

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 934 941**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12Q 1/6827 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.11.2010 E 20200970 (0)**

97 Fecha y número de publicación de la concesión europea: **23.11.2022 EP 3783110**

54 Título: **Análisis genómico fetal a partir de una muestra biológica materna**

30 Prioridad:

05.11.2009 US 25856709 P

06.11.2009 US 25907509 P

10.09.2010 US 38185410 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
28.02.2023

73 Titular/es:

**THE CHINESE UNIVERSITY OF HONG KONG
(50.0%)**

**Office of Research and Knowledge Transfer
Services, Room 301 Pi Ch'iu Building Shatin
New Territories Hong Kong, HK y
SEQUENOM, INC. (50.0%)**

72 Inventor/es:

**LO, YUK MING DENNIS;
CHAN, KWAN CHEE;
CHIU, WAI KWUN ROSSA y
CANTOR, CHARLES**

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 934 941 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

Análisis genómico fetal a partir de una muestra biológica materna

5 La presente invención se refiere en general al análisis de un genoma fetal basado en una muestra materna, y más particularmente a la determinación de todo o partes del genoma fetal basándose en un análisis de fragmentos genéticos en la muestra materna.

10 El descubrimiento de ácidos nucleicos fetales libres de células en el plasma materno en 1997 abrió nuevas posibilidades para el diagnóstico prenatal no invasivo (Lo YMD et al, Lancet, 1997; 350: 485-487; y Patente de EE.UU. 6.258.540). Esta tecnología se ha traducido rápidamente en aplicaciones clínicas, con la detección de genes o secuencias heredadas paternalmente derivadas del feto, p. ej. para la determinación fetal del sexo, la determinación del estado fetal del RhD y la determinación de si el feto ha heredado una mutación heredada paternalmente (Amicucci P et al, Clin Chem, 2000; 46: 301-302; Saito H et al, Lancet, 2000; 356: 1170; y Chiu RWK et al, Lancet, 2002; 360: 998-1000). El progreso reciente en el campo ha permitido el diagnóstico prenatal de aneuploidías cromosómicas fetales, tales como la trisomía 21, a partir del análisis de ácidos nucleicos de plasma materno, Lo YMD et al., Nat Med, 2007; 13: 218-223; Tong YK et al, Clin Chem, 2006; 52: 2194-2202; Publicación de patente de EE.UU. 2006/0252071; Lo YMD y otros, Proc Natl Acad Sci USA, 2007; 104: 131 16-13121; Chiu RWK y otros, Proc Natl Acad Sci USA, 2008; 105: 20458-20463; Fan HC y otros, Proc Natl Acad Sci, 2008; 105: 16266-16271; Publicación de patente de EE.UU. 2007/0202525; y Publicación de patente de EE.UU. 2009/0029377).

25 Otra área de progreso reciente significativo es el uso de métodos de conteo de una sola molécula, tal como la PCR digital, para el diagnóstico prenatal no invasivo de enfermedades monogénicas en las que la madre y el padre llevan ambos la misma mutación. Esto se ha conseguido mediante el análisis de la dosis de mutación relativa (RMD) en el plasma materno (Solicitud de Patente de EE.UU. 2009/0087847; Lun FMF y otros, Proc Natl Acad Sci USA, 2008; 105: 19920-19925; y Chiu RWK y col., Trends Genet 2009; 25: 324-331).

30 El documento US 2004/137470 divulga un método no invasivo para la detección de anomalías cromosómicas en un feto. Chiu et al. (Trends in Genetics, Elsevier Science Publishers, vol. 25, n.º 7, páginas 324-331) divulga métodos analíticos para la detección no invasiva de enfermedades monogénicas fetales y aneuploidías cromosómicas.

35 Lun et al. (Proceedings of the National Academy of Sciences of the United States of America, vol. 105, n.º 50, páginas 19920-19925) divulgan un enfoque digital de dosificación relativa de mutaciones para el diagnóstico prenatal no invasivo de enfermedades monogénicas.

40 Lo (Annals of the New York Academy of Sciences, New York Academy of Sciences, Estados Unidos, 27 de agosto de 2007, páginas 140-143) analiza distintos enfoques que permiten utilizar los ácidos nucleicos fetales circulantes para la detección prenatal de aneuploidías cromosómicas.

45 Sin embargo, tales métodos utilizan el conocimiento previo de posibles mutaciones para analizar partes específicas del genoma, y por lo tanto no pueden identificar mutaciones latentes o poco comunes o enfermedades genéticas. Por lo tanto, es deseable proporcionar nuevos métodos, sistemas y aparatos que puedan identificar todo o partes de un genoma fetal usando técnicas no invasivas.

Según un aspecto de la invención se proporciona un método para determinar una concentración fraccional de ADN fetal en una muestra biológica tomada de una mujer embarazada con un feto, teniendo el feto un padre y una madre que es la mujer embarazada, teniendo la madre un genoma materno, teniendo el feto un genoma fetal, en donde la muestra biológica contiene una mezcla de ácidos nucleicos maternos y fetales. El método comprende analizar una pluralidad de moléculas de ácido nucleico de la muestra biológica, en donde el análisis de una molécula de ácido nucleico incluye: identificar una localización de la molécula de ácido nucleico en el genoma humano; y determinar un alelo respectivo de la molécula de ácido nucleico. El método comprende además determinar uno o más primeros loci, en donde el genoma fetal es heterocigótico en cada primer loci de tal manera que el genoma fetal tiene un primer y segundo alelo respectivo en ese primer loci, y en donde el genoma materno es homocigótico en cada primer loci de tal manera que el genoma materno tiene dos del segundo alelo respectivo en ese primer loci, siendo el primer alelo distinto del segundo alelo, en donde determinar que un locus específico es uno de los uno o más primeros loci incluye: determinar un valor de corte para un número de recuentos predichos del primer alelo respectivo en el locus específico, prediciendo el valor de corte si el genoma materno es homocigótico y el genoma fetal es heterocigótico, en donde el valor de corte se determina basándose en una distribución estadística de números de recuentos para distintas combinaciones de homocigosidad y heterocigosidad en el locus específico; basándose en el análisis de la pluralidad de moléculas de ácido nucleico, detectar el primer alelo respectivo y el segundo alelo respectivo en el locus específico; determinar un número de recuentos reales del primer alelo respectivo basándose en el análisis de la pluralidad de moléculas de ácido nucleico de la muestra biológica; y determinar que el locus específico es uno de los primeros loci cuando el número de recuentos reales es inferior al valor de corte. El método comprende además para al menos uno de los primeros loci: determinar un primer número

P de recuentos del respectivo primer alelo y un segundo número Q de recuentos del respectivo segundo alelo; y calcular la concentración fraccional basándose en los números primero y segundo.

Otras realizaciones de la invención están dirigidas a sistemas, aparatos, y medios legibles por ordenador asociados con los métodos descritos a continuación. En una realización, el medio legible por ordenador contiene instrucciones para recibir datos y analizarlos, pero no instrucciones para dirigir una máquina para crear los datos (por ejemplo, secuenciar moléculas de ácido nucleico). En otra realización, el medio legible por ordenador contiene instrucciones para dirigir una máquina para crear los datos. En una realización, un producto de programa de ordenador comprende un medio legible por ordenador que almacena una pluralidad de instrucciones para controlar un procesador para realizar una operación para los métodos descritos en el presente documento. Las realizaciones también se dirigen a sistemas informáticos configurados para realizar las etapas de cualquiera de los métodos descritos en el presente documento, potencialmente con distintos componentes que realizan una etapa respectiva o un grupo respectivo de etapas.

La referencia a las partes restantes de la memoria descriptiva, incluyendo los dibujos y las reivindicaciones, mostrará otras características y ventajas de las realizaciones de la presente invención. Otras características y ventajas, así como la estructura y el funcionamiento de las diversas realizaciones de la presente invención, se describen con detalle a continuación con respecto a los dibujos adjuntos. En los dibujos, los números de referencia iguales pueden indicar elementos idénticos o funcionalmente similares.

La FIG. 1 es un diagrama de flujo de un método 100 para determinar al menos una parte del genoma de un feto no nacido de una mujer embarazada según las realizaciones de la presente invención.

La FIG. 2 muestra dos haplotipos para el padre y dos haplotipos para la madre para un segmento particular de su respectivo código genómico según las realizaciones de la presente invención.

La FIG. 3 muestra los dos tipos de SNPs en los haplotipos parentales de la FIG. 2 según las realizaciones de la presente invención.

Las Figs. 4A y 4B muestran un análisis para determinar los haplotipos fetales para los dos tipos de SNP de acuerdo con las realizaciones de la presente invención.

Las Figs. 5A y 5B muestran el análisis de la comparación de cantidades relativas (por ejemplo, recuentos) de fragmentos para cada locus y si un resultado de la comparación es para clasificar un haplotipo particular como heredado o no según las realizaciones de la presente invención.

La FIG. 6 ilustra el efecto de cambiar la relación de probabilidad para la clasificación de SPRT según las realizaciones de la presente invención.

La FIG. 7 es un diagrama de flujo de un método 700 para determinar al menos una porción del genoma de un feto no nacido de una mujer embarazada heredado del padre de acuerdo con las realizaciones de la presente invención.

La FIG. 8 es un diagrama de flujo de un método 800 para determinar al menos una porción del genoma de un feto no nacido en una región en la que la madre y el padre son heterocigóticos según las realizaciones de la presente invención.

La FIG. 9 muestra haplotipos de un padre y una madre que son ambos heterocigotos en una región genómica particular según las realizaciones de la presente invención.

La FIG. 10 es un diagrama de flujo que ilustra un método 1000 para determinar la concentración fraccionada del material fetal en una muestra materna de acuerdo con las realizaciones de la presente invención.

La FIG. 11 es un diagrama de flujo de un método para determinar si un locus es informativo según las realizaciones de la presente invención.

Las FIG. 12A y 12B muestran la distribución prevista de los recuentos para el alelo T (el alelo menos abundante en los escenarios (a) y (c)) para los tres escenarios con una concentración de ADN fetal fraccional supuesta del 20% y 5%, respectivamente, de acuerdo con las realizaciones de la presente invención.

Las Figs. 13A, 13B y 14 muestran las distribuciones previstas para los recuentos del alelo menos abundante para una concentración fraccionada de ADN fetal del 20%, cada uno para diferentes conteos totales de moléculas correspondientes a un SNP según las realizaciones de la presente invención.

Las FIG. 15A y 15B muestran ejemplos de haplotipos de referencia, haplotipos parentales tomados de los haplotipos de referencia, y un haplotipo fetal resultante según las realizaciones de la presente invención.

La FIG. 16 es un diagrama de flujo de un método 1600 para determinar al menos parte de un genoma fetal cuando se conoce un conjunto de haplotipos de referencia, pero los haplotipos parentales no se conocen, según las realizaciones de la presente invención.

La FIG. 17 muestra un ejemplo de determinación de loci informativos a partir del análisis de fragmentos de ADN de una muestra materna según las realizaciones de la presente invención.

La FIG. 18 muestra los tres haplotipos de referencia (Hap A, Hap B y Hap C) y los alelos paternos.

La FIG. 19 muestra la determinación del haplotipo parental a partir de los alelos paternos según las realizaciones de la presente invención.

La FIG. 20 muestra la deducción de los genotipos maternos del análisis de la muestra materna de acuerdo con las realizaciones de la presente invención.

La FIG. 21 muestra una realización para determinar los haplotipos maternos de los genotipos maternos y los haplotipos de referencia de acuerdo con las realizaciones de la presente invención.

La FIG. 22 muestra los haplotipos maternos determinados y el haplotipo heredado paternalmente según las realizaciones de la presente invención.

La FIG. 23 muestra los diferentes tipos de loci (alfa (A) y beta (B)) para los haplotipos maternos en relación con el haplotipo paterno según las realizaciones de la presente invención.

La FIG. 24 es un diagrama de flujo que ilustra un método 2400 de identificación de una mutación de novo en el genoma de un feto no nacido de una mujer embarazada.

La FIG. 25A muestra el número absoluto y los porcentajes de SNPs que muestran diferentes combinaciones de genotipo para el padre, la madre y el feto (CVS) según las realizaciones de la presente invención.

La FIG. 25B muestra una tabla que enumera las estadísticas de alineación de las primeras 20 celdas de flujo.

La FIG. 26 es una tabla que muestra las concentraciones fraccionarias del ADN fetal calculadas para SNPs mediante dos métodos según las realizaciones de la presente invención.

La FIG. 27A muestra un gráfico que ilustra el porcentaje observado de SNPs en este subconjunto en el que se puede ver un alelo fetal a partir de los datos de secuenciación para las primeras 20 células de flujo analizadas, y la FIG. 27B muestra un gráfico de la cobertura frente al número de lecturas según las realizaciones de la presente invención.

Las FIG. 28A y 28B muestran gráficas de la correlación entre la cobertura de alelos paternalmente heredados y el número de lecturas de secuencias asignables y el número de secuencias de células de flujo, respectivamente, según las realizaciones de la presente invención.

La FIG. 29A muestra la correlación entre la tasa de falsos positivos y el número de células de flujo secuenciadas, y la FIG. 29B muestra la correlación entre la tasa de falsos positivos y el número de células de flujo secuenciadas de acuerdo con las realizaciones de la presente invención.

La FIG. 30 muestra la cobertura de los SNPs específicos del feto para diferentes números de células de flujo analizadas de acuerdo con las realizaciones de la presente invención.

La FIG. 31 muestra la precisión del análisis de Tipo A cuando se utilizaban datos de 10 células de flujo de acuerdo con las realizaciones de la presente invención.

La FIG. 32 muestra la precisión del análisis de Tipo B cuando se utilizaban datos de 10 células de flujo de acuerdo con las realizaciones de la presente invención.

La FIG. 33 muestra la precisión del análisis de Tipo A cuando se utilizaban los datos de 20 células de flujo de acuerdo con las realizaciones de la presente invención.

La FIG. 34 muestra la precisión del análisis de Tipo B cuando se utilizaban los datos de 20 células de flujo de acuerdo con las realizaciones de la presente invención.

Las FIG. 35A y 35B muestran lecturas con mutaciones y con una secuencia de tipo salvaje en los codones 41/42 según las realizaciones de la presente invención.

La FIG. 36 muestra una tabla de un análisis RHDO de Tipo A mientras que los del análisis RHDO de Tipo B se muestran en la FIG. 37 según las realizaciones de la presente invención.

Las FIGS. 38A y 38B muestran los resultados de la clasificación SPRT para el caso PW226 como ejemplo.

La FIG. 39 muestra una tabla que resume los resultados del análisis de RHDO para los cinco casos según las realizaciones de la presente invención.

La FIG. 40 muestra un gráfico de la profundidad de secuenciación frente al número de células de flujo secuenciadas de acuerdo con las realizaciones de la presente invención.

La FIG. 41 muestra un gráfico de los tamaños de las secuencias fetales y totales para el genoma completo, y la FIG. 42A-42C muestra gráficas similares individualmente para cada cromosoma según las realizaciones de la presente invención.

La FIG. 43 muestra un diagrama de bloques de un sistema informático del ejemplo 4300 utilizable con el sistema y los métodos según las realizaciones de la presente invención.

La expresión "*muestra biológica*", tal como se utiliza en la presente memoria, se refiere a cualquier muestra que se toma de un sujeto (por ejemplo, un ser humano, tal como una mujer embarazada) y contiene una o más moléculas de ácido nucleico de interés.

La expresión "*ácido nucleico*" o "*polinucleótido*" se refiere a un ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN) y un polímero del mismo en forma de cadena única o bicatenaria. A menos que esté específicamente limitada, la expresión abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares como el ácido nucleico de referencia y se metabolizan de una manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también implica implícitamente variantes conservativamente modificadas de la misma (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, SNPs y secuencias complementarias, así como la secuencia explícitamente indicada. Específicamente, las sustituciones de codones degenerados se pueden lograr generando secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) está sustituida con residuos de base mixta y/o de desoxiinosina (Batzet et al., Nucleic Acid Res. 19: 5081 (1991); Ohtsuka et al., J. Biol. Chem. 260: 2605-2608 (1985); y Rossolini et al., Mol. Cell. Probes 8: 91-98 (1994)). La expresión "ácido nucleico" se utiliza indistintamente con el gen, el ADNc, el ARNm, el ARN pequeño no codificante, el micro ARN (miARN), el ARN que interactúa con el Piwi y el ARN de horquilla corta (shRNA) codificado por un gen o locus.

El término "*gen*" significa el segmento de ADN implicado en la producción de una cadena polipeptídica o un producto de ARN transcrito. Puede incluir regiones que preceden y siguen a la región de codificación (líder y tráiler), así como secuencias intermedias (intrones) entre segmentos de codificación individuales (exones).

La expresión "*secuencia de ácido nucleico clínicamente relevante*" (también denominada secuencia o cromosoma diana) tal como se utiliza en la presente memoria descriptiva, puede referirse a una secuencia polinucleotídica que corresponde a un segmento de una secuencia genómica más grande cuyo potencial de desequilibrio está siendo probado o a la propia secuencia genómica más grande. Un ejemplo es la secuencia del cromosoma 21. Otros ejemplos incluyen los cromosomas 18, 13, X e Y. Todavía otros ejemplos incluyen las secuencias genéticas mutadas o polimorfismos genéticos o variaciones del número de copias que un feto puede heredar de uno o ambos de sus padres, o como una mutación de novo en el feto. En algunas realizaciones, se pueden usar múltiples secuencias de ácido nucleico clínicamente relevantes, o creadores equivalentemente múltiples de la secuencia de ácido nucleico clínicamente relevante, para proporcionar datos para detectar el desequilibrio. Por ejemplo, los datos de cinco secuencias no consecutivas en el cromosoma 21 pueden usarse de una manera aditiva para la determinación de un posible desequilibrio cromosómico 21, reduciendo efectivamente el volumen de muestra necesario a 1/5.

La expresión "*basado en*" como se usa en este documento significa "*basado al menos en parte en*" y se refiere a un valor (o resultado) que se utiliza en la determinación de otro valor, tal como ocurre en la relación de una entrada de un método y la salida de ese método. El término "*derivar*", tal como se usa en la presente memoria, también se refiere a la relación de una entrada de un método y la salida de ese método, tal como ocurre cuando la derivación es el cálculo de una fórmula.

El término "*parámetro*", tal como se utiliza en la presente memoria, significa un valor numérico que caracteriza un conjunto de datos cuantitativos y/o una relación numérica entre conjuntos de datos cuantitativos. Por ejemplo, una relación (o función de una relación) entre una primera cantidad de una primera secuencia de ácido nucleico y una segunda cantidad de una segunda secuencia de ácido nucleico es un parámetro.

Como se usa en este documento, el término "*locus*", o su forma plural "*loci*", es una localización o dirección de cualquier longitud de nucleótidos (o pares de bases) que tiene una variación a través de los genomas.

La expresión "*desequilibrio secuencial*", tal como se utiliza en la presente memoria, significa cualquier desviación significativa definida por al menos un valor de corte en una cantidad de la secuencia de ácido nucleico clínicamente relevante a partir de una cantidad de referencia. Un desequilibrio de secuencia puede incluir el desequilibrio de la dosis cromosómica, desequilibrio alélico, desequilibrio de la dosis de mutación, desequilibrio de la dosis del haplotipo y otros desequilibrios similares. Como ejemplo, un desequilibrio de dosificación de alelos o mutaciones puede ocurrir cuando un feto tiene un genotipo diferente de la madre, creando de este modo un desequilibrio en un locus particular de la muestra.

La expresión "*aneuploidía cromosómica*", tal como se utiliza en la presente memoria, significa una variación en la cantidad cuantitativa de un cromosoma de la de un genoma diploide. La variación puede ser una ganancia o una pérdida. Puede implicar la totalidad de un cromosoma o una región de un cromosoma.

El término "*haplotipo*", como se usa en la presente memoria, se refiere a una combinación de alelos en múltiples loci que se transmiten juntos en el mismo cromosoma o región cromosómica. Un haplotipo puede referirse a tan sólo un par de loci o a una región cromosómica, o a un cromosoma entero. El término "*alelos*" se refiere a secuencias alternativas de ADN en el mismo locus genómico físico, que pueden o no resultar en diferentes rasgos fenotípicos. En cualquier organismo diploide particular, con dos copias de cada cromosoma (excepto los cromosomas sexuales en un sujeto humano masculino), el genotipo de cada gen comprende el par de alelos presentes en ese locus, que son los mismos en homocigotos y diferentes en heterocigotos. Una población o especie de organismos típicamente incluye múltiples alelos en cada locus entre varios individuos. Un locus genómico donde más de un alelo se encuentra en la población se denomina un sitio polimórfico. La variación alélica en un locus es mensurable como el número de alelos (es decir, el grado de polimorfismo) presente, o la proporción de heterocigotos (es decir, la tasa de heterocigosidad) en la población. Como se usa en este documento, el término "*polimorfismo*" se refiere a cualquier variación interindividual en el genoma humano, independientemente de su frecuencia. Ejemplos de tales variaciones incluyen, pero no se limitan a, polimorfismos de nucleótidos sencillos, polimorfismos simples de repetición en tándem, polimorfismos de inserción-delección, mutaciones (que pueden ser causantes de enfermedades) y variaciones del número de copias.

Se puede proporcionar una construcción de un mapa genético parcial o una secuencia genómica completa de un feto no nacido sobre la base de los haplotipos de las secuencias polimórficas de sus padres. El término "*haplotipo*", como se usa en la presente memoria, se refiere a una combinación de alelos en múltiples loci que se transmiten juntos en el mismo cromosoma o la región cromosómica. Por ejemplo, las realizaciones pueden analizar fragmentos de ADN a partir de una muestra materna (que contiene ADN materno y fetal) para identificar alelos en ciertos loci específicos (puntos de referencia). Las cantidades de fragmentos de ADN de los respectivos alelos en estos loci pueden entonces analizarse conjuntamente para determinar las cantidades relativas de los haplotipos para estos loci y determinar con ello qué haplotipos han sido heredados por el feto de los genomas maternos y/o paternos. Al identificar los haplotipos fetales, se puede determinar el genotipo fetal en un locus individual dentro de la región genómica correspondiente incluyendo los loci especificados. En diversas realizaciones, los loci donde los padres son una combinación específica de homocigotos y heterocigotos se pueden analizar de una manera para determinar regiones del genoma fetal. En una implementación, los haplotipos de referencia que son representativos de haplotipos comunes en la población se utilizan junto con el análisis de los fragmentos de ADN de la muestra materna para determinar el genoma materno y paterno.

Un ejemplo de una aplicación de una realización para determinar al menos parte de un genoma fetal podría ser para pruebas de paternidad comparando el genotipo o haplotipo fetal deducido con el genotipo o haplotipo del supuesto padre. Otro ejemplo es detectar una o más mutaciones de novo que el feto ha adquirido, o detectar eventos de recombinación meiótica que se han producido durante la producción de gametos de sus padres. Estos son los gametos que han fertilizado, y el cigoto resultante se ha desarrollado en el feto.

Además, algunas realizaciones también pueden permitir que la secuencia genómica del feto no nacido se determine en cualquier resolución deseada. Por ejemplo, en ciertas aplicaciones, las realizaciones pueden permitir determinar la secuencia genómica completa o cercana a la completa del feto. En una realización, la resolución de la secuencia genómica fetal que puede determinarse depende de la resolución del conocimiento de los genomas del padre y la madre, junto con la información de secuenciación de la muestra biológica materna que contiene los ácidos nucleicos fetales. En el caso de que se conozcan las secuencias genómicas completas o cercanas a completas del padre y la madre, se podría deducir la secuencia genómica completa o cercana a la completa del feto no nacido.

En otras realizaciones, sólo se deducen las secuencias genómicas de regiones seleccionadas dentro del genoma, por ejemplo, para el diagnóstico prenatal de trastornos genéticos, epigenéticos seleccionados (tales como trastornos de impresión) o trastornos cromosómicos. Ejemplos de trastornos genéticos a los que puede aplicarse una realización incluyen las hemoglobinopatías (tal como beta-talasemia, alfa-talasemia, anemia de células falciformes, enfermedad de la hemoglobina E), fibrosis quística y trastornos relacionados con el sexo (tales como hemofilia y

distrofia muscular de Duchenne). Otros ejemplos de mutaciones que pueden detectarse utilizando una realización se pueden encontrar en la Herencia Mendeliana en Línea en el Hombre (www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi).

