5
1215,7	1202,2	1139,8	1195,3	1247,6	1162,4	1195,4
5141,4	5141,1	4647,9	4884,0	5419,2	4491,4	5066,3
8788,4	8589,7	8054,1	7676,0	9180,7	8239,9	7860,0
11789,9	11721,3	11387,4	11299,3	12930,0	12800,3	11749,7
14003,2	13876,2	13700,7	12946,6	14557,5	14022,1	13684,1
15911,6	15468,6	14791,0	15421,7	16203,6	16097,6	15506,3
18219,5	18507,0	17671,1	16562,4	18738,9	18920,1	17571,3
19453,4	19745,8	18751,0	18557,6	20749,5	20360,6	19775,2
20405,2	20347,0	19747,6	19406,0	21205,3	21409,5	20303,8

Tabla 4: Medias de los contextos con muestra (c) con 40 por ciento de ADN de niño

ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H2))	ch 22
14453,4	14433,9	13747,4	13574,8	15284,4	15340,9	14299,5
13022,7	13352,4	12687,0	11874,1	13907,0	13834,7	13265,3
12211,6	11774,2	10772,2	10840,1	12664,4	13763,4	11278,3
11674,6	11651,0	10614,0	11718,0	11951,0	11943,3	11750,2
9652,1	9865,7	9356,0	9078,6	9857,3	10296,7	9339,5
7456,1	6979,7	7149,0	6720,3	8212,8	8899,0	6930,3
7521,5	7710,6	7260,6	6449,0	7607,0	6434,7	7132,0
4117,2	4415,2	4005,3	4324,8	4473,7	3862,9	4128,1
860,4	849,9	833,1	864,3	863,7	894,7	846,3
1360,4	1364,5	1275,6	1349,8	1398,1	1353,4	1345,3
7185,7	7019,5	6493,3	6789,4	7586,5	6193,1	6964,6
12481,0	12200,9	11442,4	11061,2	13030,5	11304,5	11423,5
11840,0	11758,5	11671,3	11148,7	13512,2	14165,6	11967,1
15575,7	15291,0	15286,0	14449,9	16439,0	16386,7	15302,0
19001,6	18522,8	17684,2	18689,6	19325,3	19377,0	18569,2
19158,2	19308,5	18786,6	17259,8	19780,8	21111,1	18409,1
21537,7	21797,0	20716,9	20383,8	23039,6	23429,0	21894,6

ch16 (H11)	ch17 (H11)	ch18 (H11)	ch19 (H11)	ch20 (H11)	ch21 (H2))	ch 22
23523,5	23412,2	22665,6	22280,1	24448,8	25148,6	23313,8

En una realización, la identificación de los haplotipos parentales (fase parental) puede utilizarse para estimar las localizaciones de recombinación que determinan qué haplotipos están presentes en el niño. La identificación de qué haplotipo parental está presente en cada posición en el niño determina el genotipo del niño. Esto puede tener como resultado varianzas de modelo inferiores porque las posiciones con diferentes genotipos de niño dejarán de ser un promedio. Ciertos métodos descritos en el presente documento pueden modificarse para detectar trisomías meióticas cuando ambos haplotipos de los progenitores están presentes.

Algunas realizaciones

En algunas realizaciones de la presente invención, un método para determinar el estado de ploidía de uno o más cromosomas en un individuo diana puede incluir cualquiera de las siguientes etapas y combinaciones de las mismas:

En algunas realizaciones, se pueden obtener los datos genéticos del individuo diana y de uno o más individuos emparentados. En una realización, los individuos emparentados incluyen ambos progenitores del individuo diana. En una realización, los individuos emparentados incluyen hermanos del individuo diana. En una realización, los individuos emparentados pueden incluir los progenitores y uno o más abuelos. Estos datos genéticos para individuos se pueden obtener a partir de datos *in silico*; pueden ser información de salida de un método informático diseñado para limpiar datos genéticos, o puede derivarse de otras fuentes. En algunas realizaciones de la divulgación, se pueden obtener los datos genotípicos de los progenitores y, opcionalmente, sincronizarlos, empleando los métodos que se encuentran en las tres solicitudes de patente Rabinowitz 2006, 2008 y 2009, a las que se ha hecho referencia a lo largo de la presente solicitud. Puede utilizarse cualquier número de métodos para obtener los datos genotípicos parentales siempre y cuando el conjunto de SNP medidos en la muestra mixta de ADN fetal y materno esté suficientemente solapado con el conjunto de SNP para el que se conoce ese genotipo parental.

La amplificación del ADN, proceso que transforma una pequeña cantidad de material genético en una cantidad mayor de material genético que contiene un conjunto de datos genéticos similar, se puede realizar a través de una amplia variedad de métodos, entre los que se incluyen, sin limitarse a ellos, reacción en cadena de la polimerasa (PCR), PCR mediada por ligando, PCR de cebador oligonucleótido degenerativa, amplificación por desplazamiento múltiple, técnicas de amplificación específicas de alelo, sondas de inversión molecular (MIP), sondas padlock, otras sondas circularizables y combinaciones de las mismas. Se pueden emplear muchas variantes de otros protocolos convencionales, como por ejemplo, aumentar o disminuir los períodos de tiempo de ciertas etapas del protocolo, aumentar o disminuir la temperatura de ciertas etapas, aumentar o disminuir las cantidades de diversos reactivos, etc. La amplificación del ADN transforma la muestra inicial de ADN en una muestra de ADN que es similar en el conjunto de secuencias, pero que es mucho mayor en cantidad. En algunos casos, no se requiere amplificación.

Los datos genéticos del individuo diana y/o del individuo emparentado se pueden transformar desde un estado molecular a un estado electrónico midiendo el material genético apropiado con el empleo de herramientas y/o técnicas tomadas del grupo que incluye, sin limitarse a ellas: genotipificación de micromatrices, ensayo de genotipificación de SNP TAQMAN de APPLIED BIOSCIENCE'S, el sistema de genotipificación ILLUMINA, por ejemplo la plataforma ILLUMINA BEADARRAY utilizando por ejemplo el chip 1M-DUO, una plataforma AFFYMETRIX GENOTYPING, como AFFYMETRIX 6.0 GENECHIP, matriz GENFLEX TAG de AFFYMETRIX, otras micromatrices de genotipificación. Puede utilizarse un método de secuenciación completo superior, como la secuenciación de ADN de Sanger, pirosecuenciación, la plataforma ILLUMINA SOLEXA, GENOME ANALYZER de ILLUMINA, o plataforma de secuenciación 454 de APPLIED BIOSYSTEM, plataforma de secuenciación TRUE SINGLE MOLECULE SEQUENCING de HELICOS, o cualquier otro método de secuenciación, hibridación *in situ* fluorescente (FISH), hibridación genómica de matriz comparada (CGH), otras plataformas de genotipificación de alto rendimiento y combinaciones de las mismas. Todos estos métodos, transforman físicamente los datos genéticos almacenados en una muestra de ADN en un conjunto de datos genéticos que se almacenan normalmente en el dispositivo de memoria en ruta para ser procesados.

Se puede medir cualquier dato genético del individuo pertinente analizando las sustancias tomadas de un grupo que incluye, sin limitarse a ellos: tejido diploide en bloque del individuo, una o más células diploides del individuo, una o más células haploides del individuo, uno o más blastómeros del individuo diana, material genético extracelular encontrado en el individuo, material genético extracelular del individuo encontrado en sangre materna, células del individuo encontradas en la sangre materna, uno o más embriones creados a partir de (un) gameto(s) del individuo emparentado, uno o más blastómeros tomados de dicho embrión, material genético extracelular encontrado en el individuo emparentado, material genético que se sabe que proviene del individuo emparentado y combinaciones de los mismos.