También se pueden usar algunas realizaciones para determinar una concentración fraccionada de ADN fetal, que puede realizarse sin ningún conocimiento previo de los genomas específicos de los padres. Un análisis similar también puede usarse para determinar la profundidad de cobertura necesaria para una determinación precisa del genoma fetal. Por lo tanto, esta determinación de cobertura se puede utilizar para estimar la cantidad de datos que se deben analizar para obtener resultados precisos.

I. INTRODUCCIÓN

Cuando se usa una muestra materna (por ejemplo, plasma o suero) como el material para elucidar el haplotipo fetal, puede haber dos desafíos principales. Un primer desafío es que el plasma o suero materno consiste en una mezcla de ADN fetal y materno, siendo el ADN fetal la población menor. Se ha determinado que el ADN fetal representa una concentración media-mediana de entre el 5% y el 10% del total de ADN en el plasma materno en los dos primeros trimestres del embarazo (Lo YMD y otros *Am J Hum Genet* 1998; 62: 768-775; Lun FMF et al *Clin Chem* 2008; 54: 1664-1672). A medida que el ADN es liberado por las células sanguíneas maternas durante el proceso de coagulación sanguínea, la concentración fraccionada del ADN fetal en el suero materno puede ser incluso menor que en el plasma materno. Por lo tanto, en algunas realizaciones, el plasma materno se prefiere sobre el suero materno.

Un segundo desafío es que el ADN fetal y el ADN materno en el plasma materno consisten en fragmentos cortos (Chan KCA et al *Clin Chem* 2004; 50: 88-92). De hecho, el ADN derivado del feto es generalmente más corto que el ADN derivado de la madre en el plasma materno. La mayor parte del ADN fetal en el plasma materno tiene menos de 200 pb de longitud. Usando tales fragmentos cortos de ADN de plasma, puede resultar difícil construir el haplotipo de polimorfismos genéticos a lo largo de las largas distancias genómicas. Los desafíos mencionados anteriormente para el plasma materno y el suero también se aplican a la detección de ADN fetal en la orina materna (Botezatu I et al *Clin Chem* 2000; 46: 1078-1084). El ADN fetal sólo representa una fracción menor del ADN en la orina de una mujer embarazada, y el ADN fetal en la orina materna también consiste en fragmentos cortos de ADN.

A. Secuenciación y análisis de la muestra materna

Un enfoque que algunas realizaciones han tomado para abordar el primer reto es usar un método que permita el genotipado cuantitativo de ácidos nucleicos obtenidos a partir de la muestra biológica materna con alta precisión. En una realización de este enfoque, la precisión se logra mediante el análisis de un gran número (por ejemplo, millones o billones) de moléculas de ácido nucleico. Además, la precisión se puede mejorar mediante el análisis de moléculas de ácido nucleico individuales o la amplificación clonal de moléculas de ácido nucleico individuales. Una realización utiliza una secuenciación masivamente paralela de ADN, tal como, pero sin limitarse a la realizada por la plataforma Illumina Genome Analyzer (Bentley DR et al. *Nature* 2008; 456: 53-59), la plataforma Roche 454 (Margulies M et al. *Nature* 2005; 437: 376-380), la plataforma ABI SOLiD (McKernan KJ et al. *Genome Res* 2009; 19: 1527-1541), la plataforma de secuenciación de una sola molécula de Helicos (Harris TD et al. *Science* 2008; 320: 106-109), la secuenciación en tiempo real utilizando moléculas de polimerasa única (*Science* 2009; 323: 133-138) y la secuenciación de nanoporos (Clarke J. et al. *Nat Nanotechnol.* 2009; 4: 265-70). En una realización, la secuenciación masivamente paralela se realiza sobre un subconjunto aleatorio de moléculas de ácido nucleico en la muestra biológica.

En algunas realizaciones, puede ser beneficioso obtener una secuencia leída de cada molécula tan larga como sea posible. Una limitación de la longitud de las lecturas de secuenciación que se puede conseguir es la naturaleza de las moléculas de ácido nucleico en la muestra biológica materna. Por ejemplo, se sabe que la mayoría de las moléculas de ADN en el plasma materno consisten en fragmentos cortos (Chan KCA et al *Clin Chem* 2004; 50: 88-92). Además, la longitud de lectura debe equilibrarse con la fidelidad del sistema de secuenciación a largas longitudes de lectura. Para algunos de los sistemas mencionados anteriormente, podría ser preferible obtener secuencias de ambos extremos de la molécula, la llamada secuenciación de extremo pareado. Como ilustración, un enfoque consiste en realizar 50 pb de secuenciación desde cada extremo de una molécula de ADN, dando como resultado un total de 100 pb de secuencia por molécula. En otra realización, puede realizarse 75 pb de secuenciación desde cada extremo de una molécula de ADN, dando como resultado un total de 150 pb de secuencia por molécula.

Después de realizar la secuenciación, las secuencias se alinean de nuevo con un genoma humano de referencia. Como las realizaciones dilucidar las variaciones genómicas heredadas por un feto no nacido de sus padres, el algoritmo de alineación puede ser capaz de hacer frente a las variaciones de secuencia. Un ejemplo de dicho paquete de software es el software Efficient Large-Scale Alignment of Nucleotide Databases (ELAND) producido por Illumina. Otro ejemplo de dicho paquete de software es el SOAP (programa de alineación de oligonucleótidos cortos) y SOAP2 (Li R et al. *Bioinformatics* 2008; 24: 713-714; Li R et al. *Bioinformatics* 2009; 25: 1966-1967).

La cantidad de secuenciación de ADN que puede necesitar ser realizada puede depender de la resolución a la que el mapa genético fetal o la secuencia genómica fetal puede necesitar ser construida. En general, cuantas más moléculas se secuencien, mayor será la resolución. Otro determinante de la resolución del mapa genético fetal o de la secuencia genómica fetal en un nivel o profundidad de secuenciación del ADN es la concentración fraccionada de ADN fetal en la muestra biológica materna. En general, cuanto mayor sea la concentración fraccionada de ADN fetal, mayor será la resolución del mapa genético fetal o secuencia genómica fetal que puede ser elucidada a un nivel dado de secuenciación de ADN. Como la concentración fraccional de ADN fetal en el plasma materno es mayor que en el suero materno, el plasma materno es un tipo de muestra biológica materna más preferido que el suero materno para algunas realizaciones.

El rendimiento de los métodos basados en secuenciación mencionados anteriormente puede aumentarse con el uso de la indexación o la codificación de códigos de barras. Así, una muestra o un índice o código de barras específico para el paciente se puede añadir a fragmentos de ácido nucleico en una biblioteca de secuenciación del ácido nucleico particular. A continuación, se mezclan una serie de tales bibliotecas, cada una con una muestra o índice o código de barras específico del paciente, y se secuencian entre sí. Después de las reacciones de secuenciación, los datos de secuenciación pueden recogerse de cada muestra o paciente basándose en el código de barras o índice. Esta estrategia puede aumentar el rendimiento y, por tanto, la rentabilidad de las realizaciones de la presente invención.

En una realización, las moléculas de ácido nucleico en la muestra biológica pueden seleccionarse o fraccionarse antes de la genotipificación cuantitativa (por ejemplo, secuenciación). En una variante, las moléculas de ácido nucleico se tratan con un dispositivo (por ejemplo, un microarray) que puede unirse preferentemente a moléculas de ácido nucleico a partir de loci seleccionados en el genoma (por ejemplo, la región del cromosoma 7 que contiene el gen CFTR). Entonces, la secuenciación puede realizarse preferentemente sobre moléculas de ácido nucleico capturadas por el dispositivo. Este esquema permitirá dirigir la secuencia hacia la región genómica de interés. En una realización de este esquema se puede usar un sistema de captura de secuencias de Nimblegen (www.nimblegen.com/products/seqcap/index.html) o un Sistema Agilent SureSelect Target Enrichment (www.opengenomics.com/SureSelect_Target_Enrichment_System), o plataformas similares. En algunas realizaciones, las moléculas de ácido nucleico de las regiones seleccionadas del genoma se someten a secuenciación aleatoria.

En otra realización, la región genómica de interés en la muestra biológica puede amplificarse primero mediante un conjunto o conjunto múltiple de cebadores de amplificación. A continuación, el genotipado cuantitativo, por ejemplo, la secuenciación, puede realizarse sobre los productos amplificados. En una implementación de este esquema, se puede usar el sistema RainDance (www.raindancetech.com/technology/pcr-genomics-research.asp). En algunas realizaciones, las moléculas de ácido nucleico amplificadas se someten a secuenciación aleatoria.

También se puede realizar una etapa de fraccionamiento por tamaños sobre las moléculas de ácido nucleico en la muestra biológica. Dado que se sabe que el ADN fetal es más corto que el ADN materno en el plasma materno (Li y col Clin Chem 2004; 50: 1002-1011; Solicitud de Patente de EE.UU. 20050164241; Solicitud de Patente de EE.UU. 20070202525), la fracción de menor tamaño molecular se puede cosechar y luego ser utilizada para la genotipificación cuantitativa, por ejemplo, secuenciando. Tal fracción contendría una concentración fraccional más alta del ADN fetal que en la muestra biológica original. De este modo, la secuenciación de una fracción enriquecida en ADN fetal puede permitir construir el mapa genético fetal o deducir la secuencia genómica fetal con una mayor resolución a un nivel particular de análisis (por ejemplo, profundidad de secuenciación), que si una muestra no enriquecida ha sido usada. Esto puede, por lo tanto, hacer la tecnología más rentable. Como ejemplos de métodos para el fraccionamiento por tamaños, podría utilizarse (i) electroforesis en gel seguida por la extracción de moléculas de ácido nucleico a partir de fracciones específicas de gel; (ii) matriz de unión a ácidos nucleicos con afinidad diferencial para moléculas de ácido nucleico de diferentes tamaños; o (iii) sistemas de filtración con retención diferencial para moléculas de ácido nucleico de diferentes tamaños.

En otra realización más, se podrían analizar preferencialmente moléculas de ácido nucleico de un tamaño específico o rango de tamaños después de la secuenciación del ácido nucleico. Por ejemplo, se podría realizar una secuenciación en pares en la que se secuencian ambos extremos de una molécula de ADN. A continuación, las coordenadas genómicas de ambos extremos podrían ser asignadas de nuevo a un genoma humano de referencia. Entonces se podría deducir el tamaño de la molécula restando las coordenadas genómicas de ambos extremos. Una forma de realizar esta secuenciación en pares es utilizar el protocolo de secuenciación en pares del Illumina Genome Analyzer. Otro método para deducir el tamaño de una molécula de ADN es secuenciar toda la molécula de ADN. Esto se hace más fácilmente mediante plataformas de secuenciación con longitudes de lectura relativamente largas, como la plataforma Roche 454 (Marguelis et al Nature 2005; 437: 376-380) y la molécula única de Pacific Biosciences, en tiempo real (SMRT™) (Eid et al Science 2009; 323: 133-138). Después de la deducción del tamaño de las moléculas de ácido nucleico, se podría optar por centrar el análisis subsiguiente en moléculas de menos de un corte de tamaño particular, con lo que se enriquecería en la concentración fraccionada de ADN fetal. El análisis de este subconjunto de moléculas puede permitir que el mapa genético fetal o las secuencias genómicas fetales se deduzcan con menos moléculas analizadas después de la selección por tamaños que si este procedimiento no se ha

realizado. En una realización, se usa un corte de tamaño de 300 pb. En aún otras realizaciones, se podría usar un corte de tamaño de 250 pb, 200 pb, 180 pb, 150 pb, 125 pb, 100 pb, o 75 pb.

B. Uso de genomas parentales como andamios

Para abordar el segundo desafío, algunas realizaciones pueden usar haplotipos de los cromosomas de la madre como un "andamio". Los haplotipos de los cromosomas del padre también se pueden utilizar como otro "andamio". Este andamio puede compararse con la información genética del feto obtenida de la muestra materna que contiene el ADN fetal. Esta información genética fetal puede usarse para determinar cómo se ha erigido el andamio de la madre y/o el padre en el genoma fetal, usando de este modo las partes componentes del andamio para determinar el genoma fetal resultante.

Los haplotipos parentales se pueden construir a partir del ADN genómico del padre y la madre, y de otros miembros de la familia, p. ej., un hermano del feto en el embarazo actual. Es posible que la disponibilidad de los haplotipos parentales se vuelva cada vez más común, en vista de la reducción en los costos de la secuenciación genómica. En un escenario, si uno o ambos padres ya tienen su genoma secuenciado y sus haplotipos en una o más regiones cromosómicas se han determinado, entonces esta información se puede utilizar como el mencionado andamio.

Se puede utilizar cualquier plataforma de genotipificación conocida por los expertos en la técnica que pueda interrogar variaciones de secuencias en el genoma, incluyendo secuenciación de ADN, microarrays, sondas de hibridación, técnicas basadas en fluorescencia, técnicas ópticas, códigos de barras moleculares y formación de imágenes de molécula única (Geiss GK et al. Nat Biotechnol 2008; 26: 317-325), Análisis de molécula única, PCR, PCR digital, espectrometría de masas (como la plataforma Sequenom MassARRAY), etc. Como ejemplo más extremo, la secuencia de ADN del padre y de la madre puede determinarse por secuenciación de ADN del genoma completo usando un método de secuenciación masivamente paralelo (p. ej., Bentley DR et al. Nature 2008; 456: 53-59; McKernan KJ et al. Genome Res 2009; 19: 1527-1541). Un ejemplo de variaciones de secuencias que pueden ser de interés son los polimorfismos de un solo nucleótido (SNP). Un método particularmente preferido para determinar los genotipos parentales es mediante análisis de microarrays de SNPs a una escala de todo el genoma, o en regiones genómicas seleccionadas, p. ej., aquellas que contienen genes cuyas mutaciones pueden causar enfermedades genéticas (tales como genes en el grupo de beta-globina, o el gen regulador de la conductancia transmembrana de la fibrosis quística (CFTR)). Además de las variaciones de secuencia, también se pueden usar variaciones de número de copias. Las variaciones de secuencias y las variaciones del número de copias se denominan características genéticas polimórficas (PMF).

En un aspecto, los genotipos maternos en los cromosomas o regiones cromosómicas de interés pueden construirse en haplotipos. Una forma en la que esto se puede realizar es mediante el análisis de otros miembros de la familia relacionados con la madre, p. ej., un hijo o una hija de la madre, un padre, un hermano, etc. Otra forma en la que pueden construirse los haplotipos es mediante otros métodos bien conocidos por los expertos en la técnica mencionados anteriormente.

La información del genotipo puede extenderse entonces a la información del haplotipo de los padres en comparación con la información del genotipo de otros miembros de la familia, por ejemplo, un hermano del feto del embarazo actual, o de los genotipos de los abuelos, etc. Los haplotipos de los padres también pueden construirse por otros métodos bien conocidos por los expertos en la técnica. Ejemplos de tales métodos incluyen métodos basados en análisis de molécula única tales como PCR digital (Ding C y Cantor CR. Proc Natl Acad Sci USA 2003; 100: 7449-7453; Ruano G et al. Proc Natl Acad Sci USA 1990; 87: 6296-6300), haplotipificación de esperma (Lien S et al. Curr Protoc Hum Genet 2002; Capítulo 1: Unidad 1.6) y técnicas de imagen (Xiao M et al. Hum Mutat 2007; 28: 913-921). Otros métodos incluyen los basados en la PCR de alelo específico (Michalatos-Beloin S et al. Nucleic Acids Res 1996; 24: 4841-4843; Lo YMD et al. Nucleic Acids Res 1991; Nucleic Acids Res. 19: 3561-3567), la clonación y la digestión con enzimas de restricción (Smirnova AS et al. Immunogenetics 2007; 59: 93-8), etc. Sin embargo, otros métodos se basan en la estructura de desequilibrio de distribución y enlace de bloques de haplotipos en la población que permiten inferir el haplotipo materno a partir de evaluaciones estadísticas (Clark AG. Mol Biol Evol 1990; 7: 111-22; 10:13-9; Salem RM et al. Hum Genomics 2005; 2: 39-66).

C. Uso de la información genómica de la muestra materna para ensamblar el andamio

En una realización, para determinar cuáles de los cromosomas maternos se han pasado al feto, se usa un método de dosis de haplotipo relativo (RHDO). Un principio general de este enfoque es el siguiente para un ejemplo de donde la madre es heterocigótica para cada uno de los polimorfismos genéticos. Por lo tanto, hay dos haplotipos, y la dosis relativa de estos haplotipos sería 1:1. Sin embargo, en la muestra materna, la presencia de una pequeña proporción de ADN fetal podría alterar la dosis relativa de haplotipo. Esto se debe a que el feto habría heredado la mitad de su complemento de haplotipo de la madre y la otra mitad del padre. Además, para cada cromosoma, el feto podría haber heredado un "mosaico" de haplotipos que se han originado de uno u otro de los cromosomas homólogos de cada uno de los padres, dependiendo de la ocurrencia de la recombinación meiótica. Todos estos factores pueden desviar la dosis relativa de haplotipo de la relación 1: 1 en el ADN constitucional materno. Por tanto,

para un cromosoma o región cromosómica dada, los alelos constituyentes de estos haplotipos se pueden buscar a partir de datos analíticos (por ejemplo, datos de secuenciación) generados a partir de la muestra materna.

A continuación, se puede realizar un procedimiento estadístico para determinar la dosis relativa de haplotipo, o si uno de estos haplotipos está sobre-representado respecto al otro haplotipo. El umbral de clasificación para este procedimiento estadístico puede ajustarse dependiendo de la concentración fraccional de ADN fetal. En general, una mayor concentración fraccionada de ADN fetal puede permitir que el umbral se alcance con menos moléculas. El umbral de clasificación también puede ajustarse dependiendo del número de fragmentos clasificados con éxito que se desee alcanzar a través del genoma o de las regiones genómicas de interés. En una realización, puede usarse la prueba de relación de probabilidad secuencial (SPRT).

En una realización, una dosificación de mutación relativa (RMD), como se describe en la Solicitud de Patente de EE.UU. 2009/0087847) puede usarse para determinar una cantidad relativa de un alelo en polimorfismos particulares de la madre. Estas cantidades relativas pueden usarse para determinar un haplotipo del feto (por ejemplo, cuando los polimorfismos están en loci consecutivos o enlazados). En una implementación de este enfoque dirigido está el uso de la reacción en cadena de la polimerasa (PCR) para amplificar las secuencias específicas de las partes seleccionadas del genoma para el análisis RMD. Para ampliar este enfoque RMD para determinar la herencia fetal en una gran región genómica o en todo el genoma, es necesario un gran volumen de la muestra materna.

En una realización que usa la secuenciación aleatoria, las regiones genómicas de interés no están dirigidas específicamente. Por lo tanto, el número de secuencias obtenidas en las regiones genómicas de interés puede no ser tan numeroso como en un enfoque dirigido (a menos que se realice una secuenciación muy profunda). Sin embargo, los recuentos se pueden agrupar a través de una serie de polimorfismos vinculados, para lograr la potencia estadística necesaria con fines de diagnóstico. Una implicación práctica de usar esta realización secuencial es que puede ahorrar costes evitando la necesidad de una secuenciación excesivamente profunda. También requiere una entrada de una menor cantidad de muestra materna que los enfoques basados en PCR digital.

Además, puede ser deseable llevar a cabo dicho análisis RHDO en bloques. En otras palabras, cada cromosoma puede ser analizado en uno, o preferiblemente más de un bloque. En un aspecto, este último puede permitir que se observe la recombinación meiótica. Por ejemplo, un haplotipo de un segmento de un cromosoma particular del feto puede parecer haber venido de uno de los cromosomas homólogos maternos, mientras que el otro segmento del mismo cromosoma fetal parece poseer el haplotipo del otro cromosoma homólogo materno. Un análisis SPRT puede permitir que esta segmentación se realice.

Por ejemplo, el análisis SPRT puede realizarse en SNPs vecinos que demuestren la configuración requerida del genotipo parental (es decir, siendo el padre homocigótico y la madre heterocigótica) partiendo de un extremo de un cromosoma. Esto continuará hasta que el análisis SPRT haya indicado que uno de los haplotipos maternos es predominante en los datos analíticos de plasma materno (por ejemplo, datos de secuenciación). A continuación, el análisis SPRT puede ser "restablecido" y empezar de nuevo desde el siguiente SNP vecino que demuestra la configuración requerida del genotipo parental. Esto puede continuar de nuevo hasta que el análisis SPRT haya indicado una vez más que uno de los haplotipos maternos es predominante en los datos analíticos de plasma materno (por ejemplo, datos de secuenciación). Este proceso puede continuar hasta el último SNP seleccionado en dicho cromosoma. Entonces, estos diversos segmentos de haplotipos determinados por SPRT en el cromosoma pueden compararse con los haplotipos de los dos cromosomas homólogos en el genoma de la madre. Se observa una recombinación meiótica cuando los segmentos de haplotipos en el feto parecen haber cambiado de un cromosoma homólogo materno a otro. Este sistema también puede funcionar incluso si hay más de una recombinación meiótica por cromosoma.

Como se describe más adelante, el análisis RHDO también puede llevarse a cabo para regiones genómicas en las que el padre y la madre son ambos heterocigotos para los polimorfismos genéticos constituyentes. Este escenario es particularmente útil para la situación cuando el padre y la madre comparten una copia mutante del gen de la enfermedad del mismo origen ancestral, tal como cuando son consanguíneos, o cuando la mutación predominante para la enfermedad se debe a un gran efecto fundador (es decir, la mayoría de los individuos con la mutación ha heredado el mismo haplotipo de un fundador ancestral común de la población). Por lo tanto, los haplotipos del padre y la madre en esta región se pueden utilizar para deducir el haplotipo fetal.

II. CONSTRUCCIÓN DEL GENOMA FETAL A PARTIR DEL GENOMA MATERNO

Se describe ahora la construcción de un mapa genético fetal o la elucidación de la secuencia genómica fetal con conocimiento explícito de los genomas parentales.

A. Método

La FIG. 1 es un diagrama de flujo de un método 100 para determinar al menos una porción del genoma de un feto no nacido de una mujer embarazada. El feto tiene un padre y una madre que es la mujer embarazada. El padre tiene

un genoma paterno con dos haplotipos y la madre tiene un genoma maternal con dos haplotipos. El método 100 analiza las moléculas de ácido nucleico (fragmentos) de una muestra biológica obtenida de la mujer embarazada para determinar el genoma del feto. El método 100 se describe principalmente para el ejemplo de cuando el padre es homocigótico y la madre es heterocigótica en una pluralidad de loci, mientras que otros ejemplos describen otras realizaciones.

El método 100 y cualquiera de los métodos descritos en el presente documento pueden realizarse total o parcialmente con un sistema informático que incluye un procesador, que puede configurarse para llevar a cabo las etapas. Así, las realizaciones se dirigen a sistemas informáticos configurados para llevar a cabo las etapas de cualquiera de los procedimientos descritos en la presente memoria, potencialmente con diferentes componentes que realizan una etapa respectiva o un respectivo grupo de etapas. Aunque se presentan como pasos numerados, los pasos de los métodos de la presente invención pueden realizarse al mismo tiempo o en un orden diferente. Adicionalmente, partes de estos pasos pueden usarse con porciones de otras etapas de otros métodos. Además, todas o partes de una etapa pueden ser opcionales. Además, cualquiera de los pasos de cualquiera de los métodos puede realizarse con módulos, circuitos u otros medios para llevar a cabo estos pasos.

En la etapa 110, se identifica una primera pluralidad de loci en la que el genoma materno es heterocigótico. En una realización, esta determinación puede realizarse en parte de un genotipado del padre y la madre a nivel del genoma o en loci genómicos seleccionados de interés. En otras realizaciones, la determinación de la primera pluralidad de loci puede hacerse durante un análisis de la muestra materna, que se describe en secciones posteriores.