En algunas realizaciones, se puede crear un conjunto de al menos una hipótesis de estado de ploidía para cada uno de los cromosomas de interés del individuo diana. Cada una de las hipótesis del estado de ploidía puede referirse a un estado de ploidía posible del cromosoma o el segmento de cromosomas del individuo diana. El conjunto de hipótesis puede incluir alguno o todos los estados de ploidía posible que se puede esperar que tenga el cromosoma del individuo diana. Algunos de los estados de ploidía posibles pueden incluir nulisomía, monosomía, disomía, disomía uniparental, euploidía, trisomía, trisomía apareada, trisomía sin aparear, trisomía materna, trisomía paterna, tetrasomía equilibrada (2:2) tetrasomía, tetrasomía desequilibrada (3:1), otras aneuploidías y pueden implicar adicionalmente translocaciones desequilibradas, translocaciones equilibradas, translocaciones de Robertson, recombinaciones, deleciones, inserciones, cruces y combinaciones de ellos.

En algunas realizaciones, el conjunto de probabilidades determinadas puede combinarse después. Esto puede entrañar para cada hipótesis el promedio o la multiplicación de las probabilidades según se determinan mediante cada técnica y también puede implicar la normalización de la hipótesis. En algunas realizaciones, las probabilidades pueden combinarse bajo el supuesto de que son independientes. El conjunto de productos de probabilidades para cada hipótesis en el conjunto de hipótesis se produce como probabilidades combinadas de las hipótesis.

En algunas realizaciones de la divulgación, las probabilidades determinadas tal como se determinan según el método descrito en el presente documento se pueden combinar con probabilidades de otras hipótesis que se calculan a través de métodos de diagnóstico que funcionan con preceptos completamente diferentes. Por ejemplo, algunos métodos utilizados para diagnóstico prenatal implican la medición de los niveles de determinadas hormonas en la sangre materna, en los que se correlacionan dichas hormonas con diversas anomalías genéticas. Algunos ejemplos de ellos son el cribado serológico en el tercer trimestre, la prueba triple y la prueba cuádruple. Algunos métodos implican la medición de las dimensiones y otras cualidades del feto que se pueden medir mediante ecografía, por ejemplo, translucencia nucal. Algunos de estos métodos sirven para calcular la probabilidad de que el feto sea euploide, o sufra trisomía, especialmente trisomía 18 y/o trisomía 21. En el caso en el que se utilicen múltiples métodos para determinar la probabilidad de un resultado dado, no siendo definitivos ninguno de los métodos por sí mismos, es posible combinar la información dada por los métodos para obtener una predicción que sea más precisa que la de cualquiera de los métodos por separado. Por ejemplo, en la prueba triple, la combinación de la información dada por las tres hormonas diferentes puede tener como resultado una predicción de las anomalías genéticas que sea más precisa que la que pudiera predecir los niveles de cualquiera de las hormonas por separado. En algunas realizaciones, el método implica la medición de los niveles de alfa-fetoproteína (AFP) en la sangre materna. En algunas realizaciones el método implica la medición de los niveles de estriol sin conjugar (UE₃) en la sangre materna. En algunas realizaciones, el método puede implicar la medición de los niveles de la fracción beta de gonadotropina coriónica humana (β-hCG) en la sangre materna. En algunas realizaciones, el método puede implicar la medición de los niveles de antígeno trofoblástico invasivo (ITA) en la sangre materna. En algunas realizaciones, el método puede implicar la medición de los niveles de inhibina-A en la sangre materna. En algunas realizaciones, el método puede implicar la medición de los niveles de proteína A de plasma asociada al embarazo (PAPP-A) en la sangre materna. En algunas realizaciones, el método puede implicar la medición de otras hormonas o marcadores serológicos maternos en la sangre materna. En algunas realizaciones, se han podido hacer algunas de las predicciones utilizando otros métodos. En algunas realizaciones, algunas de las predicciones se han hecho utilizando una prueba integrada completa como por ejemplo una que combine ultrasonido con análisis de sangre aproximadamente a las 10-14 semanas del embarazo y un segundo análisis aproximadamente a las 15-20 semanas. En algunas realizaciones, el método implica la medición de la translucencia nucal (NT) del feto según se mide por ultrasonido. En algunas realizaciones, el método implica el uso de los niveles de medida de las hormonas mencionadas para hacer predicciones. En algunas realizaciones, el método implica una combinación de los métodos mencionados.

La información de salida del método descrito en el presente documento se puede combinar con la información de salida de una pluralidad de otros métodos. Existen muchas formas de combinar las predicciones, por ejemplo, es posible convertir las mediciones de hormonas en un múltiplo de la media (MoM) y a continuación, en las relaciones probables (LR). Igualmente, se pueden transformar otras mediciones en LR utilizando el modelo mixto de distribuciones NT. Las LR para NT y los marcadores bioquímicos podrían multiplicarse por la edad y el riesgo relacionado con la gestación para derivar el riesgo para varias afecciones, como trisomía 21. Los índices de detección (DR) y los índices de falso positivo (FPR) se podrían calcular tomando las proporciones con riesgo por encima de un umbral de riesgo dado.

Una realización puede implicar una situación en la que se miden cuatro niveles de hormona, conociéndose la distribución de la probabilidad en torno a dichas hormonas $p(x_1, x_2, x_3, x_4|e)$ para el caso euploide y $p(x_1, x_2, x_3, x_4|a)$ para el caso aneuploide. A continuación, es posible medir la distribución de la probabilidad para las mediciones de ADN, $g(y|e)$ y $g(y|a)$ para los casos euploide y aneuploide, respectivamente. Dando por supuesto que son independientes, dado el supuesto de euploide/aneuploide, se podría combinar como $p(x_1, x_2, x_3, x_4|a)g(y|a)$ y $p(x_1, x_2, x_3, x_4|e)g(y|e)$ y multiplicar después cada uno por los $p(a)$ y $p(e)$ previos dada la edad maternal. Después, es posible seleccionar el caso de la probabilidad máxima. En una realización, es posible evocar el teorema de límite central para suponer que la distribución en $g(y|a)$ o e es gaussiana y medir la media y las desviaciones típicas examinando múltiples muestras. En otra realización, se podría suponer que no son independientes dado el resultado y recoger suficientes muestras como para estimar la distribución conjunta $p(x_1, x_2, x_3, x_4|a)$ o e .

En una realización, el estado de ploidía para el individuo diana se determina como el estado de ploidía que está asociado con la hipótesis cuya probabilidad es la máxima. En algunos casos, una hipótesis tendrá una probabilidad combinada normalizada superior a 90 %. Cada hipótesis está asociada con un estado de ploidía o conjunto de estados de ploidía y se puede seleccionar el estado de ploidía asociado con la hipótesis cuya probabilidad combinada normalizada es superior a 90 %, o algún otro valor umbral como 50 %, 80 %, 95 %, 98 %, 99 % o 99,9 %, como el umbral requerido para denominar una hipótesis como estado de ploidía determinado.