En la etapa 120, se determinan cada uno de los dos haplotipos maternos que cubren la primera pluralidad de loci. Como se mencionó anteriormente, el genoma materno podría obtenerse a partir de la secuenciación directa. En otras realizaciones, el genotipado puede hacerse en una pluralidad de loci y luego utilizar un genoma mapeado de alguien que se espera tenga un genoma similar, p. ej., de un miembro de la familia o de un genoma de referencia que es común en una población igual o similar. En una realización, la etapa 120 puede realizarse primero para todas o partes del genoma materno y entonces el genoma materno puede ser investigado para encontrar loci en los que la madre sea heterocigótica.

En un aspecto, no es esencial construir los haplotipos de los cromosomas del padre. Sin embargo, si los haplotipos paternos pudieran construirse, entonces se podría obtener información adicional de los resultados de la secuenciación. Una tal información adicional incluye el hecho de que se puede realizar un análisis de dosificación de haplotipo relativo para regiones en las que ambos progenitores son heterocigotos. Otra información adicional que se puede obtener si el haplotipo paterno está disponible es la información referente a la recombinación meiótica que implica a uno o más cromosomas paternos, y determinar si los alelos de la enfermedad ligados a tales polimorfismos se han pasado al feto.

En la etapa 130, se determina un alelo heredado por el feto del padre en cada uno de la primera pluralidad del loci. Algunas realizaciones usan loci genómicos que son homocigóticos para el padre, pero heterocigotos para la madre (como se mencionó en la etapa 110). Por lo tanto, si el padre es homocigótico en los loci, entonces se conoce el alelo que es heredado del padre. El genotipado del padre para determinar los loci en los que el padre es homocigótico puede determinarse en cualquiera de las formas descritas en el presente documento. En una realización, la determinación de la primera pluralidad de loci puede determinarse basándose en el genotipado del padre y la madre para encontrar los loci en los que el padre es homocigótico y en los que la madre es heterocigótica.

En otra realización, una segunda pluralidad de loci del genoma paterno que son heterocigóticos puede usarse para determinar el haplotipo paterno heredado por el feto en la primera pluralidad de loci en los que el padre es homocigótico. Por ejemplo, si el genoma materno es homocigótico en la segunda pluralidad de loci, se pueden identificar alelos que están presentes en el genoma paterno en los respectivos de la segunda pluralidad de loci y ausentes en el genoma materno. El haplotipo paterno heredado puede entonces ser identificado como el haplotipo con los alelos identificados, y ser usado para determinar el alelo heredado del padre en la primera pluralidad de loci. Estos aspectos de la determinación de un haplotipo paterno se discuten con más detalle a continuación.

En la etapa 140, se analizó una pluralidad de moléculas de ácido nucleico de una muestra biológica obtenida de la mujer embarazada. La muestra contiene una mezcla de ácidos nucleicos maternos y fetales. La muestra biológica materna puede ser tomada y luego ser recibida para su análisis. En una realización, se utiliza plasma materno y suero. En otras realizaciones, pueden usarse sangre materna, orina materna, saliva materna, líquido de lavado uterino o células fetales obtenidas a partir de la sangre materna.

En una realización, el análisis de una molécula de ácido nucleico incluye identificar una localización de la molécula de ácido nucleico en el genoma humano y determinar un alelo de la molécula de ácido nucleico en el locus individual. De este modo, una realización puede realizar el genotipado cuantitativo usando los alelos determinados de las moléculas de ácido nucleico del mismo locus. Se puede utilizar cualquier método que permita determinar la localización genómica y el alelo (información sobre el genotipo) de las moléculas de ácido nucleico en la muestra biológica materna. Algunos de estos métodos se describen en las solicitudes de EE.UU. 12/178.181 y 12/614350, y la solicitud titulada "Size-Based Genomic Analysis."

En la etapa 150, en base a los alelos determinados de las moléculas de ácido nucleico, se determinan las cantidades de alelos respectivos en cada uno de la primera pluralidad de loci. En una realización, las cantidades pueden ser el número de alelos de cada tipo en un primer lugar. Por ejemplo, seis A y cuatro T. En otra realización, una cantidad puede ser una distribución de tamaños de las moléculas de ácido nucleico que tienen un alelo particular. Por ejemplo, una cantidad relativa también puede incluir una distribución de tamaños de los fragmentos con un genotipo particular, que puede transportar una cantidad relativa de fragmentos a ciertas longitudes. Tales cantidades relativas también pueden proporcionar información sobre qué genotipo está en el genoma fetal, ya que los fragmentos fetales tienden a ser más pequeños que los fragmentos maternos. Algunos ejemplos de cantidades y métodos se describen en las solicitudes de EE.UU. 12/178.181 y 12/614350, y la solicitud titulada "Size-Based Genomic Analysis."

En una realización, las cantidades relativas de los alelos en un locus pueden proporcionar información sobre qué genotipo fue heredado por el feto (por ejemplo, después de que un conjunto de datos haya alcanzado suficiente potencia estadística). Por ejemplo, las cantidades relativas pueden usarse para determinar si se produce un desequilibrio de secuencia con respecto a los genotipos de la madre en un locus. Las solicitudes de patente relacionadas citadas anteriormente proporcionan ejemplos de realizaciones para detectar un desequilibrio de secuencia en un locus o región particular.

En la etapa 160, se comparan cantidades relativas de los alelos respectivos de las moléculas de ácido nucleico en más de un locus de la primera pluralidad de loci. En algunas realizaciones, las cantidades de cada alelo en cada locus de la primera pluralidad de loci que comprenden los haplotipos se agregan antes de realizar una comparación. Las cantidades agregadas de los haplotipos parentales pueden entonces compararse para determinar si un haplotipo está sobrerrepresentado, igualmente representado o sub-representado. En otras realizaciones, las cantidades para los alelos en un locus se comparan, y se utilizan comparaciones en múltiples loci. Por ejemplo, se puede agregar un valor de separación (por ejemplo, una diferencia o una relación), que se puede utilizar en una comparación con un valor de corte. Cada una de estas realizaciones puede aplicarse a cualquiera de las etapas de comparación descritas en la presente memoria.

En diversas realizaciones, las cantidades relativas pueden ser un recuento de un número de cada fragmento con un alelo particular en un locus particular, un recuento de un número de fragmentos de cualquier locus (o cualquier loci en una región) en un haplotipo particular, y un valor estadístico del recuento (por ejemplo, un promedio) en un locus particular o en un haplotipo particular. Por tanto, en una realización, la comparación puede ser una determinación de un valor de separación (por ejemplo, una diferencia o una relación) de un alelo frente a otro alelo en cada loci.

En la etapa 170, basándose en la comparación, se puede determinar el haplotipo heredado por el feto no nacido de la madre en la porción del genoma cubierta por la primera pluralidad de loci. En una realización, para determinar cuál de los cromosomas maternos ha sido pasado sobre el feto, se usa un método de dosificación relativa de haplotipo (RHDO), por ejemplo, como se mencionó anteriormente. Como la madre es heterocigótica para cada uno de los primeros loci, los primeros loci corresponden a dos haplotipos para la región genómica de los primeros loci. La dosis relativa de estos haplotipos sería de 1: 1 si la muestra fuera sólo de la madre. Las desviaciones o la ausencia de desviaciones de esta relación pueden usarse para determinar el haplotipo del feto heredado de la madre (y del padre, que se aborda con más detalle más adelante). Por lo tanto, para un cromosoma o región cromosómica dada, los alelos constituyentes de estos haplotipos se pueden buscar a partir de los datos analíticos (por ejemplo, datos de secuenciación) generados en la etapa 130.

Puesto que se analizan una pluralidad de loci y se comparan con el haplotipo de la madre, las secuencias entre los loci pueden atribuirse a un haplotipo particular. En una realización, si varios loci coinciden con un haplotipo particular, entonces se puede suponer que los segmentos de secuencia entre los loci son iguales que los del haplotipo materno. Debido a la ocurrencia de recombinación meiótica, el haplotipo final heredado por el feto puede consistir en un mosaico de 'segmentos de haplotipos' originarios de uno de estos dos cromosomas homólogos. Las realizaciones pueden detectar tal recombinación.

La resolución en la que se podría detectar tal recombinación depende del número y distribución de los marcadores genéticos que se haya determinado en el ADN constitucional del padre y la madre, y el umbral que se utiliza en el subsiguiente análisis bioinformático (utilizando por ejemplo el SPRT). Por ejemplo, si la comparación sugiere que el alelo heredado de la madre en cada uno de un primer conjunto de loci consecutivos corresponden al primer haplotipo, entonces se determina que el primer haplotipo se hereda para la localización genómica correspondiente al primer conjunto de loci. Si un segundo conjunto de loci consecutivos sugieren que el segundo haplotipo se hereda, entonces se determina que el segundo haplotipo se hereda para la localización genómica correspondiente al segundo conjunto de loci.

En una realización, cuando se analizan una pluralidad de loci, el haplotipo puede determinarse con mayor precisión. Por ejemplo, los datos estadísticos de un loci pueden no ser determinantes, pero cuando se combinan con los datos estadísticos de otros loci, se puede determinar qué haplotipo se hereda. En otra realización, cada loci puede

analizarse independientemente para hacer una clasificación y, a continuación, las clasificaciones pueden analizarse para proporcionar una determinación de qué haplotipo se hereda para una región dada.

En una realización, puede realizarse un procedimiento estadístico para determinar la dosificación de haplotipo relativa (por ejemplo, si uno de estos haplotipos está sobre-representado respecto al otro haplotipo). El umbral de clasificación para este procedimiento estadístico puede ajustarse dependiendo de la concentración fraccional de ADN fetal. En general, una mayor concentración fraccionada de ADN fetal puede permitir que el umbral se alcance con menos moléculas. El umbral de clasificación también puede ajustarse dependiendo del número de segmentos clasificados con éxito que se desee alcanzar a través del genoma o de las regiones genómicas de interés.

Haciendo referencia de nuevo a la FIG. 1, en el paso 180, el genoma fetal puede ser analizado para mutaciones. Por ejemplo, se pueden usar realizaciones para buscar un panel de mutaciones que causan enfermedades genéticas en una población particular. Ejemplos de mutaciones que pueden detectarse usando realizaciones se pueden encontrar en la Herencia Mendeliana en Línea en el Hombre (www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi). Estas mutaciones pueden ser buscadas durante las etapas 140-160; o como un paso separado como se describe en este documento. Por ejemplo, en familias en las que el padre es portador de una o más mutaciones que están ausentes en la madre, entonces la mutación o las mutaciones podrían ser buscadas a partir de los datos analíticos (por ejemplo, datos de secuenciación) de la muestra biológica materna.

Aparte de la detección de la mutación real, también podrían buscarse marcadores genéticos polimórficos que están vinculados al alelo mutante o de tipo salvaje en el padre o la madre. Por ejemplo, el análisis de RHDO puede revelar que el feto ha heredado el haplotipo de la madre que se sabe que lleva una mutación para una enfermedad. Las realizaciones de la invención también pueden usarse para el diagnóstico prenatal no invasivo de enfermedades causadas por deleciones de regiones cromosómicas, p. ej., la supresión del Sudeste Asiático que causa la alfa-talasemia. En el escenario en el que tanto el padre como la madre son portadores de la deleción, si el feto es homocigótico para la deleción y si se realiza una secuenciación masivamente paralela en el ADN plasmático materno, entonces debería haber una reducción en las frecuencias de las secuencias de ADN procedentes de la región suprimida en el plasma materno.

B. Ejemplo

Esta sección describe un ejemplo de realizaciones (por ejemplo, del método 100) aplicadas al polimorfismo de un solo nucleótido (SNP) en el que la madre es heterocigótica. Los alelos SNP en el mismo cromosoma forman un haplotipo, con la madre teniendo un par homólogo de cada cromosoma, y por lo tanto dos haplotipos. Para ilustrar cómo se realiza tal determinación, considérese un segmento en el cromosoma 3, por ejemplo, como se muestra en la FIG. 2.

La FIG. 2 muestra dos haplotipos para el padre y dos haplotipos para la madre de un segmento particular de su respectivo código genómico. Cinco SNPs se encontraron dentro de este segmento en el que el padre y la madre fueron homocigotos y heterocigotos, respectivamente, para todos los 5 de estos SNPs. Los dos cromosomas homólogos del padre poseían el mismo haplotipo (Hap), es decir, A-G-A-A-G (de arriba a abajo en la Figura 2). Por simplicidad, los haplotipos paternos se llaman Hap I y Hap II, teniendo en cuenta que ambos son idénticos para este conjunto de 5 SNPs. Para la madre se observaron dos haplotipos: Hap III, A-A-A-G-G y Hap IV, G-G-G-A-A.

Los SNPs en este ejemplo podrían ser clasificados en dos tipos. La FIG. 3 muestra los dos tipos de SNP según las realizaciones de la presente invención. El tipo A consiste en aquellos SNP en los que los alelos paternos eran los mismos que en el haplotipo materno III. El tipo B consiste en aquellos SNPs en los que los alelos paternos eran los mismos que en el haplotipo materno IV.

Estos dos tipos de SNPs pueden requerir un manejo matemático ligeramente diferente. Por lo tanto, en el escenario de Tipo A, la herencia fetal del haplotipo III daría lugar a la sobre-representación del haplotipo III, en relación con el haplotipo IV, en el plasma materno (Figura 4A). Por ejemplo, si se mira sólo un SNP 410 para facilitar la discusión, el alelo A es heredado del padre, y si Hap III es heredado de la madre, entonces el feto aportará dos alelos A a la muestra, lo que causará una sobre-representación de A. Si el feto hubiera heredado el haplotipo IV, no se vería ninguna sobre-representación, ya que el feto también sería heterocigótico con A y G en el locus.

Por otro lado, en el escenario Tipo B, la herencia fetal del haplotipo III resultaría en una representación igual de haplotipo III y haplotipo IV en plasma materno (Figura 4B). Por ejemplo, si se mira respecto a SNP 420, la herencia de G del padre y A como parte de Hap III haría que el feto aportara cantidades iguales de A y G en SNP 420, al igual que la madre. Si el feto hubiese heredado el haplotipo IV, se observaría una sobrerrepresentación, como se desprende de la discusión anterior.

Las Figs. 5A y 5B muestran el análisis de comparación de las cantidades relativas (por ejemplo, recuentos) de fragmentos para cada locus y si un resultado de la comparación es clasificar un haplotipo particular como heredado o no. Cualquier localización genómica en la que haya un SNP que se ajuste a una de estas configuraciones de genotipo del padre y la madre (p. ej., escenarios tipo A o Tipo B) se puede utilizar para este ejemplo. A partir de los

datos de secuenciación plasmática materna, se puede enfocar el número de moléculas secuenciadas correspondientes a un alelo particular del SNP. Puede utilizarse un análisis SPRT (u otro método de comparación) para determinar si había algún desequilibrio alélico entre estos alelos (Lo YMD y otros, Proc Natl Acad Sci USA 2007; 104: 13116-13121).

La FIG. 5A muestra un análisis para los SNP de tipo A. Como se muestra, para cada SNP, una comparación SPRT de las cantidades relativas (por ejemplo, como se define por un valor de separación) con un valor de corte proporciona una clasificación. En una realización, si se alcanzó el umbral de clasificación para SPRT, entonces se concluyó la herencia fetal de un haplotipo materno particular. A continuación, se puede restablecer el cálculo del análisis SPRT. Entonces, un análisis puede moverse sobre un SNP vecino que se ajusta a la configuración de genotipo requerida, ya sea desde la dirección telomérica a centromérica, o viceversa; y el nuevo análisis SPRT puede comenzar con este SNP siguiente.

Por otra parte, en una realización, si la clasificación para SPRT no se alcanzaba con el SNP, entonces también se puede pasar a un SNP vecino de una manera similar, excepto que los recuentos para el siguiente SNP se pueden agregar al anterior y que luego se puede volver a realizar SPRT. Este proceso puede continuar hasta que se haya alcanzado el umbral de clasificación. La FIG. 5A y la FIG. 5B ilustran el funcionamiento de este proceso para los análisis de Tipo A y de Tipo B. En una realización, las clasificaciones se analizan juntas para hacer una clasificación total para una región. Por ejemplo, si se obtiene una clasificación para un primer grupo de SNPs y para el siguiente grupo de SNPs, la clasificación de los dos puede ser comparada para ver si la clasificación es consistente.

La FIG. 6 ilustra el efecto de cambiar la relación de probabilidad para la clasificación SPRT (Zhou W et al. Nat Biotechnol 2001; 19: 78-81; Karoui NE et al. Statist Med 2006; 25: 3124-33). En general, una proporción de probabilidad más baja para la clasificación, por ejemplo, 8, puede permitir que la clasificación se haga más fácilmente. Esto puede resultar en un mayor número de regiones clasificadas dentro del genoma. Sin embargo, cabe esperar que algunas de esas regiones se clasifiquen erróneamente. Por otra parte, una mayor probabilidad de clasificación, por ejemplo, 1200, sólo puede permitir la clasificación cuando se han puntuado más SNPs. Esto puede resultar en un número menor de regiones clasificadas dentro del genoma. Se puede esperar que el número y la proporción de regiones clasificadas erróneamente sean menores cuando se compara con situaciones en las que se utilizaba un umbral de clasificación más bajo.

En una realización, se hace una clasificación solamente si dos clasificaciones consecutivas de SPRT dan como resultado el mismo haplotipo (denominado algoritmo de "dos bloques consecutivos"). En un aspecto, el algoritmo de "dos bloques consecutivos" puede aumentar la precisión de la clasificación. En algunas realizaciones, para cualquier tramo de secuencia, una realización puede realizar primero un análisis SPRT para SNPs de Tipo A, y luego hacer otro análisis SPRT para los SNPs de Tipo B. En una realización, se puede considerar el escenario para un tramo de secuencia para el cual los SNP de Tipo A y Tipo B forman dos grupos de entrelazamiento de hitos genéticos (p. ej. SNPs). En las realizaciones que utilizan el algoritmo de "dos bloques consecutivos", los dos bloques pueden ser de tipos diferentes.

Los resultados de SPRT de los análisis de Tipo A y de Tipo B pueden permitir comprobar la concordancia o discordancia en los resultados de su clasificación. Para mejorar la precisión de la clasificación, una realización ("enfoque de entrelazado") sólo podría hacer una clasificación si tanto el análisis del tipo A como el del tipo B para una determinada región genómica puede dar resultados coherentes. Si los dos análisis dan resultados discordantes, se pueden ver los resultados de clasificación de las dos regiones contiguas de clasificación próximas a la región, una en el extremo centromérico y la otra en el extremo telomérico. Si estas dos regiones contiguas producen resultados concordantes, entonces se puede clasificar la primera región como un haplotipo continuo con estas dos regiones. Si estas dos regiones contiguas no producen resultados concordantes, entonces se puede pasar a las siguientes dos regiones contiguas hasta que se vea la concordancia. Una variante de este tema es moverse en una sola dirección y tomar los resultados de la clasificación de la siguiente, o dos, o incluso más regiones contiguas como los resultados de la región original en cuestión. El principio general es utilizar los resultados de clasificación de regiones genómicas adyacentes para confirmar los resultados de clasificación de una región en particular.

III. DETERMINACIÓN DE LOS ALELOS PATERNOS HEREDADOS POR EL FETO

La FIG. 7 es un diagrama de flujo de un método 700 para determinar al menos una parte del genoma de un feto no nacido de una mujer embarazada heredada del padre. El método 700 analiza las moléculas de ácido nucleico (fragmentos) de una muestra biológica obtenida de la mujer embarazada para determinar el genoma del feto. La muestra contiene una mezcla de ácidos nucleicos maternos y fetales.

En la etapa 710, cada una de una pluralidad de moléculas de ácido nucleico de la muestra biológica se analiza para identificar una localización de la molécula de ácido nucleico en el genoma humano y para determinar un tipo de alelo de la molécula de ácido nucleico. De este modo, los genotipos de las moléculas de ácido nucleico en una localización particular (locus) se pueden determinar en una realización. Cualquiera de los métodos descritos anteriormente y en otra parte se pueden usar para este análisis.

En la etapa 720, se determina una primera pluralidad de loci en la que el genoma paterno es heterocigótico y el genoma materno es homocigótico. En una realización, la primera pluralidad de loci se obtiene determinando los genomas paterno y materno. Los genomas se pueden extraer para loci genómicos en los que el padre es heterocigótico y la madre es homocigótica.

En la etapa 730, el haplotipo heredado por el feto no nacido del padre en la porción del genoma cubierta por la primera pluralidad de loci se determina basado en los genotipos determinados en la primera pluralidad de loci. En una realización, el alelo de cada uno de estos loci que posee el padre, pero que está ausente en el genoma de la madre, se busca en los datos analíticos (por ejemplo, datos de secuenciación). La combinación de estos alelos indicaría los haplotipos de los cromosomas que el feto ha heredado del padre.

En otra realización, si se conocen los haplotipos de cada uno de los cromosomas o las regiones cromosómicas de interés en el genoma del padre, también se puede determinar entonces dónde se ha producido la recombinación meiótica durante la espermatogénesis en el padre. Por lo tanto, la recombinación meiótica paterna se observa cuando el haplotipo de un tramo de ADN en un cromosoma paternalmente heredado difiere entre el feto y el padre. La inclusión de tal información de recombinación puede ser útil cuando los datos analíticos (por ejemplo, los datos de secuenciación) se usan para el diagnóstico prenatal de una enfermedad genética mediante el análisis de ligamiento a polimorfismos genéticos.

IV. EL PADRE Y LA MADRE SON HETEROCIGÓTICOS PARA UNA REGIÓN GENÓMICA

Las realizaciones pueden referirse a un escenario en el que el padre y la madre sean heterocigotos para una región genómica. Este escenario puede ser particularmente relevante en las familias en las que el padre y la madre sean consanguíneos. Cuando una enfermedad se asocia con una mutación predominante que ha resultado de un gran efecto fundador también puede ser relevante. En tales circunstancias, es de esperar que, si el padre y la madre del feto no nacido son ambos portadores del gen mutante, entonces el haplotipo del cromosoma que lleva la copia mutante del gen puede ser esencialmente idéntico, a excepción de la ocurrencia de eventos de recombinación meiótica. Este tipo de análisis puede ser especialmente útil para enfermedades autosómicas recesivas tales como la fibrosis quística, beta-talasemia, anemia de células falciformes y enfermedad de hemoglobina E.

La FIG. 8 es un diagrama de flujo de un método 800 para determinar al menos una porción del genoma de un feto no nacido en una región en la que la madre y el padre son heterocigóticos según las realizaciones de la presente invención.

En la etapa 810, se determinan una primera pluralidad de loci en la que el padre y la madre son ambos heterocigóticos. En una realización, los primeros loci pueden determinarse por cualquiera de los métodos mencionados en este documento. Por ejemplo, todas o regiones de los genomas parentales pueden secuenciarse, o diferentes partes genotipadas para encontrar los primeros loci. Así, cada uno de los dos haplotipos paternos y cada uno de los dos haplotipos maternos en la primera pluralidad de loci pueden ser conocidos.

Como ejemplo, la FIG. 9 muestra haplotipos de un padre y una madre que son ambos heterocigóticos en una región genómica particular. Como se muestra, ambos padres tienen un gen mutante (alelo) en la región 1. Específicamente, Hap I del padre y Hap III de la madre tienen el gen mutante. También como se muestra, el padre y la madre pueden tener cada uno la otra copia del cromosoma que lleva la copia de tipo salvaje del gen. Específicamente, Hap II del padre y Hap IV de la madre tienen el gen de tipo salvaje. Por lo tanto, este ejemplo tiene relevancia en determinar si un feto ha heredado un gen mutante. Los cromosomas del padre y la madre que llevan el gen de tipo salvaje tienen un haplotipo idéntico en la vecindad inmediata del gen, pero pueden tener haplotipos divergentes más alejados del gen. Como este cromosoma probablemente tendría un origen ancestral diverso, este cromosoma probablemente no tendría haplotipos idénticos entre el padre y la madre a lo largo de todo el cromosoma.

En la etapa 820, se determina una segunda pluralidad de loci en la que el padre es heterocigótico, pero en el que la madre es homocigótica. Como se muestra, la primera y segunda pluralidad de loci están en el mismo cromosoma. La región 2 muestra dichos segundos loci. La región 2 se puede elegir de tal manera que el padre sea heterocigótico para uno o más SNPs en esta región mientras que la madre sea homocigótica en esta región.

En la etapa 830, se pueden analizar fragmentos de una muestra de la mujer embarazada para identificar una localización en el genoma humano y un genotipo. La localización se puede usar para determinar si un fragmento (molécula de ácido nucleico) incluye uno o más de los primeros loci o uno o más de los segundos loci. Esta información se puede utilizar para determinar el haplotipo heredado del padre y el haplotipo heredado de la madre.

En la etapa 840, cuyos dos haplotipos paternos han sido heredados por el feto, se determina analizando los genotipos determinados de la pluralidad de moléculas de ácido nucleico de la muestra biológica en al menos uno de los segundos loci. Por ejemplo, los alelos de SNP que están presentes únicamente en el genoma del padre, pero ausentes en el genoma de la madre, tales como el alelo T marcado con * y el marcador de alelo A por + en la Fig. 9, se pueden buscar a partir de los datos analíticos (por ejemplo, localización y genotipo resultantes de la etapa 710) de la muestra biológica materna. Como puede hacerse para el método 700, si el alelo T marcado con * se detecta a

partir del plasma materno, entonces significa que el haplotipo II (Hap II) es heredado por el feto del padre. Por el contrario, si el alelo A marcado por + se detecta a partir de plasma materno, entonces significa que Hap I es heredado por el feto del padre.

En la etapa 850, comparación de las cantidades relativas de los genotipos determinados de moléculas de ácido nucleico en más de una de la primera pluralidad de loci. En una realización, las cantidades en cada locus se agregan y se comparan las cantidades relativas de los haplotipos maternos. Las cantidades relativas pueden referirse a números contados, distribuciones de tamaños y cualquier otro parámetro que pueda transmitir información sobre qué genotipo está en el genoma fetal en un locus particular.

En la etapa 860, basado en el haplotipo paterno determinado heredado por el feto y basado en la comparación de las cantidades relativas, determinación del haplotipo heredado por el feto no nacido de la madre en la porción del genoma cubierta por la primera pluralidad de loci. Por lo tanto, se puede llevar a cabo un análisis RHDO (por ejemplo, como se describió anteriormente) de SNPs en la Región 1 a partir de los datos analíticos de la muestra biológica materna para determinar cuál de los dos haplotipos maternos ha sido heredado por el feto tomando el haplotipo paterno heredado por el feto en la Región 2. En una realización, se supone que no hay recombinación entre las Regiones 1 y 2 cuando estas regiones se pasan de los padres al feto.

Por ejemplo, considérese el escenario cuando se ha determinado que el feto ha heredado Hap I del padre a través del análisis de la Región 2. Entonces, la herencia fetal de Hap III (que es idéntica a Hap I en la Región 1) de la madre resultará en la sobrerrepresentación de Hap III relativa a Hap IV en el plasma materno. Por el contrario, si el feto ha heredado Hap IV de la madre, entonces se observará una representación igual de Hap III y Hap IV en el plasma materno.

Como otro ejemplo, considérese el escenario cuando se ha determinado que el feto ha heredado Hap II del padre a través del análisis de la Región 2. Entonces, la herencia fetal de Hap IV (que es idéntica a Hap II en la Región 1) de la madre dará lugar a la sobre-representación de Hap IV relativa a Hap III en el plasma materno. Por el contrario, si el feto ha heredado Hap III de la madre, entonces la representación igual de Hap III y Hap IV se observará en el plasma materno.

En las secciones anteriores, se ha deducido el genoma fetal y la concentración fraccional de ADN fetal utilizando los datos obtenidos de la secuenciación del ADN plasmático materno, así como la información del genotipo de los padres del feto. En las secciones siguientes, se describen realizaciones para deducir la concentración fraccional de ADN fetal y el genotipo fetal sin información previa de los genotipos-haplotipos maternos y paternos.

V. DETERMINACIÓN DE LA CONCENTRACIÓN DE ADN FETAL FRACCIONAL

En algunas realizaciones, una etapa opcional es determinar una concentración fraccionada del ADN fetal. En varios aspectos, esta concentración fraccionada puede guiar la cantidad de análisis (por ejemplo, la cantidad de secuenciación requerida) o permitir estimar la precisión del análisis para una cantidad dada de datos (por ejemplo, profundidad de la cobertura de la secuencia del genoma). La determinación de la concentración fraccional de ADN fetal también puede ser útil para determinar un corte para determinar una clasificación de qué haplotipo y/o genotipo se heredan.

En una realización, la concentración fraccionada de ADN fetal se puede determinar minando los datos analíticos (por ejemplo, como se puede obtener en las etapas 140 y 710) para loci que son homocigóticos para el padre y para la madre, pero con alelos diferentes. Por ejemplo, para un SNP con dos alelos, A y G; el padre puede ser AA y la madre puede ser GG, y viceversa. Para tales loci, el feto sería un heterocigoto obligado. En el ejemplo anterior, el genotipo fetal sería AG, y una proporción del alelo A en la muestra materna se puede usar para determinar la concentración fraccionada de ADN fetal. En otra realización, se puede hacer un análisis estadístico para determinar un locus donde la madre sea homocigótica y el feto sea heterocigótico. De esta manera, no se necesita información previa sobre el genoma de la madre o el genoma paterno.

Como alternativa a la minería de los datos analíticos, la concentración fraccional de ADN fetal también puede determinarse mediante otro enfoque, tal como el uso de ensayos de PCR, ensayos de PCR digital o ensayos basados en espectrometría de masas, en un panel de marcadores genéticos polimórficos (Lun FMF et al Clin Chem 2008; 54: 1664-1672). Otra alternativa es utilizar uno o más loci genómicos que presentan diferente metilación del ADN entre el feto y la madre (Poon LLM et al. Clin Chem 2002; 48: 35-41; Chan KCA y col. Clin Chem 2006; 52: 2211-2218; Patente de EE.UU. 6.927.028). Todavía otra alternativa es usar una concentración de ADN fetal fraccionada aproximada determinada a partir de una población de referencia, p. ej., a una edad gestacional similar. Sin embargo, como la concentración de ADN fetal fraccional podría variar de muestra a muestra, se puede esperar que este último enfoque sea menos preciso que si la concentración se mide específicamente para la muestra que se está probando.

A. Determinación de la concentración fraccional para el heterocigoto obligatorio

En las realizaciones en las que el feto es un heterocigoto obligado, se puede determinar la concentración fraccionada de ADN fetal usando la siguiente serie de cálculos (por ejemplo, usando secuenciación masivamente paralela). Sea p los recuentos del alelo fetal que está ausente del genoma materno. Sea q los recuentos del otro alelo, es decir, el alelo que es compartido por los genomas materno y fetal. La concentración fraccionada de ADN fetal se da por la siguiente ecuación:

$$\frac{2p}{p+q}$$

En una implementación, este cálculo se puede realizar sobre los datos acumulativos a través de diferentes loci genéticos polimórficos o características genéticas polimórficas que cumplen la configuración del genotipo parental (por ejemplo, ambos progenitores son homocigóticos, pero para alelos diferentes).

B. Determinación basada en SNP informativos

La concentración fraccionada de ADN fetal también puede determinarse para cualquier locus en el que la madre sea homocigótica y el feto sea heterocigótico, y no sólo cuando la madre sea homocigótica para un alelo y el padre sea homocigótico para un alelo diferente. Ambos métodos proporcionan si un locus es informativo. La expresión "SNP informativo" se puede usar en diferentes contextos dependiendo de qué información se desea. En un contexto, la información es un alelo en el genoma fetal en un lugar particular que no está presente en el genoma materno en ese locus. Por lo tanto, el subconjunto de SNPs que la madre es homocigota y el feto es heterocigoto puede ser referido como "SNPs informativos" para el contexto de la determinación de la concentración de ADN fetal. Los casos en los que la madre y el feto son heterocigotos, pero por lo menos un alelo diferente, también se pueden utilizar como un SNP informativo. Sin embargo, los SNP trialélicos son relativamente poco frecuentes en el genoma.

La FIG. 10 es un diagrama de flujo que ilustra un método 1000 para determinar la concentración fraccionada de material fetal en una muestra materna de acuerdo con las realizaciones de la presente invención. En la etapa 1010, se pueden analizar fragmentos de una muestra de la mujer embarazada para identificar una localización en el genoma humano y un tipo de alelo (que puede conducir a una determinación del genotipo en la localización). En una realización, los fragmentos se analizan secuenciando una pluralidad de moléculas de ácido nucleico a partir de la muestra biológica obtenida de la mujer embarazada. En otras realizaciones, se puede usar PCR en tiempo real o PCR digital.

En la etapa 1020, se determina que uno o más primeros loci son informativos. En algunas realizaciones, el genoma materno es homocigótico, pero se detecta un alelo no materno en la muestra en un locus informativo. En una realización, el genoma fetal es heterocigótico en cada primer loci y el genoma materno es homocigótico en cada primer loci. Por ejemplo, el genoma fetal puede tener un primer y segundo alelo respectivos (p. ej., TA) en un primer locus, y el genoma materno puede tener dos del respectivo segundo alelo (p. ej., AA) en el primer locus. Sin embargo, tales loci pueden no ser conocidos a priori, por ejemplo, en situaciones en las que el feto no es un heterocigoto obligado.

En una realización para determinar un locus informativo, se consideran los SNPs en los que la madre es homocigótica. Para los SNPs que la madre es homocigótica, el feto es homocigótico para el mismo alelo o es heterocigótico. Por ejemplo, si un SNP es polimórfico para A y T, y la madre tiene un genotipo de AA, el genotipo del feto es AA o TA. En este caso, la presencia del alelo T en la muestra de plasma materno indicaría que el genotipo fetal es TA en lugar de AA. Ciertas realizaciones pueden dirigir cuánto de una presencia del alelo T indica un genotipo de TA calculando un corte necesario, como se describe a continuación.

En la etapa 1030, para al menos uno de los primeros loci, se determina un primer número p de recuentos del primer alelo respectivo y un segundo número q de recuentos del segundo alelo respectivo. En una realización, los recuentos de los alelos fetales específicos (el alelo T) y los alelos compartidos (el alelo A) en el plasma materno se pueden determinar mediante una variedad de métodos, por ejemplo, pero sin limitarse a PCR en tiempo real, PCR digital y secuenciación masiva paralela.

En la etapa 1040, la concentración fraccional se calcula sobre la base del primer y segundo números. En una realización, en una mujer embarazada con genotipo AA y el genotipo de su feto TA, la concentración fraccional de ADN fetal (f) se puede calcular usando la ecuación: $f = 2 \times p / (p + q)$, donde p representa el recuento para el alelo fetal específico (alelo T) y q representa los recuentos para el alelo compartido por la madre y el feto (alelo A).

En otra realización, mediante el uso de SNPs informativos múltiples, la concentración fraccionada de ADN fetal en el plasma materno se puede estimar con mayor precisión. Para el uso de los recuentos de alelos de SNPs múltiples (un total de n SNPs), la concentración fraccional de ADN fetal (f) se puede calcular usando la ecuación

$$f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)}$$

dónde p_i representa los recuentos para el alelo fetal-específico para el SNP $_i$ informativo; q_i representa los recuentos para el alelo compartido por la madre y el feto para el SNP $_i$ informativo; y n representa el número total de SNP informativos. El uso de los recuentos de alelos de SNPs múltiples puede aumentar la precisión de la estimación de la concentración de ADN fetal fraccionada.

C. Concentración fraccional sin información genética explícita de los padres

Se describe ahora un método para determinar la concentración fraccional de ADN fetal en una muestra de plasma materno que no requiere información previa con respecto a los genotipos del feto y la madre. En una realización, la identificación de SNP informativos se realiza a partir de los recuentos de diferentes alelos en estos loci SNP en plasma materno. De este modo, se puede usar el método 1000, junto con la determinación de los SNP informativos basándose en las realizaciones descritas a continuación. En primer lugar, se proporciona una descripción de las probabilidades para ayudar a comprender un cálculo de un corte que se utiliza para identificar SNP informativos.

En una realización, la probabilidad de detectar el alelo fetal específico sigue la distribución de Poisson. La probabilidad (P) de detectar el alelo fetal específico se puede calcular usando la siguiente ecuación: $P = 1 - \exp(-f \times N/2)$, donde f representa la concentración fraccional de ADN fetal en la muestra de plasma materno, N representa el número total de moléculas correspondientes a este locus SNP particular; y $\exp()$ representa la función exponencial. En un aspecto, P se puede considerar una distribución esperada ya que no es una distribución resultante de medir una cantidad de moléculas a través de muchas muestras. En otras realizaciones, pueden usarse otras distribuciones.

Suponiendo que la concentración fraccional de ADN fetal es del 5% (un valor típico para el primer trimestre de embarazo) y que se analizan 100 moléculas (maternas + fetales) correspondientes a este locus SNP (equivalente a la cantidad contenida en 50 genomas diploides), la probabilidad de detectar el alelo fetal específico (el alelo T) es $1 - \exp(-0,05 \times 100/2) = 0,92$. La probabilidad de detectar el alelo fetal específico aumentaría con la concentración fraccionada de ADN fetal y el número de moléculas que se analizan para el locus SNP. Por ejemplo, si la concentración fetal de ADN es del 10% y se analizan 100 moléculas, la probabilidad de detectar el alelo fetal específico es 0,99.

Por lo tanto, en un locus de SNP para el cual la madre es homocigótica, la presencia de un alelo diferente del materno en el plasma materno puede indicar que el SNP es "informativo" para el cálculo de la concentración fraccional de ADN fetal. La probabilidad de que falte cualquier SNP informativo puede depender del número de moléculas analizadas. En otras palabras, para cualquier confianza deseada de detectar los SNP informativos, el número de moléculas que necesita ser analizado para obtener una precisión deseada se puede calcular de acuerdo con la función de probabilidad de Poisson.

Utilizando el análisis anterior, algunas realizaciones pueden determinar si un locus es informativo o no cuando no se conoce el genotipo de la madre. En una realización, se identifican loci en los que se detectan dos alelos diferentes en la muestra de plasma materno. Por ejemplo, para un locus SNP con dos alelos posibles A y T, tanto el alelo A como el alelo T se detectan en el plasma materno.

La FIG. 11 es un diagrama de flujo de un método 1100 para determinar si un locus es informativo según las realizaciones de la presente invención. En una realización, el método 1100 puede usarse para implementar la etapa 1020 del método 1000. En otra realización, una etapa del método 1100 es determinar un valor de corte basado en una distribución estadística y otro utiliza el valor de corte para determinar si un locus (SNP) es informativo.

En la etapa 1110, se determina un valor de corte para un número de recuentos previstos del primer alelo respectivo en el locus específico. En una implementación, el valor de corte predice si el genoma materno es homocigótico y el genoma fetal es heterocigótico. En una realización, el valor de corte se determina basándose en una distribución estadística del número de cuentas para diferentes combinaciones de homocigosidad y heterocigosidad en el locus específico. Por ejemplo, se puede predecir una distribución de frecuencia alélica utilizando la función de distribución de Poisson.

En la etapa 1120, sobre la base de un análisis de las moléculas de ácido nucleico de la muestra materna (por ejemplo, desde la etapa 1010), se detectan un primer alelo y un segundo alelo en el locus. Por ejemplo, un conjunto de fragmentos podría ser asignado al locus que se analiza y ser detectado el primer alelo o el segundo alelo. El primer alelo puede corresponder a uno de los respectivos primeros alelos de la etapa 1020, y el segundo alelo puede corresponder a uno de los respectivos segundos alelos. En una realización, si no se detectan dos alelos diferentes, entonces se sabe que el locus no es informativo.

En la etapa 1130, se determina un número de cuentas reales del primer alelo respectivo en el locus basado en el análisis de las moléculas de ácido nucleico. Por ejemplo, se pueden contar los resultados de secuenciación de la pluralidad de moléculas de ácido nucleico para determinar el número de veces que un fragmento que tiene un genotipo del primer alelo está mapeado al locus.

En la etapa 1140, el locus se identifica como uno de los primeros loci basándose en una comparación del número de cuentas reales con el valor de corte. En un aspecto, se puede utilizar un valor de corte para diferenciar entre tres posibilidades: (a) la madre es homocigótica (AA) y el feto es heterocigótico (AT); (B) la madre es heterocigótica (AT) y el feto es heterocigótico (AT); y (c) la madre es heterocigótica (AT) y el feto es homocigótico para (AA) o (TT). A título ilustrativo, los ejemplos siguientes asumen que el genotipo fetal es AA en el escenario (c). Sin embargo, el cálculo sería el mismo si el genotipo del feto fuera TT. Un locus informativo tendría la posibilidad (a).

En una realización, el locus se identifica como uno de los primeros loci cuando el número de cuentas reales es menor que el valor de corte. En otra realización, también se puede usar un umbral inferior para asegurarse de que no se produjo un mapeado no esencial.

La realización para determinar el corte se describe ahora. Basándose en la concentración de ADN fetal fraccional fisiológicamente posible (esta información está disponible a partir de estudios previos) y el número total de moléculas correspondientes al locus SNP, la distribución de los recuentos alélicos puede predecirse para los tres posibles escenarios anteriores. Basándose en la distribución prevista, se puede determinar un valor de corte para interpretar los conteos alélicos observados en el plasma materno para determinar si un SNP es "informativo" (es decir, el escenario (a)) o no.

La concentración fraccional de ADN fetal oscila típicamente entre el 5% y el 20% en el inicio del embarazo y oscila entre el 10% y el 35% en el final del embarazo (Lun et al., Micro fluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem 2008; 54: 1664-72). Por tanto, en una realización, se determinaron las distribuciones previstas de los recuentos alélicos para una concentración fraccional de 5% y 20% de ADN fetal.

La FIG. 12A muestra la distribución prevista de los recuentos para el alelo T (el alelo menos abundante en los escenarios (a) y (c)) para los tres escenarios con una concentración de ADN fetal fraccional de 20%. La FIG. 12B muestra la distribución prevista de los recuentos para el alelo T (el alelo menos abundante para los escenarios (a) y (c)) para los tres escenarios con la suposición de 5% de ADN fetal. En ambos modelos predichos, se supone que un total de 200 moléculas se analizaron para el locus SNP.

Usando la presencia de 40 recuentos del alelo menos abundante (el alelo T) como corte, las tres posibilidades pueden ser discriminadas estadísticamente. En otras palabras, para cualquier locus SNP con dos alelos detectados en el plasma materno y con un total de 200 moléculas analizadas, si la frecuencia alélica del alelo menor (el alelo menos abundante) es menor de 40, el locus SNP puede ser considerado "informativo". Para las concentraciones fraccionales de ADN fetal del 5% y del 20%, la diferenciación de SNPs "informativos" (escenario (a)) de los SNPs para los cuales la madre es heterocigótica (los escenarios (b) y (c)) sería 100% exacta.

En la práctica, el número total de moléculas detectadas puede ser diferente para diferentes SNPs. Para cada locus SNP, se puede construir una curva de distribución específica prevista teniendo en cuenta el número total de moléculas detectadas en la muestra de plasma materno que cubre el locus SNP. En otras palabras, el recuento de corte para determinar si un SNP es informativo o no puede variar entre SNPs y depende del número de veces que el locus SNP ha sido contado.

La siguiente tabla muestra los recuentos de alelos de tres loci SNP en plasma materno para una muestra de plasma materno que fue secuenciada. Para cada uno de los tres SNPs, se detectan dos alelos diferentes en la muestra de plasma materno. El número total de recuentos detectados en el plasma materno correspondiente a estos tres SNP es diferente.

Locus SNP	SNP id	Alelo (recuentos)	Alelo (recuentos)	Total no. de recuentos
1.	rs3107146	A (10)	G (163)	173
2.	rs7522344	G (9)	T (112)	121
3.	rs2272908	A (72)	G (62)	134

Las distribuciones previstas para los recuentos del alelo menos abundante para una concentración fraccionada de ADN fetal del 20% y los diferentes conteos totales de moléculas correspondientes a un SNP se muestran en las Figs. 13A, 13B y 14. Las distribuciones predichas se obtuvieron usando una concentración fetal de ADN supuesta del 20%, ya que esto representa el límite más alto de concentración de ADN fetal en el primer trimestre. Cuanto mayor es la concentración de ADN fetal, más se espera que se superponga entre las curvas de distribución del alelo menor para el cual la madre es homocigótica para el alelo principal contra aquella cuando la madre es heterocigótica. Por lo tanto, es más específico para derivar los puntos de corte para el menor número de alelos utilizando una mayor concentración de ADN fetal para la predicción de SNP informativos.

La FIG. 13A muestra una distribución predicha para los recuentos del alelo menos abundante con un número total de 173 moléculas y una concentración fraccionada de ADN fetal del 20%. En una realización, basándose en esta distribución, un criterio de corte de menos de 40 para los recuentos del alelo menos abundante puede ser adecuado para identificar los SNP informativos. Como los recuentos para el alelo A son 10, el locus SNP no. 1 se considera "informativo" para el cálculo de la concentración fraccionada de ADN fetal.

La FIG. 13B muestra una distribución predicha para los recuentos del alelo menos abundante con un número total de 121 moléculas y una concentración fraccionada de ADN fetal del 20%. En una realización, basándose en esta distribución, un valor de corte inferior a 26 para los recuentos del alelo menos abundante puede ser adecuado para identificar los SNP informativos. Como el número de recuentos para el alelo T es 9, el locus SNP no. 2 se considera "informativo" para el cálculo de la concentración fraccionada de ADN fetal.

La FIG. 12 muestra una distribución predicha para los recuentos del alelo menos abundante con un número total de 134 moléculas y una concentración fraccionada de ADN fetal del 20%. En una realización, basándose en esta distribución, un valor de corte inferior a 25 para los recuentos del alelo menos abundante puede ser adecuado para identificar los SNP informativos. Como el número de recuentos para el alelo T es 62, el locus SNP no. 3 se considera como "no informativo" y no se usaría para el cálculo de la concentración fraccional de ADN fetal.

En algunas realizaciones, usando la ecuación $f = 2 \times p / (p + q)$, la concentración fraccional de ADN fetal se puede calcular usando los recuentos de alelos para SNP 1 y 2 y combinados. Los resultados se muestran a continuación.

Cálculo basado en el locus SNP	Concentración fraccional de ADN fetal
1.	$10 \times 2 / (10 + 163) = 11,6\%$
2.	$9 \times 2 / (9 + 112) = 14,9\%$
1. y 2.	$(10+9) \times 2 / (10 + 9 + 163 + 112) = 12,9\%$

D. Determinación de la Cobertura Profunda del Genoma Fetal

Además de obtener una concentración fraccionada, las realizaciones pueden determinar un porcentaje de cobertura del genoma fetal que el procedimiento analítico (por ejemplo, la secuenciación) en la etapa 1010 ha logrado. En algunas realizaciones, pueden utilizarse loci informativos para determinar el porcentaje de cobertura. Por ejemplo, se puede usar cualquiera de los ejemplos anteriores. En una realización, pueden usarse los loci en los que el feto es un heterocigoto obligado. En otra realización, se pueden usar los loci en los que se determina que el feto es heterocigótico y la madre es homocigótica (por ejemplo, usando el método 1100).