En algunas realizaciones, la información conocida del estado de ploidía determinado puede utilizarse para adoptar una decisión clínica. Dicha información, se almacena normalmente como una organización de materia física en un dispositivo de memoria y después se puede transformar en un informe. En función del informe, se puede actuar. Por ejemplo, la decisión clínica puede ser interrumpir el embarazo; alternativamente, la decisión clínica puede ser continuar el embarazo. En algunas realizaciones, la decisión clínica puede implicar una intervención diseñada para disminuir la gravedad de la presentación fenotípica del trastorno genético.

En algunos casos, puede ser deseable incluir un gran número de individuos emparentados en el cálculo para determinar el estado genético más probable. En algunos casos es posible que no sea viable aplicar el algoritmo a todos los individuos emparentados deseables debido a los límites de la potencia computacional o el tiempo. La potencia computacional necesaria para calcular los valores de alelo más probables para la diana puede aumentarse exponencialmente con el número de esperma, blastómeros y otros genotipos de entrada de los individuos emparentados. En una realización, se pueden solventar estos problemas utilizando un método denominado "subsetting" en el que se pueden dividir las computaciones en conjuntos más pequeños, ejecutarse por separado y combinarse después. En una realización de la presente divulgación, es posible tener datos genéticos de los progenitores junto con los de diez embriones y diez espermatozoides. En esta realización, es posible ejecutar varios sub-algoritmos más pequeños, por ejemplo, tres embriones y tres espermatozoides y después agrupar los resultados. En una realización, el número de embriones hermanos utilizados en la determinación pueden ser de uno a tres, de tres a cinco, de cinco a diez, de diez a veinte o más de veinte. En una realización, el número de espermatozoides cuyos datos genéticos son conocidos puede ser de uno a tres, de tres a cinco, de cinco a diez, de diez a veinte o más de veinte. En una realización cada cromosoma puede dividirse en dos a cinco, cinco a diez, diez a veinte o más de veinte subconjuntos.

En una realización de la divulgación, es posible modificar cualquiera de los métodos descritos en el presente documento para permitir que provengan múltiples dianas del mismo individuo diana, por ejemplo, varias extracciones de sangre de la misma madre embarazada. Esto puede mejorar la precisión del modelo, ya que las múltiples mediciones genéticas pueden proporcionar más datos con los que se puede determinar el genotipo diana. En una realización, un conjunto de datos genéticos sirvió como datos primarios notificado y los otros sirvieron como datos para la doble comprobación de los datos genéticos diana primarios. En una realización, se considera en paralelo una pluralidad de conjuntos de datos genéticos, medidos cada uno de ellos a partir del material genético tomado del individuo diana, y de esta forma ambos conjuntos de datos genéticos diana sirven para ayudar a determinar qué secciones de los datos genéticos parentales, medidos con alta precisión, componen el genoma fetal.

En algunas realizaciones, la fuente del material genético utilizada para determinar el estado genético del feto puede ser células fetales, como por ejemplo glóbulos rojos nucleados del feto, aislados de la sangre materna. El método puede implicar la obtención de una muestra de sangre de la madre embarazada. El método puede implicar el aislamiento de un glóbulo rojo fetal utilizando técnicas visuales, basadas en la idea de que cierta combinación de colores está asociada de forma única con glóbulos rojos nucleados y una combinación similar de colores no está asociada con ninguna otra célula presente en la sangre materna. La combinación de colores asociada con los glóbulos rojos nucleados puede incluir el color rojo de la hemoglobina en torno al núcleo, color que puede hacerse más diferenciado por manchado, y el color del material nuclear, que se pueden manchar, por ejemplo azul. Al aislar las células de la sangre materna y extenderlas sobre un portaobjetos e identificar después los puntos en los que se ve rojo (de la hemoglobina) y azul (del material nuclear) es posible identificar la localización de glóbulos rojos nucleados. A continuación, es posible extraer los glóbulos rojos nucleados utilizando un micromanipulador, el uso de genotipificación y/o técnicas de secuenciación para medir aspectos del genotipo del material genético en dichas células.

En una realización, es posible manchar los glóbulos rojos nucleados con un colorante que solamente emite fluorescencia en presencia de hemoglobina fetal y no hemoglobina materna y eliminar de este modo la ambigüedad de si los glóbulos rojos nucleados provienen de la madre o del feto. Algunas realizaciones de la presente invención pueden implicar el manchado o marcado de otra forma del material nuclear. Algunas realizaciones de la presente divulgación pueden implicar el marcado del material nuclear fetal específicamente utilizando anticuerpos específicos de célula fetal.

Existen muchas otras formas de aislar células fetales de sangre materna, o ADN fetal de sangre materna, o de enriquecer muestras de material genético materno en presencia de material genético materno. Algunos de estos métodos se enumeran en el presente documento, pero no se pretende que sea una lista exhaustiva. En el presente documento, se enumeran algunas técnicas apropiadas para mayor conveniencia: utilización de anticuerpos

etiquetados con fluorescencia o de otra forma, cromatografía por exclusión de tamaño, etiquetas de afinidad marcadas magnéticamente o de otra forma, diferencias epigenéticas, como metilación diferencial entre las células maternas y fetales en alelos específicos, centrifugación en gradientes de densidad seguido de empobrecimiento de CD45/14 y selección de CD17 positivo de células CD45/14 positivas, gradientes Percoll simples y dobles con diferentes osmolalidades o método de lectina específico de galactosa.

En una realización de la presente divulgación, el individuo diana es un feto y se realizan diferentes mediciones del genotipo en una pluralidad de muestras de ADN desde el feto. En algunas realizaciones de la divulgación, las muestras de ADN fetal son de células fetales aisladas, pudiéndose mezclar las células fetales con células maternas. En algunas realizaciones de la invención, las muestras de ADN fetal son de ADN fetal que flota libremente, pudiéndose mezclar el ADN fetal con ADN materno que flota libremente. En algunas realizaciones, el ADN fetal puede mezclarse con ADN materno en relaciones comprendidas entre 99,9: 0,1 % y 90:10 %; 90:10 % y 50:50 %; 50:50 % y 10:90 %; o 10:90 % y 0,1:99,9 %.

En una realización de la presente divulgación, es posible utilizar una técnica con soporte informático como las descritas en la presente invención para determinar si las células son en realidad de origen fetal o no. En una realización de la presente divulgación, es posible utilizar una técnica con soporte informático como las descritas en la presente divulgación para determinar el estado de ploidía de uno, o un conjunto, de cromosomas en dichas células. En una realización de la presente divulgación, es posible utilizar después una técnica con soporte informático, como las que se han descrito en la presente divulgación para determinar el estado genético de las células. Cuando se aplican a los datos genéticos de la célula PARENTAL SUPPORT™ podría indicar si el glóbulo rojo nucleado es o no de origen fetal o materno identificando si la célula contiene o no un cromosoma de la madre y uno del padre, lo que podría indicar que es fetal, o dos cromosomas de la madre, lo que podría indicar que es materno.

En una realización, puede utilizarse el método para realizar pruebas de paternidad. Por ejemplo, dada la información genotípica basada en los 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 representa realmente el padre genético real del feto en gestación. Una forma sencilla de hacerlo consiste en examinar simplemente los contextos en los que la madre es AA y el posible padre es AB o BB. En estos casos, es posible esperar ver la contribución del padre la mitad de las veces (AA|AB) o todas (AA|BB) respectivamente. Teniendo en cuenta la ADO esperada, es fácil determinar si los SNP fetales que se observan se correlacionan o no con los del posible padre.

Una realización de la presente divulgación sería la siguiente: una mujer embarazada desea saber si su feto sufre síndrome de Down y/o si padecerá fibrosis quística y no desea dar a luz un niño que sufra ninguna de estas afecciones. El médico le extrae sangre y mancha la hemoglobina con un marcador de manera que aparece claramente rojo y mancha el material nuclear con otro marcador de manera que aparezca claramente azul. Sabiendo que los glóbulos rojos maternos carecen normalmente de núcleo, mientras que una alta proporción de células fetales contienen núcleo, puede aislar visualmente una serie de glóbulos rojos nucleados identificando las células que presentan tanto color rojo como azul. El médico recoge estas células desprendiéndolas desde el portaobjetos con un micromanipulador y las envía a un laboratorio que amplifica y genotipifica diez células individuales. Al utilizar las mediciones genéticas, el método PARENTAL SUPPORT™ puede determinar que seis de las diez células son células de sangre materna y cuatro de las diez células son células fetales. Si el niño ha nacido ya de la madre embarazada, puede utilizarse PARENTAL SUPPORT™ para determinar que las células fetales son distintas de las células del niño nacido obteniendo denominaciones de alelo fiables sobre las células fetales y demostrando que son distintas a las del niño nacido. Advuértase que el concepto de este método es similar a la realización de prueba de paternidad de la divulgación. Los datos genéticos medidos a partir de células fetales pueden ser de una calidad muy pobre, al contener muchas pérdidas de alelo, como consecuencia de la dificultad de genotipificar células únicas. El especialista clínico es capaz de utilizar el ADN fetal medido junto con la medición de los ADN de los progenitores fiable para deducir aspectos del genoma del feto con una alta precisión utilizando PARENTAL SUPPORT™, transformando así los datos genéticos contenidos en el material genético del feto en el estado genético previsto, almacenados en un ordenador. El especialista clínico puede determinar tanto el estado de ploidía del feto como la presencia o ausencia de una pluralidad de genes de interés ligados a enfermedad. El resultado es que el feto tiene euploidía y no es portador de fibrosis quística y la madre decide continuar el embarazo.

En otra realización, una pareja en la que la madre, que está embarazada y tiene una edad materna avanzada, desea saber si el feto en gestación tiene o no síndrome de Down o alguna otra anomalía cromosómica. El obstetra extrae sangre de la madre y el padre. Un técnico centrifuga la muestra materna para aislar el plasma y la capa leucocítica. Se transforman ADN en la capa leucocítica y la muestra de sangre paterna por amplificación y se transforman de nuevo los datos genéticos codificados por el material genético ampliado de datos genéticos almacenados molecularmente en datos genéticos almacenados electrónicamente ejecutando el material genético en una matriz de SNP para medir los genotipos parentales. La muestra de plasma puede tratarse posteriormente a través de un método como el de hacer correr un gel, o el uso de una columna de exclusión por tamaño, para aislar fracciones de ADN de tamaño específico. Se pueden utilizar otros métodos para enriquecer la fracción de ADN fetal en la muestra. Se puede emplear una técnica con soporte informático que incluye la divulgación descrita en el presente documento,

como PARENTAL SUPPORT™ para determinar el estado de ploidía del feto. Se determina que el feto tiene síndrome de Down. Se imprime un informe, o se envía por correo electrónico al obstetra de la mujer embarazada, que transmite el diagnóstico a la mujer. La mujer, su marido y el doctor, se reúnen y discuten las opciones. La pareja decide interrumpir el embarazo sobre la base de la información conocida de que el feto sufre un estado trisómico.

5 En otra realización, una mujer embarazada, en adelante denominada "la madre" puede decidir que quiere saber si su feto(s) es (son) portador(es) o no de anomalías genéticas y otras afecciones. Es posible que quiera asegurarse de si existe alguna anomalía importante antes de tener la seguridad de continuar el embarazo. Es posible que acuda al médico especialista en obstetricia, que puede extraerle una muestra de sangre. Asimismo, puede tomar una muestra genética, como por ejemplo, una torunda bucal desde la mejilla. También es posible que tome una muestra genética del padre del feto, como por ejemplo una torunda bucal, una muestra de esperma o una muestra de sangre. El médico puede enriquecer la fracción el ADN fetal que flota libremente en la muestra de sangre materna. El médico puede enriquecer la fracción de glóbulos rojos fetales nucleados en la muestra de sangre materna. El médico puede utilizar diversos aspectos del método descrito en el presente documento para determinar los datos genotípicos del feto. Dichos datos genotípicos pueden incluir el estado de ploidía del feto y/o la identidad de uno o una serie de alelos en el feto. Se puede generar un informe en el que se resuman los resultados del diagnóstico prenatal. El médico puede comunicarle a la madre el estado genético del feto. La madre puede decidir interrumpir el embarazo sobre la base del hecho de que el feto tiene una o más anomalías cromosómicas o genéticas o afecciones no deseables. Asimismo, puede decidir continuar el embarazo sobre la base del hecho de que el feto no tiene ninguna anomalía cromosómica o genética importante, ni ninguna afección genética de interés.

Otro ejemplo puede implicar una mujer embarazada que ha sido inseminada artificialmente con esperma de donante y está embarazada. Desea reducir al mínimo el riesgo de que el feto sea portador de una enfermedad genética. Acude a un flebotomista para extraerse sangre y se aplican las técnicas descritas en la presente divulgación para aislar tres glóbulos rojos fetales nucleados y se recoge también una muestra de tejido de la madre y el padre genético. Se amplifican material genético del feto y de la madre y el padre según sea apropiado y se genotifican utilizando el ILLUMINA INFINIUM BEADARRAY y con los métodos descritos en el presente documento se limpia y se sincroniza el genotipo parental y fetal con una alta precisión, así como para obtener denominaciones de ploidía para el feto. Se encuentra que el feto es euploide y se predicen las susceptibilidades fenotípicas a partir del genotipo fetal reconstruido y se genera un informe, que se envía al médico de la madre para que puedan decidir qué decisión clínica puede ser la mejor.

Otro ejemplo puede implicar una mujer que está embarazada, pero que al haber tenido más de una pareja sexual, no está segura de la paternidad de su feto. La mujer desea conocer quién es el padre genético del feto que lleva. Ella y una de sus parejas sexuales acuden al hospital y ambos donan una muestra de sangre. El especialista clínico utilizando los métodos descritos en la presente invención es capaz de determinar la paternidad del feto. El resultado es que el padre biológico del feto no es su pareja estable y, sobre la base de esta información, la mujer decide interrumpir el embarazo.

En algunas realizaciones de la presente invención, se puede cambiar una pluralidad de parámetros sin por ello cambiar la esencia de la presente divulgación. Por ejemplo, se pueden obtener los datos genéticos utilizando una plataforma de genotificación de alto rendimiento, o se pueden obtener a partir de cualquier método de genotificación, o se pueden simular, deducir o conocer de otra forma. Se podrían utilizar diversos lenguajes computacionales para codificar los algoritmos descritos en la presente invención y se podría utilizar diversas plataformas computacionales para ejecutar los cálculos. Por ejemplo, los cálculos se podrían ejecutar utilizando ordenadores personales, superordenadores y ordenadores paralelos.