Los fragmentos que se han asignado a los loci informativos se pueden utilizar para determinar una proporción de la cobertura. En una realización, se determina una proporción de loci de la primera pluralidad de loci en la que se detecta un primer alelo respectivo a partir de los resultados de la secuenciación. Por ejemplo, si el feto es TA en un locus y la madre es AA en el locus, entonces el alelo T debe ser detectado en los resultados de secuenciación si ese locus ha sido secuenciado. Por lo tanto, la proporción del genoma fetal que se ha secuenciado a partir de la muestra biológica se puede calcular sobre la base de esta proporción. En una realización, la proporción de los primeros loci donde se ve el alelo fetal específico puede tomarse como el porcentaje de cobertura del genoma fetal. En otras realizaciones, la proporción puede ser modificada en base a donde están los loci. Por ejemplo, se puede determinar un porcentaje de cobertura para cada cromosoma. Como otro ejemplo, el porcentaje se puede estimar en menos que la proporción si los primeros loci no forman una buena representación del genoma. Como otro ejemplo, puede proporcionarse un intervalo en el que la proporción es un extremo del intervalo. Mientras que un alto porcentaje, es decir, aproximándose al 100%, significa una cobertura cercana a completa del genoma fetal, la mayoría de las enfermedades genéticas pueden diagnosticarse con una cobertura mucho menor que 100%, p. ej. 80%, o 50%, o menos.

VI. SIN INFORMACIÓN PREVIA DEL GENOMA MATERNO Y PATERNO

En las secciones anteriores, algunas realizaciones han determinado un mapa genético de un feto (o una porción de un genoma fetal) cuando se conocen los haplotipos de la madre y los genotipos del padre. Otras realizaciones han demostrado que la concentración fraccionada de ADN fetal puede determinarse analizando el ADN plasmático materno sin conocimiento previo sobre los genotipos de la madre, el padre o el feto. En aún otras realizaciones, se describe adicionalmente un método para determinar el mapa genético de un feto (o una porción de un genoma fetal) usando el análisis RHDO sin información previa de los genotipos/haplotipos maternos y paternos.

En una realización, se utiliza la información de haplotipos de referencia (por ejemplo, comunes o conocidos) de la población a la que pertenecen los padres. Esta información se puede utilizar para deducir los haplotipos maternos y paternos. Un ejemplo se utiliza para ilustrar el principio de este método. La información relativa a tales haplotipos de referencia puede obtenerse, por ejemplo, en el sitio web del Proyecto Internacional HapMap (hapmap.ncbi.nlm.nih.gov/).

Como parte de un ejemplo ilustrativo, se supone que hay tres haplotipos de referencia (Hap A, Hap B y Hap C como se muestra en la Figura 15A) que están presentes en la población. Cada uno de estos tres haplotipos consta de 14 loci SNP y, para cada locus, hay dos posibles alelos. En este ejemplo, el padre posee Hap B y Hap C mientras que la madre posee Hap A y Hap B, como se muestra en la FIG. 15B. Este ejemplo supone que el feto hereda Hap A de la madre y Hap C del padre. Por lo tanto, el feto posee Hap A y Hap C, como se muestra en la FIG. 15B.

La FIG. 16 es un diagrama de flujo de un método 1600 para determinar al menos parte de un genoma fetal cuando se conoce un conjunto de haplotipos de referencia, pero los haplotipos parentales no se conocen, según las realizaciones de la presente invención.

En la etapa 1610, la muestra materna se puede analizar para identificar los SNPs en los que la madre es homocigótica y el feto es heterocigótico. Este análisis puede hacerse de una manera similar a la determinación de si un locus es informativo, como se ha descrito anteriormente. De este modo, en una realización, pueden usarse los métodos 1000 y/o 1100. En otras realizaciones descritas anteriormente, los genomas materno y paterno pueden analizarse para determinar la información para realizar la cartografía del genoma fetal.

La FIG. 17 muestra un ejemplo de determinación de loci informativos a partir del análisis de fragmentos de ADN de una muestra materna. Para cada uno de los 14 loci, se determinan los recuentos de los dos alelos para cada locus. Los recuentos de estos alelos se pueden determinar, por ejemplo, pero no se limitan a, la utilización de PCR en tiempo real, PCR digital y la secuenciación masivamente paralela. Para cada uno de estos loci, se detectaron dos alelos diferentes en el plasma materno. En contraste con los SNPs en los que la madre es heterocigótica, la proporción de los dos alelos sería significativamente diferente. El alelo específico del feto (el alelo que el feto hereda del padre) sería mucho menos abundante en comparación con el alelo materno. Los loci informativos 1710 están marcados en la FIG. 17.

En la etapa 1620, se deducen uno o más alelos del haplotipo paterno heredado por el feto. En una realización, cada uno de los loci 1710 puede usarse para determinar el haplotipo paterno heredado. Por ejemplo, el alelo paterno que el feto ha heredado puede ser identificado como el alelo fetal-específico para los loci 1720 porque el alelo fetal-específico es un alelo que es mucho menos abundante que el alelo materno en la muestra materna.

En la etapa 1630, los alelos paternos se comparan con los haplotipos de referencia para determinar el haplotipo heredado del padre. En ciertas realizaciones, se pueden deducir una serie de posibles haplotipos fetales, cada uno con su propia probabilidad. Uno o más de los haplotipos fetales más probables se pueden utilizar para el análisis subsiguiente, o para el diagnóstico clínico.

En el ejemplo mostrado en la FIG. 18, hay tres posibles haplotipos (Hap A, Hap B y Hap C) en la población. A partir del análisis de plasma materno, cuatro SNPs han sido identificados como homocigotos para la madre y heterocigotos para el feto, representando así los alelos paternos que el feto hereda. Los genotipos en estos cuatro SNPs encajan en el patrón de Hap C. Por lo tanto, el feto ha heredado Hap C del padre, como se muestra en la FIG. 19. En otras palabras, para todos los SNP dentro del mismo bloque de haplotipos, se pueden deducir los alelos paternos que el feto ha heredado.

En la etapa 1640, los loci (p. ej., SNPs) en los que la madre es heterocigótica pueden determinarse. En una realización, el análisis de la muestra materna puede proporcionar SNPs que la madre es heterocigótica. Por ejemplo, en cada uno de estos SNPs, se pueden detectar dos alelos diferentes en el plasma materno. En contraste con los SNPs que la madre es homocigótica y el feto es heterocigótico, que el alelo fetal específico sólo aporta una pequeña proporción de los alelos totales en el plasma materno, los recuentos de los dos alelos serían similares para los SNPs donde la madre es heterocigótica. Por lo tanto, el genotipo materno completo para todos los loci SNP dentro del bloque haplotipo podría determinarse a partir de análisis de plasma materno, por ejemplo, como se muestra en la FIG. 20.

En la etapa 1650, los haplotipos maternos se deducen de los genotipos maternos de la etapa 1640 comparando los genotipos en los loci con la información de haplotipo de la población relevante. La FIG. 21 muestra una realización para determinar los haplotipos maternos de los genotipos maternos y los haplotipos de referencia. En el ejemplo que se utiliza, la madre es homocigótica para el alelo G en el tercer locus SNP. Como sólo Hap A y Hap B cumplen este criterio, esto indica que la madre tiene una de las tres combinaciones de haplotipos, a saber, Hap A-HapA, Hap A-Hap B o Hap B-HapB. Además, como la madre es heterocigótica para A y C para el primer SNP, se puede deducir que la madre tiene la combinación de haplotipo de Hap A-Hap B. En una realización, podría resultar más de una posibilidad, y cada posibilidad podría ser probada en el paso siguiente. A partir de los análisis anteriores, se han determinado los haplotipos de la madre y el haplotipo que el feto hereda del padre. La FIG. 22 muestra los haplotipos maternos determinados y el haplotipo heredado paternalmente.

En la etapa 1660, el haplotipo materno heredado por el feto se determina a partir de los haplotipos maternos identificados en la etapa 1650 y el haplotipo heredado paternalmente identificado en la etapa 1630. Utilizando esta información, una realización puede usar el análisis RHDO para determinar qué haplotipo materno se transmite al

feto. Se puede realizar un análisis RHDO de acuerdo con cualquiera de las realizaciones descritas en la presente memoria.

En una realización, para el análisis RHDO, los SNPs en los que la madre es heterocigótica pueden dividirse en dos tipos, a saber, tipo alfa y tipo beta (por ejemplo, como se muestra en la Figura 23 y como se ha descrito anteriormente). Los SNP de tipo alfa se refieren a aquellos loci en los que el alelo paterno que pasa al feto es idéntico al alelo materno localizado en Hap A. Para los SNP de tipo alfa, si el feto hereda Hap A de la madre, el alelo de Hap A estaría sobrerrepresentado en el plasma materno. Por otra parte, si el feto hereda Hap B de la madre, los dos alelos maternos estarían igualmente representados en el plasma materno.

Los SNPs de tipo beta se refieren a aquellos loci en los que el alelo paterno que pasó al feto es idéntico al alelo materno localizado en Hap B. Para los SNPs tipo beta, si el feto hereda Hap B de la madre, el alelo en Hap B estaría sobre-representado en el plasma materno. Sin embargo, si el feto hereda Hap A de la madre, los dos alelos maternos estarían igualmente representados en el plasma materno. La posible sobrerrepresentación de los alelos Hap A o Hap B puede determinarse usando el análisis RHDO.

En algunas realizaciones, para aplicar el análisis de RHDO en una región particular sin información previa de los haplotipos maternos y genotipos paternos, se puede requerir una cobertura de los SNPs relativamente alta en el bloque de haplotipos, por ejemplo, 200 moléculas correspondientes a un locus SNP pueden necesitar ser analizadas en una realización. Esta información se puede obtener, por ejemplo, pero sin limitarse a, de PCR en tiempo real, PCR digital y secuenciación masivamente paralela. En una realización, puede utilizarse la secuenciación dirigida (por ejemplo, mediante una combinación de enriquecimiento diana y secuenciación masivamente paralela) para obtener una información cuantitativa representativa y no sesgada de diferentes alelos dentro de la región diana. Un ejemplo a continuación describe la secuenciación dirigida. Por lo tanto, este análisis RHDO puede aplicarse a los datos de secuenciación dirigida del ADN plasmático materno para determinar qué alelos -haplotipo maternos se transmiten al feto sin información previa con respecto a los genotipos-haplotipos parentales.

VII. DETECCIÓN DE MUTACIÓN DE NOVO

Algunas realizaciones pueden detectar una mutación que el feto ha adquirido. Una mutación de novo es una mutación que no es llevada por el padre o la madre, sino que se produce, por ejemplo, durante la gametogénesis del padre o de la madre o de ambos. Dicha detección tiene utilidad clínica porque las mutaciones de novo desempeñan un papel significativo en la producción de una serie de enfermedades genéticas, p. ej. la hemofilia A y la acondroplasia.

La FIG. 24 es un diagrama de flujo que ilustra un método 2400 de identificación de una mutación de novo en el genoma de un feto no nacido de una mujer embarazada. El feto tiene un padre y una madre que es la mujer embarazada, y el padre tiene un genoma paterno con dos haplotipos y la madre tiene un genoma materno con dos haplotipos, comprendiendo el método:

En la etapa 2410, se secuencian una pluralidad de moléculas de ácido nucleico de una muestra biológica obtenida de la mujer embarazada. Obsérvese que la muestra contiene una mezcla de ácidos nucleicos maternos y fetales.

En la etapa 2420, se identifica una localización de cada una de la molécula de ácido nucleico secuenciada en el genoma humano. En una realización, la asignación de las secuencias puede realizarse mediante secuenciación de extremo único o de extremo pareado. En un aspecto, una asignación al genoma humano para encontrar una ubicación no requiere una coincidencia exacta de cada uno de los nucleótidos para una localización que se encuentra.

En la etapa 2430, para cada una de al menos una parte de los lugares, se determinan una secuencia materna y una secuencia paterna en el lugar en cuestión. Por ejemplo, si se determinan 100 lugares en el paso 2420, entonces se pueden determinar los genomas materno y paterno en estas 100 ubicaciones. En una realización, las secuencias paternas se determinan a partir de una muestra del padre en lugar de usar haplotipos de referencia como se ha descrito anteriormente. Por lo tanto, todavía podría ser detectada una mutación no en un genoma de referencia. En diversas realizaciones, las secuencias maternas pueden obtenerse a partir de una muestra que sólo incluye ADN materno, o también puede obtenerse a partir de la muestra biológica, por ejemplo, utilizando métodos descritos en la presente memoria.

En la etapa 2440, se identifica una primera secuencia en la pluralidad de moléculas de ácido nucleico que no está presente en las secuencias maternas o paternas determinadas. En una realización, una comparación de la primera secuencia con las secuencias maternas o paternas determinadas requiere una coincidencia exacta. Por lo tanto, si el resultado no es exacto, entonces se considera que la primera secuencia no está presente en las secuencias maternas o paternas determinadas. De esta manera, se pueden identificar incluso ligeras mutaciones de novo ya que una mutación de novo puede ser sólo un cambio de nucleótido único. En otra realización, se requiere cierto número de fragmentos de ADN que muestran la secuencia no materna y no paterna para que la secuencia se considere como una mutación de novo. Por ejemplo, se podría usar un corte de 3 fragmentos de ADN para determinar si una secuencia, es decir, la mutación de novo, está presente o no.

En la etapa 2450, se determina una primera concentración fraccionada de la primera secuencia en la muestra biológica. Por ejemplo, el número de fragmentos de ADN que exhiben la primera secuencia podría expresarse como una proporción de todos los fragmentos de ADN detectados desde ese locus.

En la etapa 2460, se determina una segunda concentración fraccionada de ácidos nucleicos fetales en la muestra biológica usando una molécula de ácido nucleico que el feto ha heredado de su padre y que está presente en el genoma paterno pero que no está presente en el genoma materno. Dicha molécula de ácido nucleico puede contener un primer alelo en un lugar donde el padre es homocigótico y la madre es también homocigótica, pero para un alelo diferente, y por lo tanto el feto es un heterocigoto obligado. Pueden usarse loci informativos como se ha descrito anteriormente para determinar la molécula de ácido nucleico a utilizar para determinar la segunda concentración fraccionada.

En otras realizaciones, la segunda concentración fraccional puede determinarse usando otros enfoques, tales como el uso de ensayos de PCR, ensayos de PCR digital o ensayos basados en espectrometría de masas, en el cromosoma Y, un panel de polimorfismos genéticos, es decir, polimorfismos de un solo nucleótido, o polimorfismos de inserción-delección (Lun FMF et al Clin Chem 2008; 54: 1664-1672). Otra alternativa es utilizar uno o más loci genómicos que presentan diferente metilación del ADN entre el feto y la madre (Poon LLM et al. Clin Chem 2002; 48: 35-41; Chan KCA y col. Clin Chem 2006; 52: 2211-2218; Patente de EE.UU. 6.927.028).

En una realización, el diferente estado epigenético se refleja mediante diferentes patrones de metilación del ADN. Los diferentes patrones de metilación del ADN pueden implicar la familia de dominios de asociación RAS 1A (RASSF1A) o la holocarboxilasa sintetasa gen de la (biotina-(propionil-Coenzima A-carboxilasa (hidrolizante de ATP))ligasa (HLCS). La cantidad de fragmentos de ADN con el perfil de metilación del ADN fetal específico puede expresarse como una proporción de todos los fragmentos de ADN que se originan del locus diferencialmente metilado.

En la etapa 2470, la primera secuencia se clasifica como una mutación de novo si la primera y segunda concentraciones fraccionales son aproximadamente iguales. Una secuencia no materna y no paterna que se origina a partir de errores en el proceso de análisis, p. ej. la secuenciación de errores es un evento aleatorio y tiene una baja probabilidad de recurrencia. Por lo tanto, es probable que múltiples fragmentos de ADN que exhiben la misma secuencia no materna y no paterna a cantidades similares a la concentración de ADN fetal fraccional medida para la muestra sean una mutación de novo presente en el genoma fetal en lugar de surgir del error de secuenciación. En una realización, se puede usar un valor de corte para determinar si las concentraciones fraccionadas son las mismas. Por ejemplo, si las concentraciones están dentro de un valor especificado entre sí, entonces la primera secuencia se clasifica como una mutación de novo. En diversas realizaciones, el valor especificado puede ser 5%, 10% o 15%.

EJEMPLOS

I. Ejemplo 1

Para ilustrar las realizaciones de la presente invención, se analizó el siguiente caso. Se reclutó a una pareja que asistía a una clínica de obstetricia para el diagnóstico prenatal de beta-talasemia. El padre era un portador de la delección de los pares de bases -CTTT4 de los codones 41/42 del gen de la beta-globina humana. La madre embarazada era portadora de la mutación A → G en el nucleótido -28 del gen de la beta-globina humana. Se tomaron muestras de sangre del padre y la madre. Para la madre, la muestra de sangre se tomó antes de muestreo de vellosidades coriónicas (CVS) a las 12 semanas de gestación. Después de CVS, se almacenó una porción para el experimento. Un objetivo del experimento era construir un mapa genético de todo el genoma o determinar la secuencia genómica parcial o completa del feto mediante la secuenciación masivamente paralela del ADN plasmático materno.

1. Determinación de los genotipos parentales

Se extrajo el ADN de la capa leucocitaria del padre y la madre, y de la muestra CVS. Estas muestras de ADN se sometieron a análisis por el sistema Affymetrix Genome-Wide Human SNP Array 6.0. Este sistema cuenta con 1,8 millones de marcadores genéticos, incluyendo ~900.000 polimorfismos de nucleótido único (SNPs) y más de ~950.000 sondas para la detección de variaciones de número de copias. El número absoluto y los porcentajes de SNP que muestran diferentes combinaciones de genotipo para el padre, la madre y el feto (CVS) se muestran en la tabla de la FIG. 25A.

Aunque se utilizó el sistema Affymetrix en este ejemplo, en la práctica, cualquier plataforma de genotipificación conocida por los expertos en la técnica podría ser utilizada. De hecho, aparte del genotipado, el ADN de la capa leucocítica del padre y la madre también puede someterse a secuenciación, ya sea sobre una base de genoma completo o para regiones genómicas seleccionadas. Además, se podría usar cualquier fuente de ADN constitucional (por ejemplo, ADN de células bucales, ADN de folículo piloso, etc.) del padre y la madre estableciendo los genotipos parentales.

La muestra CVS se analizó para proporcionar un estándar para la comparación con el mapa genético fetal deducido del análisis de plasma materno. Además, para este experimento, el genotipo de la muestra CVS también puede usarse para construir el haplotipo de la madre para el análisis RHDO. En este escenario, el uso del genotipo CVS para este propósito de construcción de haplotipos sólo se utilizó con fines ilustrativos. En una aplicación clínica de las realizaciones, el haplotipo materno puede ser construido a través del análisis de otros individuos de la familia, por ejemplo, un descendiente anterior, un hermano, los padres u otros familiares de la madre. Los haplotipos maternos de las regiones cromosómicas de interés también pueden construirse por otros métodos bien conocidos por los expertos en la técnica, algunos de los cuales se mencionan en este documento.

Para realizaciones seleccionadas, también se pudo determinar el haplotipo del padre del feto no nacido a analizar. Esta información puede ser particularmente útil para la dosificación relativa de haplotipos para regiones cromosómicas en las que tanto el padre como la madre son heterocigotos.

2. Secuencia masiva paralela del ADN plasmático materno

El ADN plasmático obtenido de la madre se sometió a secuenciación masiva paralela utilizando la plataforma Illumina Genome Analyzer. Se realizó la secuenciación en pares de las moléculas de ADN plasmático. Cada molécula se secuenció en cada extremo para 50 pb, totalizando así 100 pb por molécula. Los dos extremos de cada secuencia se alinearon con el genoma humano repetido-no enmascarado (Hg18 NCBI.36 descargado de UCSC <http://genome.ucsc.edu>) utilizando el programa SOAP2 del Beijing Genomics Institute de Shenzhen (soap.genomics.org.cn) (Li R et al. Bioinformatics 2009, 25 (15): 1966-7) La tabla, FIG. 25B, enumera las estadísticas de alineación de las primeras 20 celdas de flujo. Así, con 20 células de flujo, más de 3.932 millones de lecturas se alinearon con el genoma humano de referencia.

3. Cálculo de las concentraciones fraccionales de ADN fetal

Como se mencionó anteriormente, la concentración fraccionada de ADN fetal en la muestra de plasma materno se puede calcular a partir de los datos de secuenciación. Una forma era analizar los SNPs en los que el padre y la madre eran homocigóticos, pero para diferentes alelos entre sí. Para tales SNPs, el feto sería un heterocigoto obligado para un alelo paternalmente heredado y uno heredado de madre. En una realización, se puede usar cualquiera de los métodos de cálculo descritos en la sección V. En este ejemplo, se realizaron cálculos sobre los datos acumulativos a través de diferentes loci genéticos polimórficos que cumplieran la configuración del genotipo parental (es decir, ambos progenitores eran homocigóticos, pero para alelos diferentes) en diferentes cromosomas. Las concentraciones fraccionales de ADN fetal calculadas para SNPs situados en cromosomas diferentes se enumeran en la columna de la derecha de la FIG. 26. Como puede verse en la tabla, las concentraciones fraccionarias determinadas para SNPs situados en cromosomas diferentes se correlacionan muy estrechamente entre sí.

Como experimento de control de calidad, también se investigaron aquellos SNPs en los que la madre era homocigótica y el padre heterocigótico, a partir del análisis Affymetrix SNP 6.0 de las muestras de la capa leucocitaria (columna media de la figura 26). Se puede observar que, a una profundidad suficiente de la secuenciación del ADN, las concentraciones fraccionales de ADN fetal medidas a partir de este análisis eran muy similares a las medidas para SNPs en las que tanto el padre como la madre eran homocigóticos, pero para alelos diferentes.

En una implementación, cuando se observaba una concordancia cercana de las concentraciones fraccionales de ADN fetal a partir de estos dos tipos de SNPs, se podía concluir que se estaba cerca de la cobertura de secuenciación completa del genoma fetal. En un aspecto, a una menor profundidad de cobertura, las concentraciones fraccionales de ADN fetal medidas para SNPs en las que la madre era homocigótica y el padre era heterocigótico eran superiores a las medidas para SNPs en las que tanto el padre como la madre eran homocigóticos, pero para alelos diferentes. A una menor profundidad de cobertura, la ausencia de un alelo paternalmente único a partir de los resultados de la secuenciación puede tener dos causas posibles: (i) que el feto no había heredado este alelo del padre; y/o (ii) que el feto había heredado este alelo del padre, pero que entonces este alelo faltaba en los resultados de la secuenciación debido a que la profundidad de la secuenciación no era suficiente.

4a. Cálculo del porcentaje de cobertura del genoma fetal

Además, como se mencionó anteriormente, el porcentaje del genoma fetal que se había analizado mediante secuenciación del ADN plasmático materno se pudo determinar mirando el subconjunto de SNPs en los que el padre y la madre eran homocigóticos, pero para alelos diferentes. En esta familia, 45.900 SNPs en la matriz Affymetrix SNP 6.0 pertenecían a este subconjunto. El porcentaje de cobertura del genoma fetal podría deducirse mediante el análisis de los datos de secuenciación del ADN plasmático para ver en qué porcentaje de este subconjunto de SNPs se podría detectar un alelo fetal por secuenciación.

La gráfica de la FIG. 27A ilustra el porcentaje observado de SNPs en este subconjunto en el que un alelo fetal podría ser visto a partir de los datos de secuenciación de las primeras 20 células de flujo analizadas. Así, un alelo fetal se pudo observar en el 94% de tales SNPs. Este grado de secuenciación correspondía a más de 3.932 millones de lecturas, cada una con 100 pb de secuencias. La gráfica de la FIG. 27B muestra la cobertura frente al número de lecturas, en lugar del número de células de flujo. Con el aumento del rendimiento de diferentes plataformas de secuenciación, se espera que el número de células de flujo o ejecuciones que se utilizarían o requerirían para generar este número de lecturas de secuencia o longitud de secuencias disminuiría en el futuro.

En algunas realizaciones, como se detectaron múltiples SNPs en cada región cromosómica o cromosomas, la cobertura del genoma fetal podría ser mucho menor que 94%, mientras que todavía proporciona un mapeo preciso del genoma. Por ejemplo, supongamos que hay 30 SNP informativos en una región cromosómica, pero se detecta un alelo fetal para sólo 20 SNPs de los 30 SNPs. Sin embargo, la región cromosómica todavía puede ser identificada con precisión con los 20 SNPs. Así, en una realización, se puede obtener una exactitud equivalente con una cobertura de menos del 94%.

4b. Cobertura del mapa genético de los alelos que el feto había heredado de su padre

Este análisis ilustrativo se centra en los alelos de SNP en los que el padre era heterocigótico y la madre era homocigótica. En esta familia, 131.037 SNPs en la plataforma Affymetrix SNP 6.0 pertenecían a esta categoría. Un subconjunto de estos SNP consistía en los 65.875 SNPs en los que la madre era homocigótica, mientras que el padre y el feto eran heterocigotos. Con el uso de 20 células de flujo, los alelos paternalmente heredados se pudieron observar en 61.875 de estos SNPs, lo que indica una cobertura del 93,9%. Este último porcentaje encajaba bien con los datos de cobertura porcentual deducidos en el párrafo anterior. La correlación entre la cobertura de alelos paternalmente heredados y el número de lecturas de secuencias asignables y el número de secuencias de células de flujo se muestran en la FIG. 28A y la FIG. 28B, respectivamente.

Para elucidar la especificidad de este enfoque para la detección de genuinos alelos fetales paternalmente heredados, los 65.162 (es decir, 131.037-65.875) se analizaron los SNPs en los que el feto había heredado alelos que eran los mismos que los poseídos por la madre. Para tales SNPs, la detección aparente de alelos diferentes de los poseídos por la madre representaría un falso positivo. Así, entre los 65.162 SNPs, sólo 3.225 falsos positivos (4,95%) se observaron cuando se analizaron 20 células de flujo. Estos falsos positivos pueden ser el resultado de errores de secuenciación o errores de genotipificación del ADN del padre o de la madre o mutaciones de novo en el feto. La correlación entre la tasa de falsos positivos y el número de células de flujo secuenciadas se muestra en la FIG. 29A.

Las tasas de falsos positivos también se pueden estimar considerando el subconjunto de SNPs que tanto el padre como la madre eran homocigóticos y con el mismo alelo. La presencia de cualquier alelo alternativo en el locus particular se consideró un falso positivo. Estos falsos positivos pueden ser el resultado de errores de secuenciación o errores de genotipificación del ADN del padre o de la madre o mutaciones de novo en el feto. Había 500.673 SNPs en este subconjunto. Con los datos de la secuencia de 20 células de flujo, se detectaron resultados falsos positivos en 48.396 SNP (9,67%). La correlación entre la tasa de falsos positivos y el número de células de flujo secuenciadas se muestra en la FIG. 29B. Esta tasa de falsos positivos fue mayor que la estimación utilizando el subconjunto de SNPs que la madre y el feto eran homocigóticos y el padre era heterocigótico. Esto se debe a que en el último subconjunto de SNPs sólo se considera que la presencia del alelo paternalmente heredado en el plasma materno es un falso positivo, mientras que en el primer subconjunto cualquier alelo distinto del alelo común compartido por el padre y la madre se considera como un resultado falso-positivo.

La FIG. 30 muestra la cobertura de los SNPs fetales específicos para diferentes números de células de flujo analizadas. Los SNPs que tanto el padre como la madre eran homocigóticos, pero con alelos diferentes, se incluyen en este análisis. El eje X es la cobertura en veces de los SNPs específicos del feto, y el eje Y es el porcentaje de SNPs con la cobertura plieguen en veces especificada. Con el aumento en el número de células de flujo que se analizan, el número medio de cobertura en veces para el SNP fetal específico aumenta. Por ejemplo, cuando se analizó una célula de flujo, la cobertura media de SNPs fue de 0,23 veces. La cobertura media aumentó a 4,52 veces cuando se analizaron 20 células de flujo.

5. Exactitud de un mapa genético heredado de su madre

La FIG. 31 muestra la precisión del análisis de Tipo A cuando se usaron datos de 10 células de flujo. La sección II.B describe realizaciones de un análisis de Tipo A y Tipo B (también denominado alfa y beta). La precisión es para la determinación correcta del haplotipo que fue heredado de la madre. La precisión se presenta por separado para cada cromosoma.

Utilizando una relación de probabilidad de 1.200 para el análisis SPRT (Zhou W et al. Nat Biotechnol 2001; 19: 78-81; Karoui NE et al. Statist Med 2006; 25: 3124-33), la precisión oscilaba entre 96% y 100%. Como se muestra, incluso con una alta proporción de probabilidad para la clasificación SPRT, un total de 2.760 segmentos a través del genoma pudieron ser clasificados. Este grado de resolución es suficiente para la mayoría de los propósitos, cuando

se considera que las recombinaciones meióticas tienen lugar a la frecuencia de uno a un número de un solo dígito por brazo cromosómico por generación. Además, se pudo ver que se podían evitar todas las clasificaciones erróneas cuando se usaba el método de entrelazado (lado derecho de la figura 31). Como se ha descrito anteriormente, el enfoque de entrelazado utiliza tanto el análisis de Tipo A como el de Tipo B.

La FIG. 32 muestra la precisión del análisis de tipo B cuando se usaban datos de 10 células de flujo. Utilizando una razón de probabilidad de 1.200 para el análisis SPRT, la precisión variaba de 94,1% a 100%. Todas las clasificaciones erróneas pudieron ser evitadas cuando se usaba el método de entrelazado (lado derecho de la figura 32), como se observa en la FIG. 31.

La FIG. 33 muestra la precisión del análisis de Tipo A cuando se usaban los datos de 20 células de flujo. Utilizando una razón de probabilidad de 1.200 para el análisis SPRT y el algoritmo de "dos bloques consecutivos", se realizaron un total de 3.780 clasificaciones y sólo 3 (0,1%) clasificaciones fueron incorrectas. La FIG. 34 muestra la precisión del análisis de Tipo B cuando se usaban los datos de 20 células de flujo. Utilizando una razón de probabilidad de 1.200 para el análisis SPRT y el algoritmo de "dos bloques consecutivos", se realizaron un total de 3.355 clasificaciones y sólo 6 (0,2%) clasificaciones fueron incorrectas. En estos ejemplos, el SPRT se realiza a través de una serie de marcadores genéticos, tales como SNPs.

II. DETERMINACIÓN PRENATAL DEL RIESGO DE BETA-TALASEMIA

En una realización, para determinar el riesgo de que el feto tenga beta-talasemia (una enfermedad autosómica recesiva) se puede determinar si el feto ha heredado alelos mutantes llevados por su padre y su madre. En este caso mencionado anteriormente, el padre es un portador de la delección de CTTT de 4 pares de bases de los codones 41/42 del gen de la beta-globina humana. La madre embarazada era portadora de la mutación A → G en el nucleótido -28 del gen de la beta-globina humana.

Para determinar si el feto ha heredado la mutación de los codones paternos 41/42, se buscaron los datos de secuenciación del ADN plasmático materno, utilizando las primeras 10 células de flujo, para esta mutación. Se encontraron un total de 10 lecturas con esta mutación (Figura 35A). Por lo tanto, el feto había heredado la mutación paterna. Además, se encontró que 62 lecturas contenían la secuencia de tipo salvaje en los codones 41/42 (Figura 35B). Por lo tanto, el porcentaje de las lecturas en esta región que contiene la mutación es 0,1398. Esta cifra está muy próxima a la concentración fraccionada de ADN fetal determinada en la FIG. 26. En una realización, el riesgo de que el feto herede la mutación paterna también se puede determinar elucidando su herencia de polimorfismos genéticos ligados a la mutación paterna.

En una realización, para determinar el riesgo de que el feto haya heredado la mutación materna-28, se realizó un análisis RHDO. En esta familia, la mutación -28 se localizó en el haplotipo IV, mientras que el alelo de tipo salvaje se localizó en el haplotipo III. Los resultados del análisis RHDO de tipo A se muestran en la FIG. 36 mientras que los del análisis RHDO de tipo B se muestran en la FIG. 37. En ambos tipos de análisis, se dedujo la herencia fetal del haplotipo III de la madre. En otras palabras, el feto había heredado el alelo de tipo salvaje de la madre. El diagnóstico final del feto fue que ha heredado la mutación de los codones 41/42 del padre y un alelo de tipo salvaje de la madre. Por lo tanto, el feto es un portador heterocigótico de beta-talasemia y por lo tanto debe ser clínicamente saludable.

III. ENRIQUECIMIENTO DE LA DIANA Y SECUENCIACIÓN DIRIGIDA

Como se discutió en las secciones anteriores, la exactitud de la estimación de la concentración de ADN fetal fraccional y la resolución del mapa genético deducido del análisis del ADN plasmático materno puede depender de la profundidad de cobertura de los loci de interés. Por ejemplo, los inventores han demostrado que un total de 200 moléculas correspondientes a un locus SNP podrían ser necesarias para determinar, con alta precisión, la concentración de ADN fetal fraccional sin información previa del genotipo materno. Los recuentos de alelos para un SNP en el plasma materno se pueden obtener, por ejemplo, pero no limitándose, con PCR en tiempo real, PCR digital y secuenciación masiva paralela.

A medida que la secuenciación masivamente paralela del ADN plasmático materno puede determinar simultáneamente el número de alelos para millones de SNPs en todo el genoma, es una plataforma ideal para el análisis del genoma a través de diferentes loci. El formato básico de la secuenciación masivamente paralela permite que diferentes regiones dentro del genoma se cubran a profundidades similares. Sin embargo, con el fin de secuenciar una región de interés particular a alta profundidad de secuenciación utilizando secuenciación aleatoriamente masiva paralela, las partes restantes del genoma (que no se pretenden analizar) tienen que ser secuenciadas en la misma medida. Por lo tanto, este enfoque podría ser costoso. Para mejorar la rentabilidad del método de secuenciación masivamente paralelo, una forma es enriquecer la región diana antes de proceder a la secuenciación. La secuenciación dirigida puede realizarse mediante la captura en fase de solución (Gnirke A, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009; 27: 182-9), captura con microarrays (por ejemplo, usando la plataforma NimbleGen) o por amplificación

dirigida (Tewhey R, et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 2009; 27: 1025-31).

La secuenciación dirigida se aplicó inicialmente para detectar variaciones genéticas de población, p. ej., para estudios de asociación genética. Por lo tanto, su aplicación actual en la investigación genómica está dirigida a resolver problemas cualitativos (por ejemplo, genotipado o detección de mutaciones). Sin embargo, la aplicación de la secuenciación dirigida en el ADN plasmático materno para fines de diagnóstico prenatal no invasivo implica consideraciones cuantitativas, cuya viabilidad no ha sido clara. Por ejemplo, el uso de secuenciación dirigida podría introducir un sesgo cuantitativo en la detección de ADN fetal y materno en el plasma materno. Además, trabajos anteriores han demostrado que el ADN fetal es más corto que el ADN materno (Chan KCA y col. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 2004; 50: 88-92). Esta diferencia de tamaño también podría introducir un sesgo cuantitativo o una eficiencia diferencial en la captura del ADN fetal y materno en el plasma materno. Tampoco se estaba seguro de la eficiencia con la que tales moléculas de ADN fragmentadas podrían ser capturadas. En las siguientes descripciones, se demuestra que la secuenciación dirigida puede lograrse mediante el enriquecimiento de diana seguido por secuenciación masivamente paralela. También se muestra que el enriquecimiento de la diana es una manera eficaz de estimar la concentración de ADN fetal fraccional en comparación con la secuenciación del genoma completo.

A. Determinación de la concentración fraccional utilizando el enriquecimiento de la diana

1. Materiales y métodos

Cuatro (M6011, M6028, M6029 y M6043) mujeres embarazadas con fetos únicos femeninos fueron reclutadas. Las muestras de sangre periférica materna se recogieron en tubos de sangre de EDTA antes de una sección por cesárea electiva en el tercer trimestre, mientras que las muestras de placenta se recogieron después de la cesárea electiva. Después de la centrifugación, el ADN de las células sanguíneas periféricas se extrajo utilizando el Mini Kit de Sangre (Qiagen). El ADN de 2,4 ml de plasma se extrajo mediante el Mini Kit de Sangre DNA DSP (Qiagen). Se extrajo el ADN genómico materno de la capa leucocitaria y se extrajo ADN genómico fetal de los tejidos placentarios. Las muestras del tercer trimestre se usaron en este ejemplo con fines ilustrativos solamente. Igualmente se pueden usar muestras del primer y segundo trimestre.

Los genotipos maternos y fetales fueron determinados por el Genome-Wide Human SNP Array 6.0 (Affymetrix). Se utilizaron 5-30 ng de ADN de plasma para cada caso para la construcción de la biblioteca de ADN mediante el kit de preparación de muestras de pares (Illumina) de acuerdo con el protocolo del fabricante de la preparación de muestras de secuenciación con inmunoprecipitación de cromatina. El ADN ligado al adaptador se purificó directamente usando columnas de giro proporcionadas en un kit de purificación de PCR QIAquick (Qiagen), sin selección de tamaño adicional. El ADN ligado al adaptador se amplificó después usando una PCR de 15 ciclos con cebadores estándar. Los cebadores fueron PCR Primer PE 1.0 y 2.0 de Illumina. Las bibliotecas de ADN se cuantificaron usando un espectrofotómetro NanoDrop ND-1000 (NanoDrop Technologies) y se ejecutaron en un Bioanalyzer 2100, usando un kit DNA 1000 (Agilent), para verificar la distribución de tamaños. Se generaron 0,6-1 µg de una biblioteca de ADN de plasma amplificado para cada muestra en un tamaño medio de aproximadamente 290 pb. La biblioteca de captura se obtuvo de Agilent y cubrió el 85% de los exones en el chrX humano (número de catálogo: 5190-1993). Para los cuatro casos en este estudio, se incubaron 500 ng de la biblioteca de ADN de plasma amplificado de cada caso con las sondas de captura durante 24 horas a 65 °C, de acuerdo con las instrucciones del fabricante. Después de la hibridación, se seleccionaron las dianas capturadas tirando hacia abajo de los híbridos biotinilados sonda-diana usando esferas magnéticas recubiertas con estreptavidina (Dynal DynaMag-2 Invitrogen) y purificándose con el Kit de Purificación de PCR MinElute (Qiagen). Finalmente, las bibliotecas de ADN diana se enriquecieron mediante amplificación por PCR de 12 ciclos con cebadores SureSelect GA PE de Agilent. Los productos de PCR se purificaron mediante el Kit QIAquick PCR Purification (Qiagen). Las bibliotecas de ADN preparadas con o sin enriquecimiento de la diana se sometieron luego a secuenciación masiva aleatoriamente paralela utilizando el Illumina Genome Analyzer Iix. Se usó un carril de secuenciación en una célula de flujo estándar para secuenciar cada biblioteca de ADN.

2. Concentración fraccional de ADN fetal sin enriquecimiento de la diana

La concentración fraccionada de ADN fetal se puede calcular sobre la base de los recuentos de alelos de los SNP informativos (es decir, SNPs en los que la madre es homocigota y el feto es heterocigoto). La tabla a continuación muestra que 120184, 110730, 107362 y 110321 SNP informativos fueron identificados en todo el genoma de los cuatro casos, mientras que 63, 61, 69 y 65 (respectivamente, en el mismo caso) cayeron dentro de la región diana en el cromosoma X. Sin el enriquecimiento de la diana, las concentraciones fraccionales de ADN fetal fueron 33,4%, 31,3%, 29,2% y 34,4% basadas en los datos de todos los SNP informativos en el genoma.

Muestra	Enriquecimiento objetivo	SNP informativo no. del genoma completo	Recuentos de alelos compartidos	Recuentos de alelos específicos fetales	Concentración fraccional de ADN fetal
M6011	No	120.184	15.309	3.064	33,4%
M6028	No	110.730	16.778	3.114	31,3%
M6029	No	107.362	19.889	3.404	29,2%
M6043	No	110.321	21.070	4.369	34,4%

3. Comparación de muestras con y sin enriquecimiento de la diana

En algunas realizaciones, la profundidad de la cobertura de secuencia representaba el número medio de veces que cada base había sido secuenciada en una región particular. En esta realización, se calculó la profundidad de secuencia de la región diana dividiendo el número total de bases secuenciadas dentro de la región diana por la longitud de región diana (3,05 Mb). Para las regiones cubiertas por el kit de enriquecimiento, la cobertura media de la secuencia fue 0,19 veces para las muestras no enriquecidas y 54,9 veces para las muestras enriquecidas, lo que indica una media de 289 veces el enriquecimiento. A esta profundidad de secuenciación, sólo el 4,0% de los alelos fetales específicos dentro de la región diana se detectaron antes del enriquecimiento diana (véase la tabla a continuación). En comparación, el 95,8% de ellos se volvieron detectables después del enriquecimiento de la diana (véase la tabla a continuación). Por lo tanto, el enriquecimiento de la diana aumentó mucho la tasa de detección de los alelos específicos fetales dentro de la región diana.

A continuación, se comparan las concentraciones fraccionales de ADN fetal basado en los recuentos de lectura de todos los SNP informativos dentro de la región diana para cada muestra, con y sin enriquecimiento. Sin el enriquecimiento diana, el número de lecturas específicas fetales variaba de 0 a 6 para las cuatro muestras (véase la tabla a continuación). Debido a la baja cobertura de la secuencia, el muestreo inadecuado de las moléculas de ADN fetal evitaría una estimación precisa de la concentración fraccional de ADN fetal. Con el enriquecimiento diana, se observó un número mucho mayor de recuentos de alelos específicos fetales (511-776) y de contajes de alelos compartidos (2570 ~ 3922) dentro de la región diana (véase la tabla a continuación). Los porcentajes de ADN fetal fueron calculados como 35,4%, 33,2%, 26,1% y 33,0%, consistente con los porcentajes de ADN fetal estimados por los datos del genoma en las muestras no enriquecidas (véase la tabla más abajo). Estos resultados indicaron que las moléculas de ADN materno y fetal se enriquecieron en un grado similar dentro de la región diana.

Muestra	Enriquecimiento de la diana	Número de SNP informativos dentro de la región seleccionada	Número de alelos específicos fetales detectables	Tasa de detección del alelo fetal específico	Recuentos de alelos compartidos	Recuentos de alelos específicos fetales	Concentración fraccional de ADN fetal
M6011	No	63	6	9,5%	13	6	63,2%
M6028	No	61	2	3,3%	6	2	50,0%
M6029	No	69	2	2,9%	11	2	30,8%
M6043	No	65	0	0,0%	15	0	0,0%
M6011	Sí	63	60	95,2%	3072	661	35,4%
M6028	Sí	61	60	98,4%	2570	511	33,2%
M6029	Sí	69	66	95,7%	3835	575	26,1%
M6034	Sí	65	61	93,9%	3922	776	33,0%

B. Determinación del genoma fetal utilizando el enriquecimiento de la diana

Una aplicación de un método RHDO es para la detección prenatal no invasiva de enfermedades genéticas heredadas de la madre. Utilizando secuenciación masivamente paralela de plasma materno sin enriquecimiento de la diana, el análisis RHDO puede determinar con precisión qué haplotipo materno se transmite al feto con un promedio de 17 SNPs cuando la profundidad de secuenciación del ADN plasmático materno es aproximadamente de 65 veces la cobertura del genoma humano. Para mejorar la rentabilidad de este enfoque, puede dirigirse selectivamente la secuenciación a regiones específicas de interés dentro del genoma y, a continuación, aplicar un análisis RHDO a los datos de secuenciación. Como ejemplo, se ha demostrado el concepto utilizando la secuenciación dirigida y el análisis RHDO del cromosoma X. Sin embargo, la secuenciación dirigida y el análisis RHDO también se pueden aplicar a todos los cromosomas, p. ej. los autosomas. En una realización, puede utilizarse un análisis RHDO como se ha descrito anteriormente para las realizaciones diana.

Se reclutaron a cinco (PW226, PW263, PW316, PW370 y PW421) mujeres embarazadas con fetos varones únicos. Se recogieron muestras de sangre periférica materna en tubos de sangre de EDTA antes del muestreo de las vellosidades coriónicas en el primer trimestre. Después de la centrifugación, el ADN de las células sanguíneas periféricas se extrajo utilizando el Mini Kit de Sangre (Qiagen). El ADN de 3,2 ml de plasma se extrajo mediante el

Mini Kit de sangre DSP DNA (Qiagen). Se extrajo el ADN genómico materno de la capa leucocitaria y se extrajo el ADN genómico fetal de las vellosidades coriónicas. Las muestras se prepararon y analizaron como se ha descrito anteriormente. Cada muestra se secuenció entonces al azar usando un carril en una celda de flujo de Illumina.

En este ejemplo, se utilizó el genotipo fetal, junto con la información de secuenciación de ácidos nucleicos de la madre, para deducir los haplotipos maternos para el cromosoma X y deducir qué haplotipo se heredó de la madre. Para cada SNP en el cromosoma X que la madre era heterocigota (es decir, un SNP informativo), el alelo que fue heredado por el feto se define como procedente del haplotipo materno 1 (Hap I) mientras que el alelo materno que no fue pasado al feto se definió como procedente del haplotipo materno 2 (Hap II). En algunas realizaciones, para aplicaciones clínicas, el genotipo fetal puede no estar disponible de antemano y los haplotipos maternos pueden determinarse o inferirse por métodos bien conocidos por los expertos en la técnica y los métodos descritos en la presente memoria. El cromosoma X se utiliza en este documento sólo con fines ilustrativos. Otros cromosomas, p. ej. los autosomas, también se pueden utilizar en dicho análisis.

Para los cinco casos descritos en este documento, todos ellos llevaban un feto único varón. Como el feto varón sólo hereda un cromosoma X de la madre, pero ningún cromosoma X del padre, el cromosoma X materno que fue pasado al feto estaría sobre-representado en el plasma materno. El análisis de RHDO se llevó a cabo del pter a qter del cromosoma X. A partir del SNP más cercano al pter del cromosoma X, el análisis SPRT puede determinar si el alelo de Hap I o Hap II estaba significativamente sobrerrepresentado estadísticamente en el plasma materno. Si ninguno de los dos haplotipos estaba estadísticamente significativamente sobre representado, los recuentos alélicos para el siguiente SNP se podían combinar para un análisis SPRT más. Los SNPs adicionales se podían combinar para el análisis hasta que el proceso SPRT identificaba uno de los haplotipos como estadísticamente y significativamente sobre-representados. El proceso de clasificación puede reiniciarse en el siguiente SNP.

Las Figs. 38A y 38B muestran los resultados de la clasificación SPRT para el caso PW226 como ejemplo. Hubo un total de nueve exitosas clasificaciones SPRT para el cromosoma X en este caso. Para cada clasificación SPRT, se mostró que los alelos de Hap I estaban sobrerrepresentados en la muestra de plasma materno, lo que indicaba que el feto había heredado Hap I de la madre. Como se ha definido que Hap I es el haplotipo que contiene los alelos pasados en el feto, los resultados de todas estas clasificaciones SPRT fueron correctas.

Los resultados del análisis RHDO para los cinco casos se resumen en la FIG. 39. El número de exitosas clasificaciones de SPRT osciló entre 1 y 9. Todas las clasificaciones de SPRT eran correctas. Una mayor fracción de concentración de ADN fetal se asoció con un mayor número de clasificaciones. Esto se debe a que el desequilibrio alélico debido a la presencia de ADN fetal se puede detectar más fácilmente cuando la concentración fraccional de ADN fetal es mayor. Por lo tanto, pueden ser necesarios menos SNPs para alcanzar una clasificación RHDO exitosa. Las regiones cromosómicas definidas se pueden dividir así en más bloques RHDO. Los resultados de los inventores confirman que el análisis RHDO puede realizarse sobre los datos de secuenciación masiva que se obtienen después del enriquecimiento de la diana.