En algunas realizaciones de la invención, se puede implementar el método en un circuito electrónico digital o en un hardware de ordenador, firmware, software o combinaciones de ellos. Los aparatos de la divulgación pueden implantarse en producto de programa informático incorporable tangiblemente en un dispositivo de almacenamiento legible por máquina para ejecutarlo con un procesador programable; y las etapas del método de la presente invención se pueden realizar a través de un procesador programable que ejecute un programa de instrucciones para realizar las funciones de la invención operando sobre datos de entrada y generando datos de salida. La invención se puede implementar ventajosamente en uno o más programas informáticos que se pueden ejecutar en un sistema programable que incluye al menos un procesador programable acoplado para recibir datos e instrucciones y para transmitir datos e instrucciones, un sistema de almacenamiento de datos, al menos un dispositivo de entrada y al menos un dispositivo de salida. Cada programa informático se puede implementar en un lenguaje de programación orientado al objeto o procedural de alto nivel o en conjunto con lenguaje máquina, si se desea; y en cualquier caso, el lenguaje puede ser un lenguaje compilado o interpretado. Entre los procesadores adecuados se incluyen a modo de ejemplo, tanto microprocesadores generales como especializados. Generalmente, un procesador recibirá instrucciones y los datos de una memoria solo de lectura y/o una memoria de acceso aleatorio. Generalmente, el ordenador incluirá uno o más dispositivos de almacenamiento en masa para almacenar archivos de datos; dichos dispositivos incluyen discos magnéticos, como por ejemplo discos duros internos y discos extraíbles; discos magneto-ópticos; y discos ópticos. Los dispositivos de almacenamiento adecuados para incorporar tangiblemente las instrucciones del programa informático y los datos incluyen todas las formas de memoria no volátil, incluyendo a modo de ejemplo dispositivos de memoria semiconductores como EPROM, EEPROM y dispositivos de memoria

instantánea; discos magnéticos como discos duros internos y discos extraíbles, discos magneto-ópticos; discos CD-ROM. Cualquiera de los anteriores puede suplementarse o llevar incorporado ASIC (circuitos integrados de aplicación específica). Los resultados se pueden sacar en forma de un informe impreso, o desplegar en la pantalla, o se pueden salvar con un dispositivo de memoria que implica el almacenamiento de información por medio de un cambio físico en el sustrato del dispositivo de memoria, como por ejemplo, los que se han enumerado anteriormente. Se puede generar un informe que describa la determinación del estado de ploidía del feto, ya sea impreso o electrónicamente que transmita la información al facultativo sanitario y/o el progenitor. Se puede adoptar una decisión clínica sobre la base de la determinación. En algunas realizaciones, se puede adoptar la decisión clínica de interrumpir un embarazo en la contingencia de que el feto sea aneuploide; el rechazo de una afección de aneuploidía en el feto proporciona la base para decidir la interrupción del embarazo y también puede incluir la ejecución de dicha decisión.

En una realización, el material genético en bruto de la madre y el padre se transforma por amplificación en una cantidad de ADN que es similar en secuencia, pero de mayor cantidad. A continuación, a través de un método de genotipificación, se transforman los datos genotípicos codificados por ácidos nucleicos en mediciones genéticas que se pueden almacenar físicamente y/o electrónicamente en un dispositivo de memoria, como los que se han descrito. Los algoritmos pertinentes que componen el algoritmo PARENTAL SUPPORT™ cuyas partes correspondientes se explican en detalle en la presente invención, se traducen en un programa informático utilizando un lenguaje de programación. A continuación, a través de la ejecución del programa informático en el hardware del ordenador, en lugar de codificar físicamente bits y bytes, dispuestos en un patrón que representa los datos de medición en bruto, se transforman en un patrón que representa una determinación de alta fiabilidad del estado de ploidía del feto. Los detalles de esta transformación se basarán en los propios datos y el lenguaje informático y el sistema de hardware utilizado para ejecutar el método descrito en el presente documento, si bien es predecible si los contextos son conocidos. A continuación, los datos que se configuran físicamente para representar una determinación de ploidía de alta calidad del feto se transforman en un informe que puede ser enviado al facultativo sanitario. Esta transformación se puede llevar a cabo utilizando una impresora o una pantalla de ordenador. El informe puede ser una copia impresa sobre papel u otro soporte adecuado, o puede ser electrónico. En el caso de un informe electrónico, se puede transmitir, se puede almacenar físicamente en el dispositivo de memoria en una localización del ordenador accesible al facultativo sanitario; también puede aparecer en pantalla para poderlo leer. En el caso de que aparezca en pantalla, los datos se pueden transformar a un formato legible provocando la transformación física de los píxeles en el dispositivo de pantalla. La transformación se puede llevar a cabo disparando físicamente los electrones en una pantalla fosforescente, alterando una carga eléctrica que cambia físicamente la transparencia de un ajuste de píxeles específico en la pantalla que puede estar situada en frente de un sustrato que emite o absorbe fotones. Esta transformación puede llevarse a cabo cambiando la orientación de nanoescala de las moléculas en un cristal líquido, por ejemplo, desde la fase nemática a la fase colestérica o esméctica, en un conjunto de píxeles específico. Esta transformación se puede llevar a cabo a través de una corriente eléctrica haciendo que los protones sean emitidos desde un conjunto de píxeles específico hecho de una pluralidad de diodos que emiten luz dispuestos en un patrón organizado. Esta transformación se puede llevar a cabo de cualquier forma utilizada para desplegar información, como por ejemplo una pantalla de ordenador, o cualquier otro dispositivo de salida o por transmisión de la información. El facultativo sanitario puede actuar en función de dicho informe, de manera que los datos de informe se transformen en una acción. La acción puede ser continuar o interrumpir el embarazo, en cuyo caso, un feto en gestación con una anomalía genética se transforma en un feto no vivo. Las transformaciones enumeradas en el presente documento se pueden agregar de manera que por ejemplo sea posible transformar el material genético de una madre embarazada y el padre, a través de una serie de etapas señaladas en la presente divulgación, en una decisión médica que consiste en abortar un feto con anomalías genéticas o que consista en continuar el embarazo. Alternativamente, es posible transformar un conjunto de mediciones genotípicas en un informe que ayude al médico a tratar a su paciente embarazada.

En una realización de la divulgación, el método descrito en el presente documento puede utilizarse para determinar el estado de ploidía de un feto incluso en los casos en los que la madre huésped, es decir, la mujer que está embarazada, no es la madre biológica del feto que lleva.

Algunos de los cálculos matemáticos en la presente divulgación plantean hipótesis concernientes a un limitado número de estados de aneuploidía. En algunos casos, por ejemplo, se espera que solamente cero, uno o dos cromosomas provengan de cada uno de los progenitores. En algunas realizaciones de la presente divulgación, las derivaciones matemáticas se pueden extender para tener en cuenta otras formas de aneuploidía, como cuadrosomía, en las que tres cromosomas provienen de uno de los progenitores, pentasomía, hexasomía, etc., sin cambiar los conceptos fundamentales de la presente divulgación. Al mismo tiempo, es posible centrarse en un número más reducido de estados de ploidía, por ejemplo, solamente trisomía y disomía. Adviértase, que las determinaciones de ploidía que indican un número no entero de cromosomas puede indicar mosaicismo en una muestra de material genético.

En algunas realizaciones de la presente divulgación, un individuo emparentado puede referirse a cualquier individuo que está genéticamente relacionado y por tanto comparte bloques de haplotipo con el individuo diana. Algunos ejemplos de individuos emparentados incluyen: padre biológico, madre biológica, hijo, hija, hermano, hermana, semi-hermano, semi-hermana, abuelo, abuela, tío, tía, sobrino, sobrina, nieto, nieta, primo, clon, el propio individuo

diana/la propia individuo diana y otros individuos con relación genética conocida con la diana. El término “individuo emparentado” también abarca cualquier embrión, feto, esperma, huevo, blastómero, blastoquiste o cuerpo polar proveniente de un individuo emparentado.