Los datos de los inventores demuestran además que el enfoque dirigido es una forma más rentable de realizar el análisis RHDO. Sin el enriquecimiento de la diana, para muestras con concentraciones similares de ADN fetal, se requirió la secuenciación de aproximadamente 5 células de flujo (es decir, 40 vías de secuenciación) (Figura 40) para alcanzar la profundidad media obtenida para las muestras mostradas en la Fig. 39. En este documento se muestra que, con el enriquecimiento de la diana, la secuenciación por un solo carril ya alcanza la profundidad de secuenciación media de unas 15 a 19 veces para la clasificación RHDO exitosa. Alternativamente, se podría conseguir un nivel de veces aún mayor de la cobertura de secuenciación con relativamente poco coste adicional cuando se usa el enriquecimiento de la diana. El nivel más alto de cobertura de secuenciación puede reducir eficazmente el tamaño de la región genómica requerida para la clasificación exitosa de RHDO y por lo tanto mejorar la resolución del análisis.

IV. ENRIQUECIMIENTO DE LA DIANA

Desde 2004 se sabe que las moléculas de ADN fetal circulantes son generalmente más cortas que el ADN materno en el plasma materno (Chan KCA et al Clin Chem 2004; 50: 88-92; Li et al Clin Chem 2004). Sin embargo, la base molecular de esta observación sigue sin resolverse. En el estudio actual de los inventores, se han generado 3,931 x 10⁹ lecturas en la muestra de plasma del estudio y se han utilizado recipientes de 1-bp en el análisis de bioinformática. El tamaño de cada molécula de ADN de plasma secuenciada se dedujo a partir de las coordenadas del genoma de los extremos de las lecturas de pares de extremo.

Para este análisis, se centró en los polimorfismos de un solo nucleótido (SNPs) en los que el padre y la madre eran homocigotos, pero para un alelo diferente. Para tales SNPs, el feto era un heterocigoto obligado. El alelo para cada SNP que el feto había heredado del padre podría ser utilizado como marcador fetal-específico. Se determinaron los tamaños de las secuencias fetales (usando los alelos específicos fetales paternalmente heredados) y totales para el genoma completo (Figura 41) e individualmente para cada cromosoma (Figura 42A-42C).

Se observó que las diferencias más significativas entre el ADN fetal y materno en el plasma materno es la reducción del pico de 166 pb, en relación con el pico de 143 pb (figura 41). Las secuencias totales más abundantes (predominantemente maternas) fueron de 166 pb de longitud. La diferencia más significativa en la distribución de tamaños entre el ADN fetal y total fue que el ADN fetal exhibió una reducción en el pico de 166 pb (Figura 41) y una prominencia relativa del pico de 143 pb. Este último probablemente correspondió al recorte de un fragmento de enlace de ~ 20-bp de un nucleosoma a su partícula núcleo de ~ 146 pb (Lewin B, en Gene IX, Jones y Bartlett, Sudbury, 2008, págs. 757-795).

A partir de aproximadamente 143 pb y por debajo, las distribuciones de ambos ADN fetal y total demostró una periodicidad de 10 pb que recuerda a los nucleosomas nucleados por nucleasa. Estos datos sugieren que los fragmentos de ADN del plasma se derivan del procesamiento enzimático apoptótico. Por el contrario, el análisis de tamaño de las lecturas que se correlaciona con el genoma mitocondrial no ligado a las histonas no mostró este patrón nucleosómico (Figura 41). Estos resultados proporcionan una explicación molecular previamente desconocida para las diferencias de tamaño conocidas entre el ADN fetal y materno usando el cromosoma Y y marcadores genéticos polimórficos seleccionados (Chan KCA et al Clin Chem 2004; 50: 88-92; Li y col Clin Chem 2004; 50: 1002-1011; Solicitud de Patente de EE.UU. 20050164241; Solicitud de Patente de EE.UU. 20070202525), y muestran que existen diferencias de tamaño en todo el genoma. La explicación más probable de esta diferencia es que las moléculas de ADN fetal circulantes consisten en más moléculas en las que el fragmento de enlace de ~ 20 pb ha sido recortado de un nucleosoma.

Teniendo en cuenta estas observaciones, hay una serie de formas en que la muestra puede ser enriquecida para el ADN fetal. En una realización, se pueden usar reactivos que se unirían preferentemente al fragmento enlazador. Se espera que tales reactivos se unan preferentemente al ADN derivado de la madre cuando se compara con el ADN derivado del feto en el plasma materno. Un ejemplo de tales reactivos es un anticuerpo. Un objetivo de tal anticuerpo es aquel que se une a la histona H1. Se sabe que la histona H1 se une al fragmento enlazador. Una aplicación de tal anticuerpo es para realizar el enriquecimiento del ADN fetal por selección negativa, es decir, a través de la inmunoprecipitación preferencial del ADN derivado de la madre en el plasma materno que contiene el fragmento enlazante, que contiene la histona H1. Además, se sabe que H1 tiene un número de variantes, algunas de las cuales exhiben variación de expresión específica de tejido (Sancho M y otros PLoS Genet 2008; 4: e1000227). Estas variantes podrían explotarse aún más para diferenciar el ADN fetal (predominantemente placentario) y materno (predominantemente hematopoyético) (Lui YYN et al Clin Chem 2002; 48: 421-427). Por ejemplo, se puede dirigir una variante de histona H1 que es predominantemente expresada por células trofoblásticas para seleccionar preferencial y positivamente ADN derivado fetal en plasma materno. Esta estrategia también se puede aplicar para otras proteínas histonas u otras proteínas nucleosómicas que exhiben patrones de expresión específicos de tejido, especialmente trofoblasto.

Dado el fuerte pico de 166 pb para el ADN materno, otra posibilidad de enriquecer el ADN fetal es diseñar un sistema para la selección negativa de fragmentos de ADN que tengan una longitud de 166 ± 2 pb. Por ejemplo, un sistema basado en electroforesis capilar o cromatografía líquida de alta resolución podría permitir la medida precisa del tamaño y la separación de moléculas de ADN. Otra forma de selección negativa es hacer esto in silico durante el análisis bioinformático de los datos de secuenciación.

Como otras especies de ADN en el plasma, p. ej. ADN tumoral (Vlassov VV y col. Curr Mol Med 2010; 10: 142-165) y ADN de órgano trasplantado (Lo YMD et al Lancet 1998; 351: 1329-1330), también se espera que comparta tales características con el ADN fetal en el plasma materno, las estrategias enumeradas en (1) y (2) anterior también podrían ser utilizadas para el enriquecimiento de estas especies de ADN.

De acuerdo con una realización, se proporciona un método para el enriquecimiento diferencial de especies de ADN en plasma o suero humano a través del direccionamiento del fragmento de unión de los nucleosomas. En una realización, el enriquecimiento se realiza mediante la eliminación de uno de los siguientes: ADN derivado de la madre o ADN derivado de células hematopoyéticas. En otra realización, el direccionamiento implica un reactivo (tal como un anticuerpo u otro tipo de proteína) que se uniría preferentemente a una proteína o componente de ácido nucleico del fragmento enlazador del nucleosoma. En otra realización, el reactivo de direccionamiento se unirá selectivamente a la histona H1 u otra proteína que se una al fragmento enlazador del nucleosoma. En otra realización, el reactivo de direccionamiento se unirá a variantes maternas o hematológicas de la histona H1 u otra proteína que se una al fragmento de unión del nucleosoma. En una realización, la eliminación del ADN se lleva a cabo mediante inmunoprecipitación o unión a una superficie sólida.

De acuerdo con otra realización, un método para el enriquecimiento diferencial de ADN fetal en plasma materno o suero incluye: (a) usar un anticuerpo que se uniría a uno o más componentes del fragmento enlazador del nucleosoma; (b) eliminar la fracción unida por inmunoprecipitación o captura a una superficie sólida; y (c) recoger la fracción no unida que contiene una concentración fraccionada incrementada del ADN fetal.

Cualquiera de los componentes o funciones del software descritos en esta solicitud puede ser implementado como código de software para ser ejecutado por un procesador usando cualquier lenguaje informático adecuado tal como, por ejemplo, Java, C++ o Perl usando, por ejemplo, técnicas convencionales u orientadas a objetivos. El código del

software puede almacenarse como una serie de instrucciones o comandos en un medio legible por ordenador para su almacenamiento y/o transmisión, medios adecuados incluyen memoria de acceso aleatorio (RAM), una memoria de sólo lectura (ROM), un medio magnético tal como un disco duro o disquete, o un medio óptico tal como un disco compacto (CD) o DVD (disco versátil digital), memoria flash, y similares. El medio legible por ordenador puede ser cualquier combinación de tales dispositivos de almacenamiento o transmisión.

Tales programas también pueden ser codificados y transmitidos usando señales portadoras adaptadas para la transmisión a través de redes cableadas, ópticas y/o inalámbricas que se ajustan a una variedad de protocolos, incluyendo Internet. Como tal, puede crearse un medio legible por ordenador de acuerdo con una realización de la presente invención usando una señal de datos codificada con dichos programas. Los medios legibles por ordenador codificados con el código de programa pueden ser empaquetados con un dispositivo compatible o proporcionados por separado de otros dispositivos (por ejemplo, a través de descarga de Internet). Cualquiera de tales medios legibles por ordenador puede residir en o dentro de un único producto de programa informático (por ejemplo, un disco duro o un sistema informático completo) y puede estar presente en o dentro de diferentes productos de programas informáticos dentro de un sistema o red. Un sistema informático puede incluir un monitor, impresora u otra pantalla adecuada para proporcionar cualquiera de los resultados mencionados en este documento a un usuario.

Un ejemplo de un sistema informático se muestra en la FIG. 43. Los subsistemas mostrados en la FIG. 43 están interconectados a través de un bus de sistema 4375. Se muestran subsistemas adicionales tales como una impresora 4374, un teclado 4378, un disco fijo 4379, un monitor 4376, que está acoplado al adaptador de pantalla 4382, y otros. Los periféricos y dispositivos de entrada-salida (E-S), que se acoplan al controlador de E-S 4371, pueden conectarse al sistema informático por cualquier número de medios conocidos en la técnica, tales como el puerto serie 4377. Por ejemplo, se puede usar el puerto serie 4377 o la interfaz externa 4381 para conectar el aparato de ordenador a una red de área amplia tal como Internet, un dispositivo de entrada de ratón o un escáner. La interconexión a través del bus de sistema permite que el procesador central 4373 se comuniquen con cada subsistema y controle la ejecución de instrucciones desde la memoria de sistema 4372 o el disco fijo 4379, así como el intercambio de información entre subsistemas. La memoria de sistema 4372 y/o el disco fijo 4379 pueden incorporar un medio legible por ordenador. Cualquiera de los valores mencionados en este documento puede ser enviado de un componente a otro componente y puede ser enviado al usuario.

Un sistema informático puede incluir una pluralidad de los mismos componentes o subsistemas, por ejemplo, conectados entre sí por una interfaz externa 4381 o por una interfaz interna. En algunas realizaciones, los sistemas informáticos, subsistemas o aparatos pueden comunicarse a través de una red. En tales casos, una computadora puede ser considerada un cliente y otra computadora un servidor, donde cada uno puede ser parte de un mismo sistema informático. Un cliente y un servidor pueden incluir múltiples sistemas, subsistemas o componentes.

REIVINDICACIONES

1. Un método para determinar una concentración fraccional de ADN fetal en una muestra biológica tomada de una mujer embarazada con un feto, teniendo el feto un padre y una madre que es la mujer embarazada, teniendo la madre un genoma materno, teniendo el feto un genoma fetal, en donde la muestra biológica contiene una mezcla de ácidos nucleicos maternos y fetales, comprendiendo el método:

analizar una pluralidad de moléculas de ácido nucleico de la muestra biológica, en donde el análisis de una molécula de ácido nucleico incluye:

identificar una localización de la molécula de ácido nucleico en el genoma humano; y
determinar un alelo respectivo de la molécula de ácido nucleico;

determinar uno o más primeros loci, en donde el genoma fetal es heterocigótico en cada primer loci, de manera que el genoma fetal tiene un respectivo primer y segundo alelo en ese primer loci, y en donde el genoma materno es homocigótico en cada primer loci, de manera que el genoma materno tiene dos del respectivo segundo alelo en ese primer loci, siendo el primer alelo distinto del segundo alelo, en donde determinar que un locus específico es uno de los uno o más de los primeros loci incluye:

determinar un valor de corte para un número de recuentos predichos del primer alelo respectivo en el locus específico, prediciendo el valor de corte si el genoma materno es homocigótico y el genoma fetal es heterocigótico, en donde el valor de corte se determina basándose en una distribución estadística de números de recuentos para distintas combinaciones de homocigosidad y heterocigosidad en el locus específico;

basándose en el análisis de la pluralidad de moléculas de ácido nucleico, detectar el primer alelo respectivo y el segundo alelo respectivo en el locus específico;

determinar un número de recuentos reales del primer alelo respectivo basándose en el análisis de la pluralidad de moléculas de ácido nucleico de la muestra biológica; y

determinar que el locus específico es uno de los primeros loci cuando el número de recuentos reales es menor que el valor de corte;

para al menos uno de los primeros loci:

determinar un primer número P de recuentos del primer alelo respectivo y un segundo número Q de recuentos del segundo alelo respectivo; y

calcular la concentración fraccional basándose en el primer y el segundo números.

2. El método de la reivindicación 1, en donde la concentración fraccional se determina como $2xp/(p+q)$.

3. El método de la reivindicación 1 o 2, en donde P y Q se determinan para una pluralidad de primeros loci, y donde

$$f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)}$$

la concentración fraccional f se determina como $\sum_{i=1}^n (p_i + q_i)$, donde p_i es el primer número del i-ésimo primer loci y q_i es el segundo número del i-ésimo primer loci.

4. El método de una cualquiera de las reivindicaciones anteriores, en donde la determinación del valor de corte incluye la determinación de una distribución estadística para una concentración fraccional máxima y mínima.

5. El método de una cualquiera de las reivindicaciones anteriores, en donde el análisis de una molécula de ácido nucleico incluye implementar en al menos una porción de las moléculas de ácido nucleico al menos una técnica seleccionada del grupo que consiste en secuenciación masiva paralela, microarray, hibridación, PCR en tiempo real, PCR digital y espectrometría de masas.

6. El método de una cualquiera de las reivindicaciones anteriores, en donde la distribución estadística comprende una distribución de Poisson.

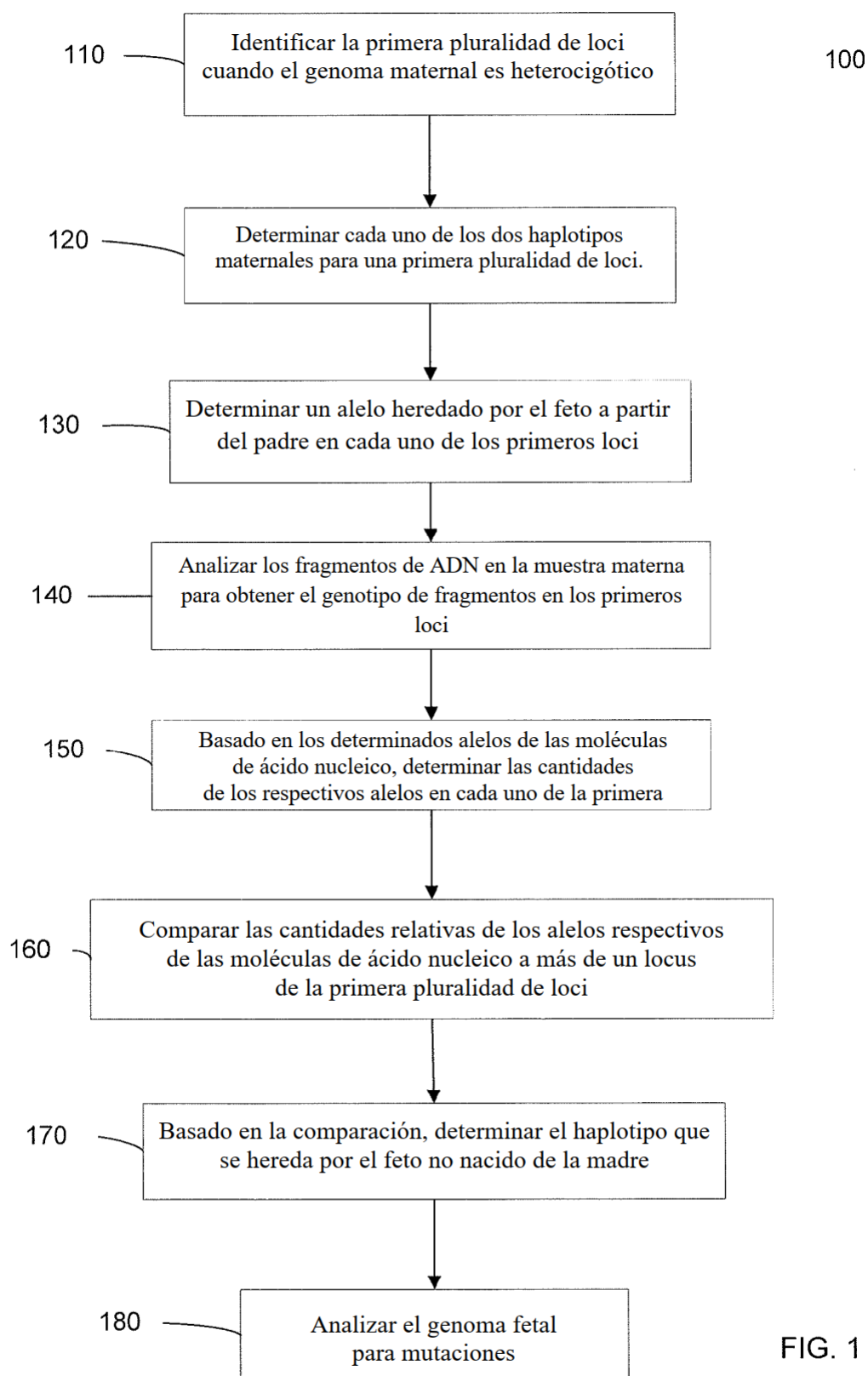
7. El método de la reivindicación 6, que comprende además:

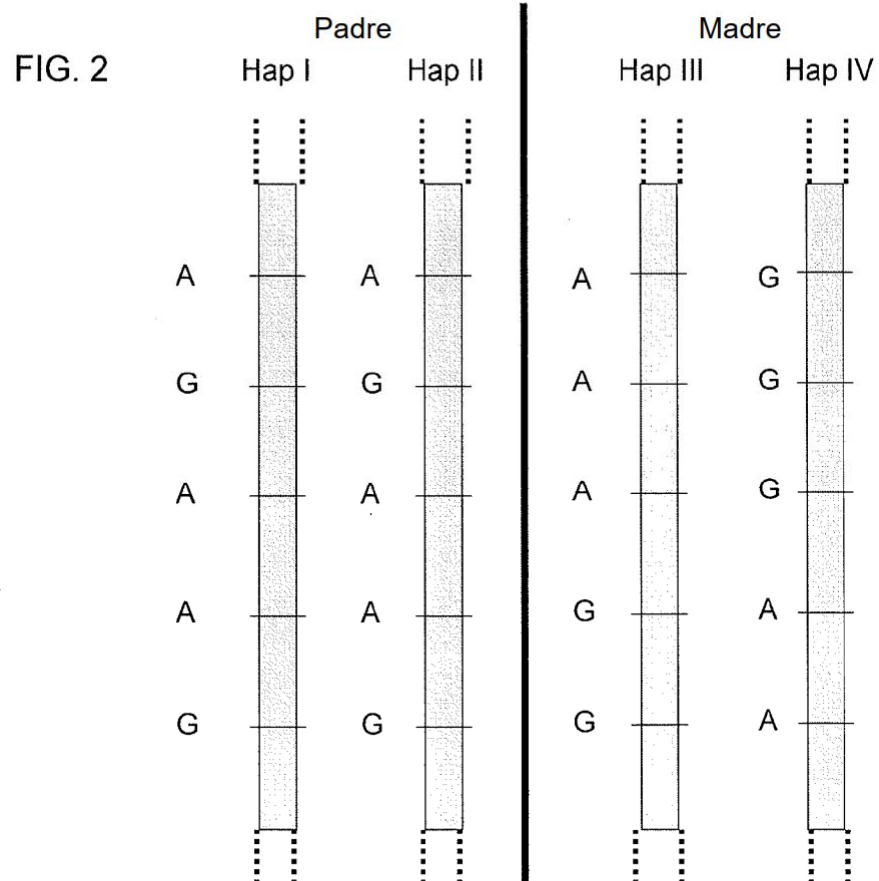
enriquecer la muestra biológica obtenida de la mujer embarazada en cuanto a moléculas de ácido nucleico en una región diana; y

secuenciar la pluralidad de moléculas de ácido nucleico de la muestra biológica enriquecida, siendo la secuenciación específica para la región diana,

en donde la identificación de la localización de la molécula de ácido nucleico en el genoma humano comprende utilizar los resultados de la secuenciación para identificar la localización de la molécula de ácido nucleico en la región diana del genoma humano.

8. El método de la reivindicación 7, en donde el enriquecimiento de la muestra biológica comprende secuenciación dirigida mediante captura en fase de solución, captura con microarrays o amplificación dirigida.
- 5 9. El método de la reivindicación 7 u 8, en donde el padre tiene un genoma paterno, en donde la determinación del uno o más primeros loci incluye:
determinar que el genoma paterno es homocigótico para el respectivo primer alelo en cada loci de los primeros loci y determinar que el genoma materno es homocigótico para el respectivo segundo alelo en el mismo locus.
- 10 10. Un programa informático que comprende un código de software capaz de ser ejecutado por un procesador de un sistema informático, estando el programa informático configurado para realizar, al ejecutarse, un método según cualquiera de las reivindicaciones anteriores.