5 En algunas realizaciones de la presente divulgación, el individuo diana puede referirse a un adulto, un joven, un feto, un embrión, un blastoquiste, un blastómero, una célula o un conjunto de células de un individuo o de una línea celular o cualquier conjunto de material genético. El individuo diana puede estar vivo, muerto, congelado o en estasis. En algunas realizaciones de la presente divulgación, teniendo en cuenta que todas las criaturas vivas o que han vivido contienen datos genéticos, los métodos se pueden aplicar igualmente a cualquier ser humano, animal o planta, vivos o muertos, que hereden o hayan heredado cromosomas de otros individuos.

10 Asimismo, es importante señalar que los datos genéticos fetales que se pueden generar midiendo el ADN amplificado de una pequeña muestra de ADN fetal se pueden utilizar para múltiples propósitos. Por ejemplo, se pueden utilizar para detectar aneuploidía, disomía uniparental, translocaciones desequilibradas, el sexo del individuo, así como para obtener una pluralidad de predicciones fenotípicas en alelos asociados a fenotipo. En algunas relaciones, se puede poner el punto de mira en afecciones genéticas en particular antes del cribado y si ciertos loci son especialmente relevantes para dichas afecciones genéticas, entonces, se puede seleccionar un conjunto de SNP más apropiado que tengan mayor probabilidad de co-segregarse con el locus de interés, aumentando así la fiabilidad de las denominaciones de alelo de interés.

15 En algunas realizaciones, la anomalía genética es un tipo de aneuploidía, como por ejemplo síndrome de Down (o trisomía 21), síndrome de Edwards (trisomía 18), síndrome de Patau (trisomía 13), síndrome de Turner (45x0) y síndrome de Klinefelter (un individuo de sexo masculino con 2 cromosomas X). Los trastornos congénitos como los que se han enumerado en la frase anterior, son normalmente rechazados y el conocimiento de que el feto sufre una o más anomalías fenotípicas puede servir como base para decidir interrumpir el embarazo.

20 En algunas realizaciones, la anomalía fenotípica puede ser una malformación de una extremidad, o un defecto del tubo neural. Entre las malformaciones de las extremidades se pueden incluir sin limitarse a ellas, amelia, ectrodactilia, focomelia, polimelia, polidactilia, sindactilia, polisindactilia, oligodactilia, braquidactilia, acondroplasia, aplasia congénita o hipoplasia, síndrome de bandas anmióticas y disostosis cleidocraneal

25 En algunas realizaciones, la anomalía fenotípica puede ser una malformación congénita del corazón. Entre las malformaciones congénitas del corazón se pueden incluir sin limitarse a ellas, ductus arterioso persistente, defecto del septo atrial, defecto del septo ventricular y tetralogía de Fallot.

30 En algunas realizaciones, la anomalía fenotípica puede ser una malformación congénita del sistema nervioso. Las malformaciones congénitas del sistema nervioso incluyen, sin limitarse solo a ellas, defectos del tubo neural (p.ej., espina bífida, meningocele, menigomielocelo, encefalocele y anencefalia), malformación de Arnold-Chiri, malformación de Dandy-Walker, hidrocefalia, microcefalia, megencefalia, lisencefalia, polimicrogiria, holoprosencefalia y agenesia del cuerpo caloso.

35 En algunas realizaciones, la anomalía fenotípica puede ser una malformación congénita del sistema gastrointestinal. Las malformaciones congénitas del sistema gastrointestinal incluyen, sin limitarse solo a ellas, estenosis, atresia y ano imperforado.

40 En algunas realizaciones, la anomalía genética es monogénica o multigénica. Las enfermedades genéticas incluyen, sin limitarse solo a ellas, síndrome de Bloom, enfermedad de Canavan, fibrosis quística, disautonomía familiar, síndrome de Riley-Day, anemia de Fanconi (grupo C), enfermedad de Gaucher, enfermedad de depósitos de glucógeno, enfermedad de orina de jarabe de arce, Mucopolidosis IV, enfermedad de Niemann-Pick, enfermedad de Tay-Sachs, talasemia Beta, anemia de células falciformes, talasemia alfa, talasemia beta, deficiencia de factor XI, ataxia de Friedreich, MCAD, enfermedad de Parkinson juvenil, Connexin26, SMA, síndrome de Rett, Fenilcetonuria, distrofia muscular de Becker, distrofia muscular de Duchennes, síndrome X frágil, Hemofilia A, inicio temprano de demencia Alzheimer, cáncer de mama ovarios, cáncer de colon, diabetes/MODY, enfermedad de Huntington, distrofia muscular miotónica, enfermedad de Parkinson – inicio temprano, síndrome de Peutz-Jeghers, enfermedad renal poliquística, distonía de torsión.

45 En algunas realizaciones, los sistemas, métodos y técnicas de la presente divulgación se utilizan en métodos para aumentar la probabilidad de implantación de un embrión obtenido por fertilización *in vitro* que esté en menor riesgo de portar una predisposición a una enfermedad genética.

50 En una realización de la presente divulgación, se describen métodos para la determinación del estado de ploidía en los que el material genético medido de la diana está contaminado con material genético de la madre, utilizando la información conocida de los datos genéticos de la madre. Esto contrasta con los métodos que pueden determinar el estado de ploidía de los datos genéticos que tienen ruido debido a mediciones insuficientes; la contaminación en dichos datos es aleatoria. Esto contrasta con los métodos que permiten determinar el estado de ploidía de un individuo diana de los datos que son difíciles de interpretar por la contaminación del ADN de individuos no

relacionados; la contaminación en esos datos es genéticamente aleatoria. En una realización, los métodos descritos en el presente documento sirven para determinar el estado de ploidía de un individuo diana cuando la dificultad de interpretación se debe a la contaminación del ADN de uno de los progenitores; la contaminación en estos datos es idéntica al menos en la mitad de a los datos diana y por lo tanto es difícil corregirlos. Para conseguir este fin, en una realización, un método de la presente divulgación aprovecha la información conocida del genotipo materno contaminante para crear un modelo de las mediciones genéticas esperadas dada una mezcla del material genético materno y de la diana, en el que no se conocen los datos genéticos de la diana a priori. Esta etapa no es necesaria cuando la incertidumbre de los datos genéticos se debe a un ruido aleatorio.

En una realización, un método para determinar el número de copias de un cromosoma de interés en un individuo diana, utilizando mediciones genotípicas obtenidas en un material genético del individuo diana, en el que se mezcla el material genético del individuo diana con el material genético de la madre del individuo diana, comprende la obtención de datos genotípicos para un conjunto de SNP de los progenitores del individuo diana; la obtención de mediciones genotípicas para un conjunto de SNP en una muestra mixta que comprende ADN del individuo diana y también ADN de la madre del individuo diana; la creación en un ordenador de un conjunto de hipótesis de estados de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las medidas genéticas de la muestra mixta y los datos genéticos de los progenitores del individuo diana; y uso de las probabilidades determinadas de cada hipótesis para determinar el número de copias más probable del cromosoma de interés en el individuo diana. En una realización, el individuo diana y los progenitores del individuo diana son sujetos de ensayo humanos.

En una realización, un método implementado por ordenador para determinar el número de copias de un cromosoma de interés en un individuo diana, empleando mediciones genotípicas obtenidas en material genético del individuo diana, en el que el material genético del individuo diana se mezcla con material genético de la madre del individuo diana, comprende la obtención de datos genotípicos para un conjunto de los SNP de los progenitores del individuo diana; la obtención de 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; la creación en un ordenador de un conjunto de hipótesis de estados de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las mediciones genéticas de la muestra mixta y de los datos genéticos de los progenitores del individuo diana; y el uso de las probabilidades determinadas de cada hipótesis para determinar el número de copias más probable del cromosoma de interés en el individuo diana.

En una realización, un método para determinar el número de copias de un cromosoma de interés de un individuo diana utilizando mediciones genotípicas obtenidas sobre material genético del individuo diana, en el que el material genético del individuo diana se mezcla con material genético de la madre del individuo diana comprende la obtención de datos genotípicos para un conjunto de SNP de la madre del individuo diana; la obtención de medidas genotípicas para un conjunto de SNP en una muestra mixta que comprende ADN del individuo diana y también ADN de la madre del individuo diana; la creación en un ordenador de un conjunto de hipótesis de estados de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las mediciones genéticas de la muestra mixta y de los datos genéticos de la madre del individuo diana; y el uso de las probabilidades determinadas de cada hipótesis para determinar el número de copias más probable del cromosoma de interés en el individuo diana.

En una realización, un método implementado por ordenador para determinar el número de copias de un cromosoma de interés de un individuo diana utilizando mediciones genotípicas obtenidas sobre un material genético del individuo diana, en el que el material genético del individuo diana se mezcla con material genético de la madre del individuo diana, comprende la obtención de datos genotípicos para un conjunto de SNP de la madre del individuo diana, la obtención de 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; la creación en un ordenador de un conjunto de hipótesis de estados de ploidía para el cromosoma de interés del individuo diana; la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las mediciones genéticas de la muestra mixta y de los datos genéticos de la madre del individuo diana; y el uso de las probabilidades determinadas de cada hipótesis para determinar el número de copias más probable del cromosoma de interés en el individuo diana.

REIVINDICACIONES

1. Un método para determinar el número de copias de un cromosoma de interés en un feto utilizando mediciones genotípicas obtenidas en ADN que flota libremente del feto encontrado en la sangre materna, en el que el ADN que flota libremente del feto se mezcla con material genético de la madre del feto, comprendiendo el método:
- 5
- obtener mediciones genotípicas para un conjunto de SNP que comprende una pluralidad de SNP del cromosoma de interés y una pluralidad de SNP de al menos un cromosoma que se espera que sea disómico en el feto en una muestra mixta que comprende el ADN que flota libremente del feto mezclado con ADN materno que flota libremente, en el que la medición del material genético se realiza en material genético que se amplifica antes de medirlo y en el que la medición genotípica se realiza en 1.000 a 20.000 SNP utilizando secuenciación de alto rendimiento;
- 10
- crear en un ordenador un conjunto de hipótesis de estados de ploidía en las que cada hipótesis de los estados de ploidía es un posible estado de ploidía del cromosoma de interés en el feto en las que el conjunto de hipótesis de los estados de ploidía comprende nueve hipótesis posibles que corresponden a diferentes estados de ploidía: H00 (0, 0); H01 (0, 1); H02 (0, 2); H10 (1, 0); H11 (1, 1); H12 (1, 2); H20 (2, 0); H21 (2, 1); y H22 (2, 2), en las que (m, f) se refieren a la hipótesis en la que m es el número de un cromosoma dado heredado de la madre y f es el número de cromosomas dado heredado del padre;
- 15
- determinar en un ordenador, la probabilidad de cada una de las hipótesis dadas las mediciones genotípicas en la muestra mixta; y
- 20
- usar las probabilidades determinadas de cada hipótesis para determinar el número de copias más probable del cromosoma de interés en el feto, en las que las probabilidades determinadas son la estimación de probabilidades de máxima probabilidad y
- 25
- en el que la muestra mixta es una muestra de sangre materna y el ADN fetal no está purificado.
2. El método de la reivindicación 1, en el que el método comprende determinar la relación entre el ADN fetal y el materno en la muestra mixta para cromosomas individuales.
- 30
3. El método de la reivindicación 1, en el que la determinación del número de copias se emplea para adoptar una decisión clínica.
4. El método de la reivindicación 3, en el que la decisión clínica es interrumpir el embarazo cuando se encuentra que el feto tiene una anomalía genética, o no interrumpir el embarazo cuando no se encuentra que el feto tiene una anomalía genética.
- 35
5. El método de la reivindicación 1, en el que el cromosoma de interés se selecciona del grupo que consiste en cromosoma 13, cromosoma 18, cromosoma 21, el cromosoma X, el cromosoma Y y las combinaciones de los mismos.
- 40
6. El método de la reivindicación 1, en el que el material genético del feto es un ADN nuclear encontrado en una o más células del feto.
7. El método de la reivindicación 1 en el que la etapa de medición del material genético se realiza en material genético que se amplifica utilizando una técnica seleccionada del grupo que consiste en reacción en cadena de la polimerasa (PCR), PCR mediada por ligando, PCR de cebador oligonucleótido degenerativa, amplificación por desplazamiento múltiple (MDA), PCR específica de alelo, técnicas de amplificación específicas de alelo, amplificación de puente, sondas padlock, sondas circularizables y combinaciones de las mismas.
- 45
8. El método de la reivindicación 1, en el que la etapa de determinar el número de copias del cromosoma de interés se realiza con fines de rastrear un estado cromosómico, en el que el estado cromosómico se selecciona del grupo que consiste en nulisomía, monosomía, disomía, disomía uniparental, euploidia, trisomía, trisomía apareada, trisomía no apareada, trisomía materna, trisomía paterna, tetrasomía, tetrasomía apareada, tetrasomía no apareada, otras aneuploidías, translocación desequilibrada, translocación equilibrada, recombinación, delección, inserción, mosaicismo y combinaciones de los mismos.
- 50
- 55
9. El método de la reivindicación 1, en el que la etapa de obtención de mediciones genotípicas comprende además la realización de una operación estadística sobre las mediciones genotípicas que actúa de forma opuesta a la tendencia de desviación del proceso de medición genotípico para corregir estadísticamente la desviación y/o aumentar la precisión de las mediciones genotípicas.
- 60
10. El método de la reivindicación 1, que comprende además la obtención de datos genotípicos para un conjunto de SNP de los progenitores del feto y la determinación en el ordenador de la probabilidad de cada una de las hipótesis dadas las mediciones genotípicas de la muestra mixta y los datos genotípicos de los progenitores del feto.
- 65

11. El método de la reivindicación 1, en el que el material genético del feto es menos de 20 % del ADN que flota libremente en la sangre o el suero.

5 12. El método de la reivindicación 1, que comprende además la obtención de probabilidades a priori de cada hipótesis de un estudio de población estadístico y la computación de la fiabilidad utilizando la regla de Bayes.

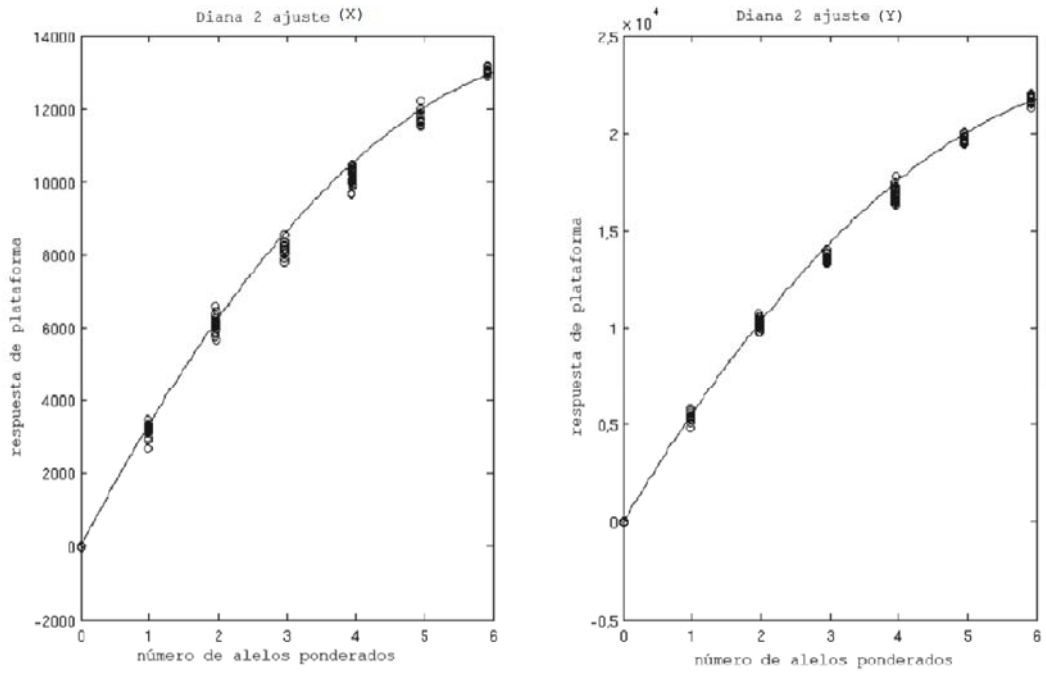


Figura 1

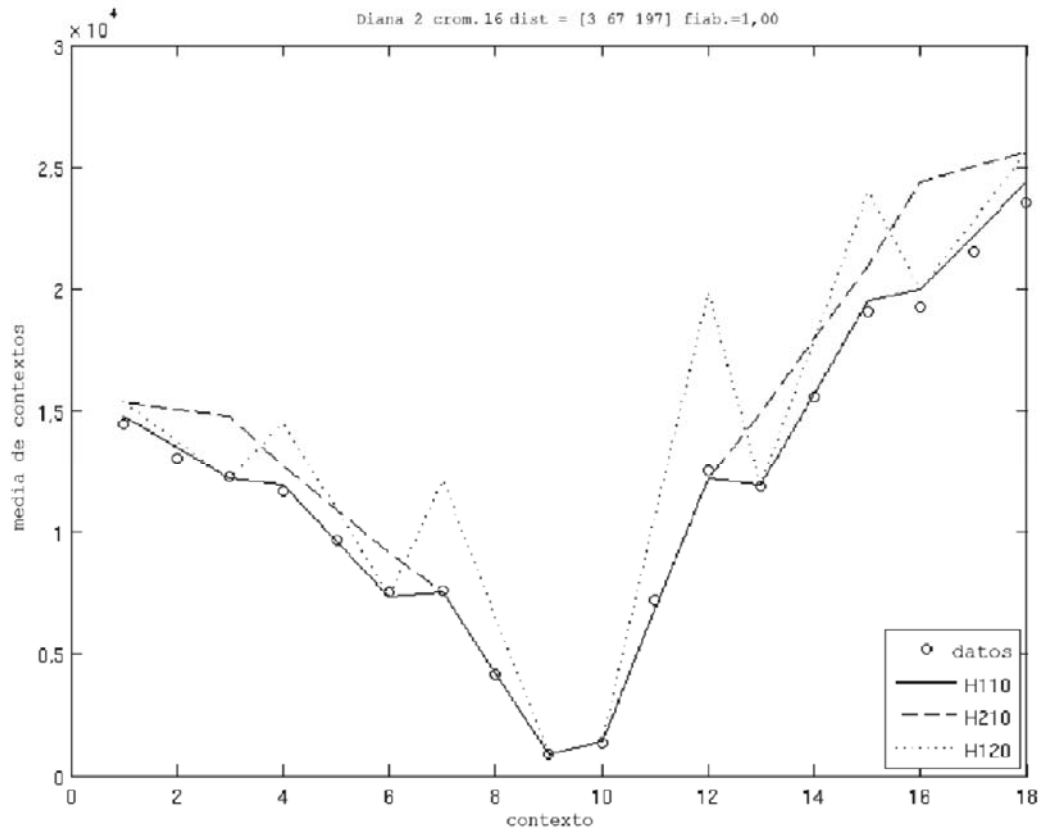


Figura 2

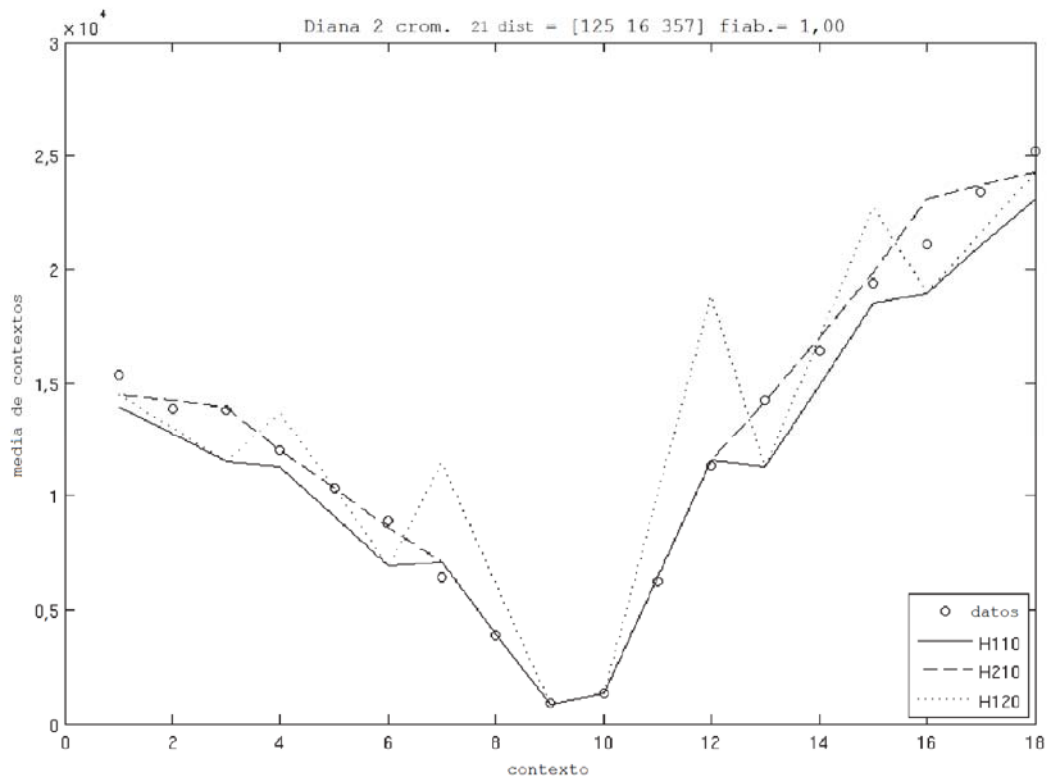


Figura 3