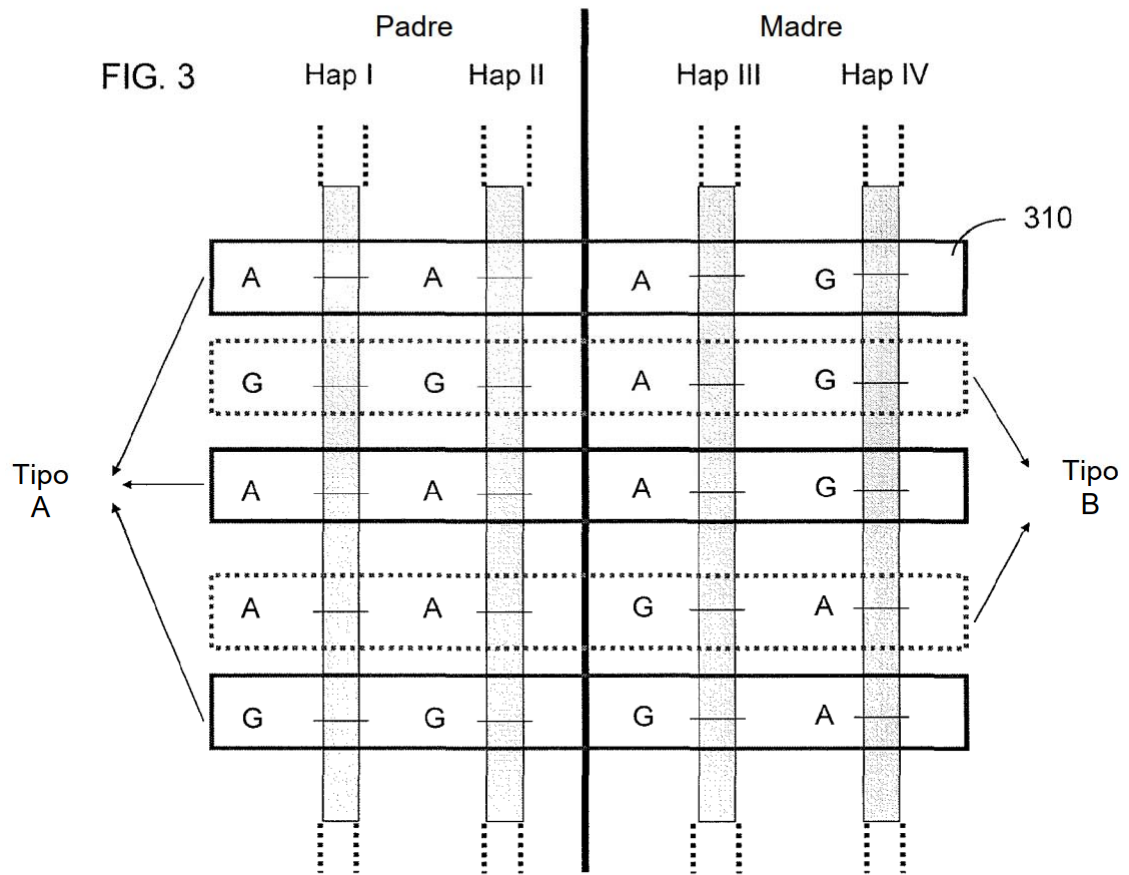


FIG. 4A

Análisis Tipo A

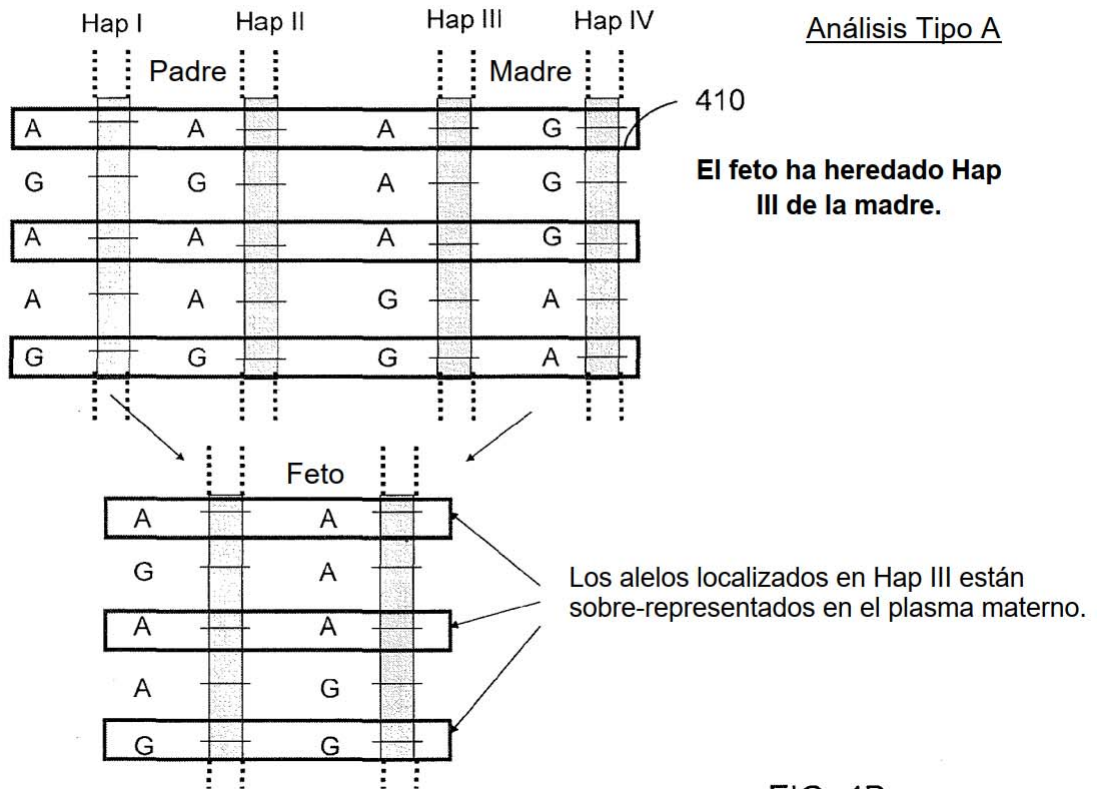


FIG. 4B

Análisis Tipo B

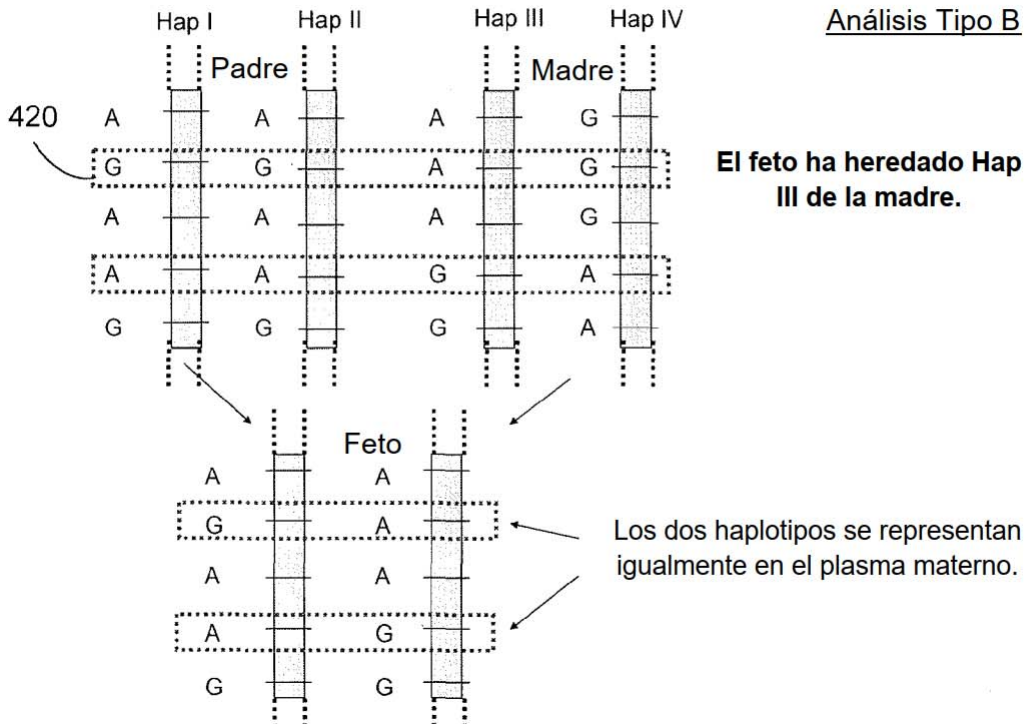


FIG. 5A: Análisis SPRT para RHDO
Análisis tipo A: alelos paternos idénticos a Hap III

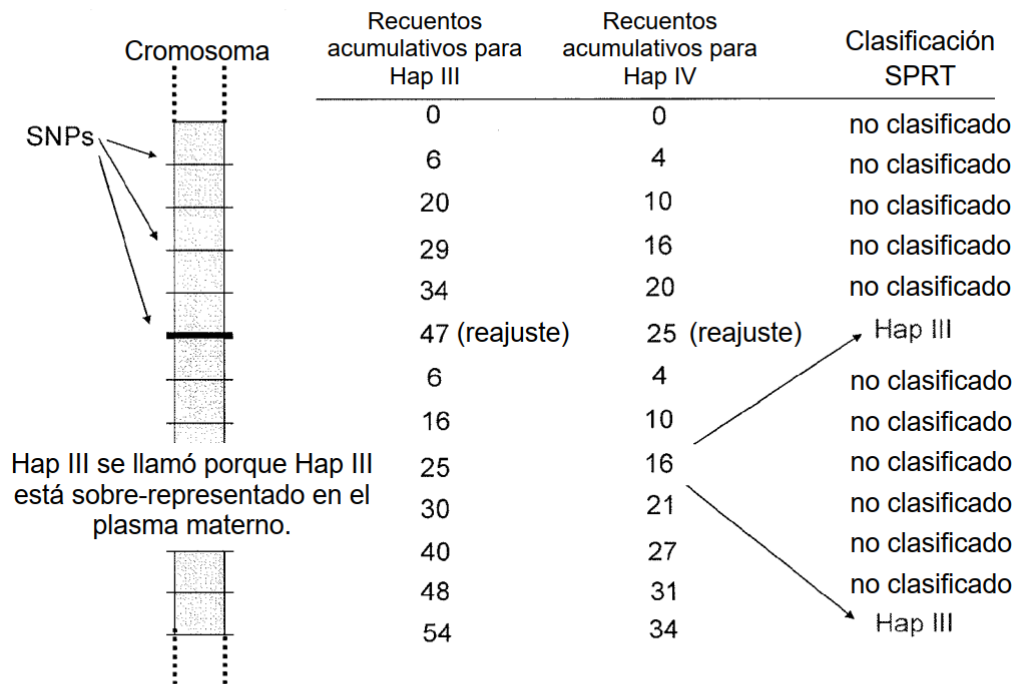


FIG. 5B: Análisis SPRT para RHDO
Análisis tipo B: alelos paternos idénticos a Hap IV

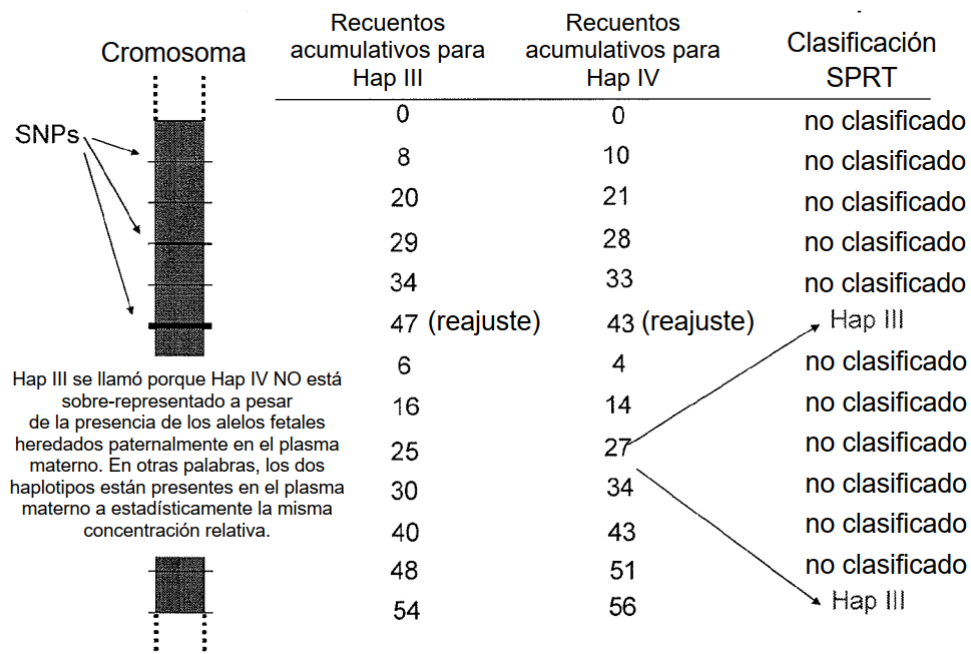
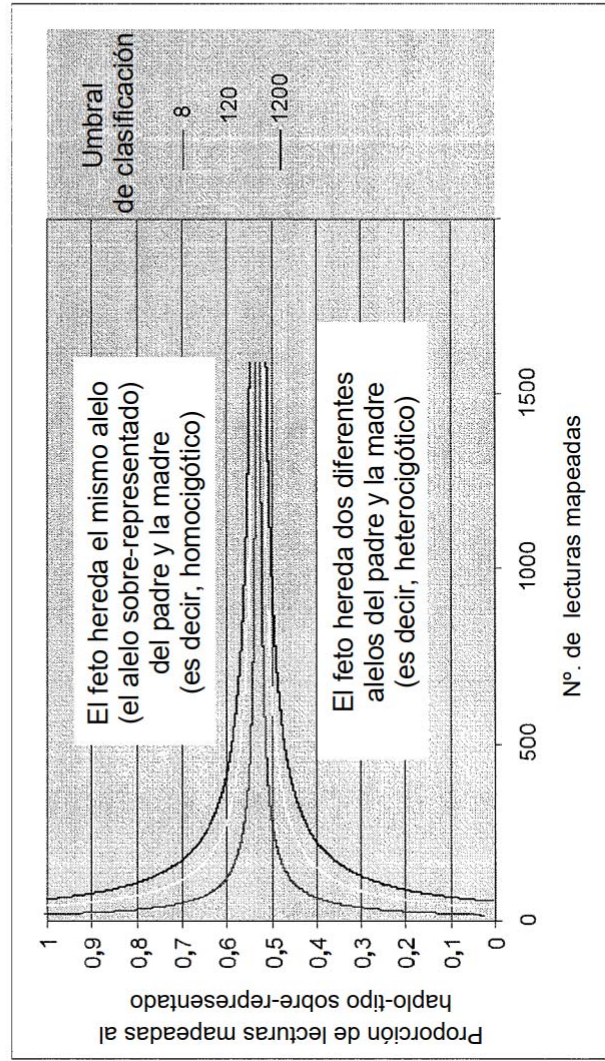


FIG. 6: Efecto de la alteración del umbral de clasificación en las curvas SPRT (alelo paterno idéntico a Hap III)



700

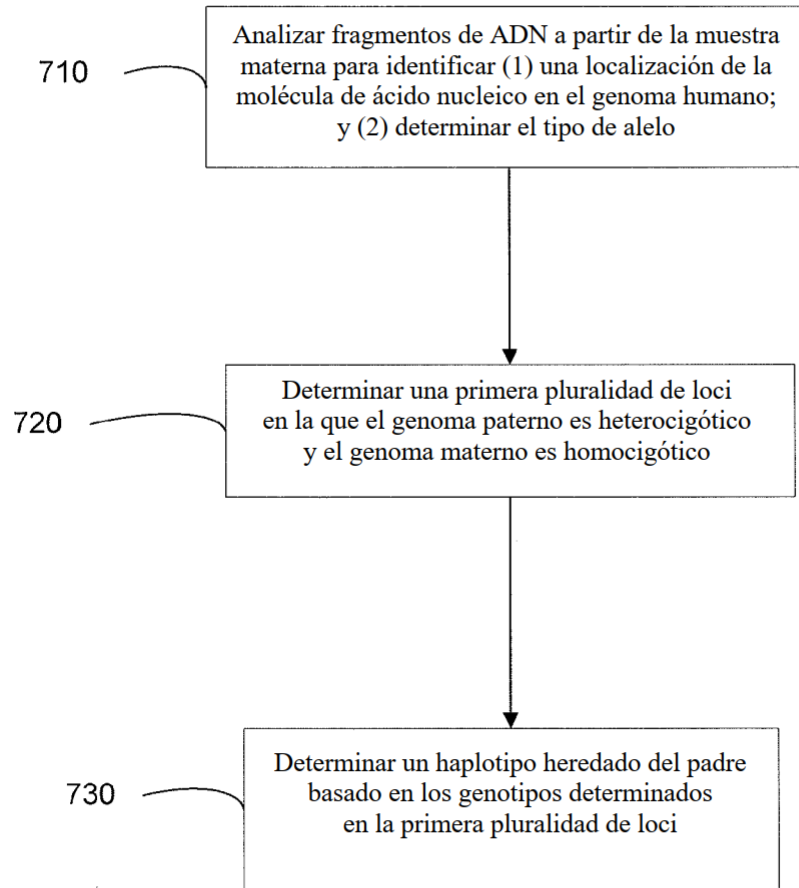


FIG. 7

800

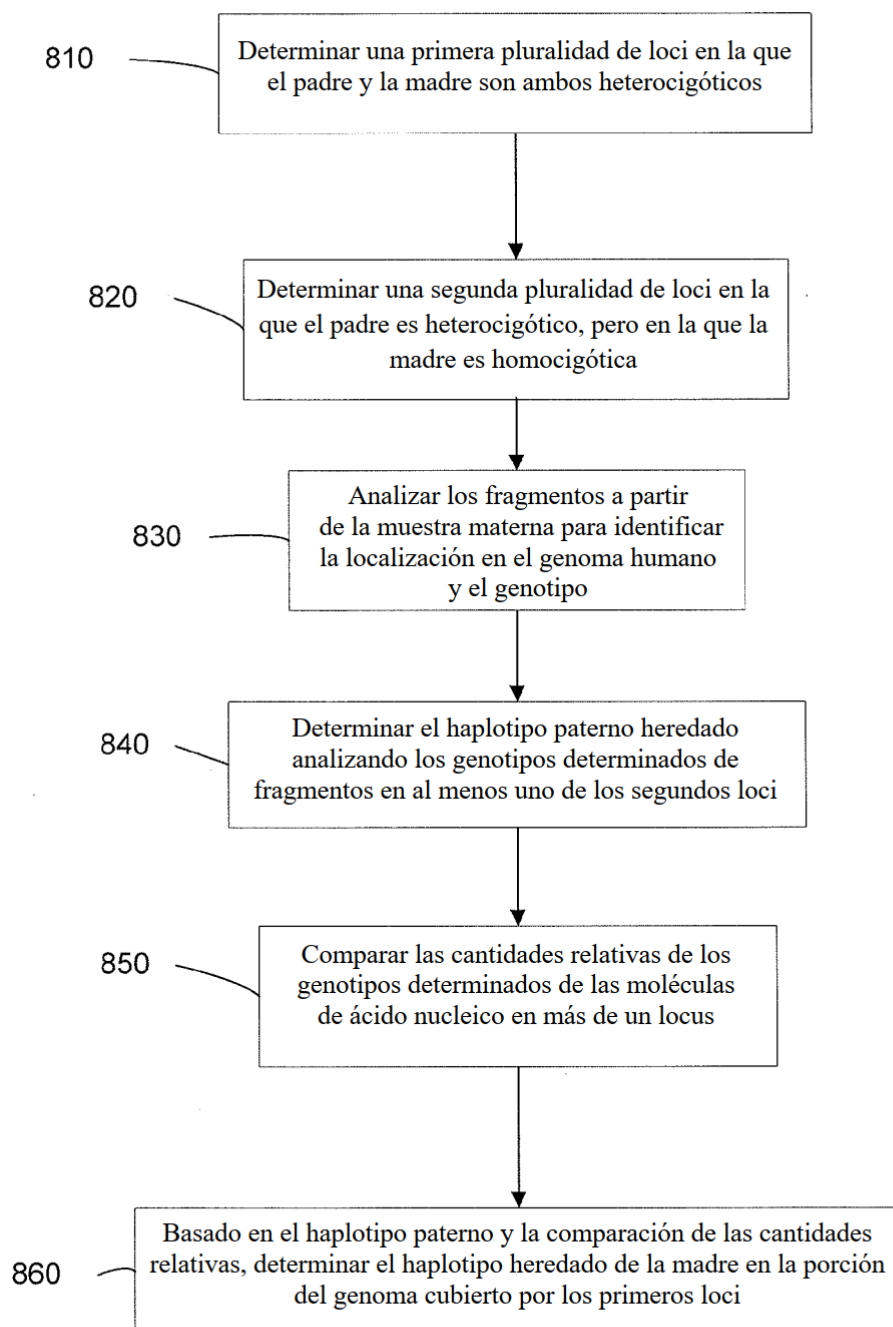
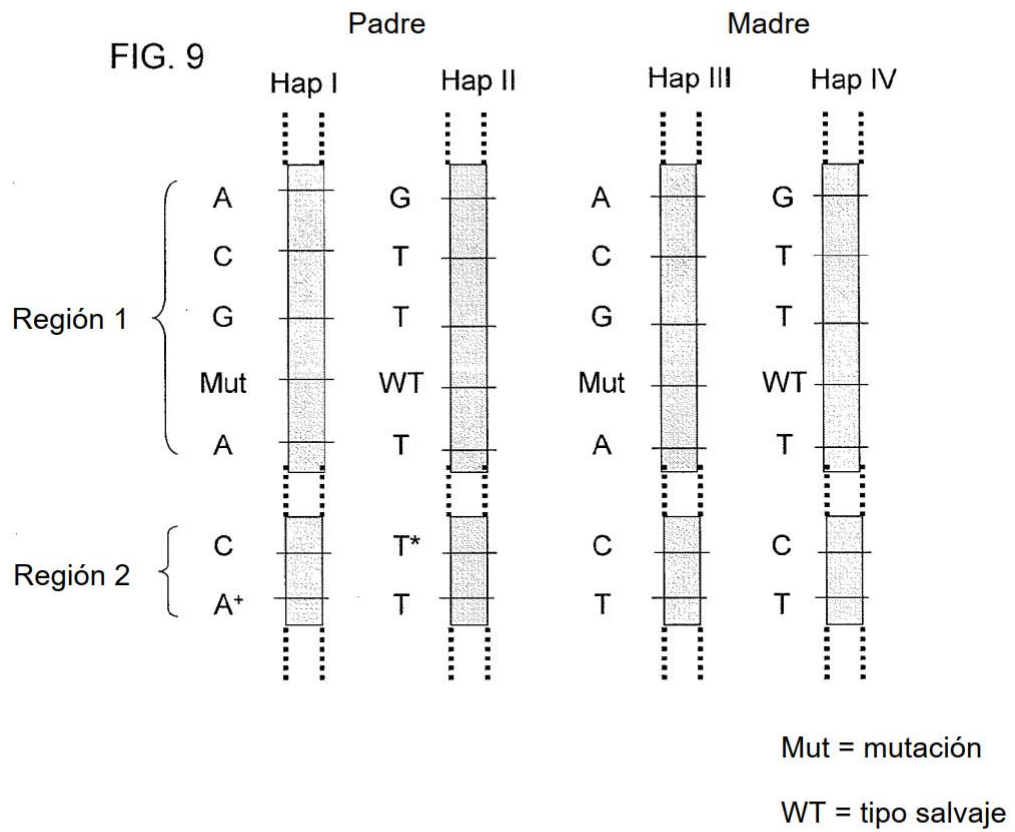


FIG. 8



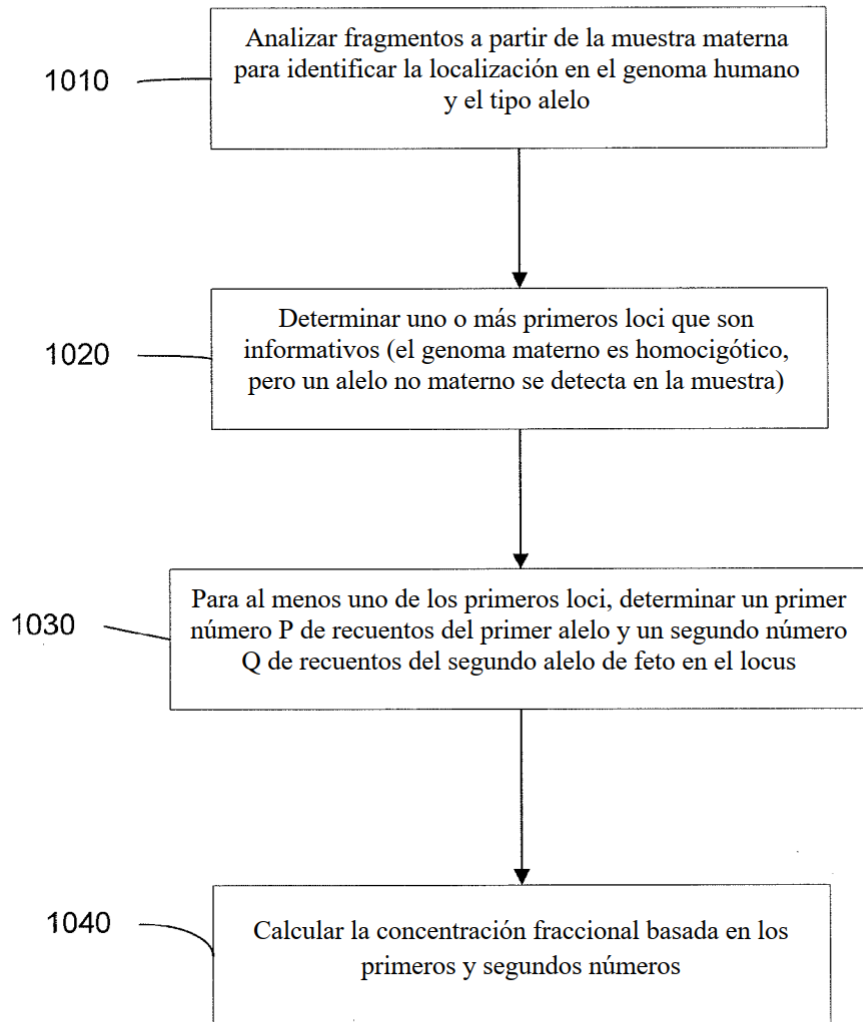


FIG. 10

1100

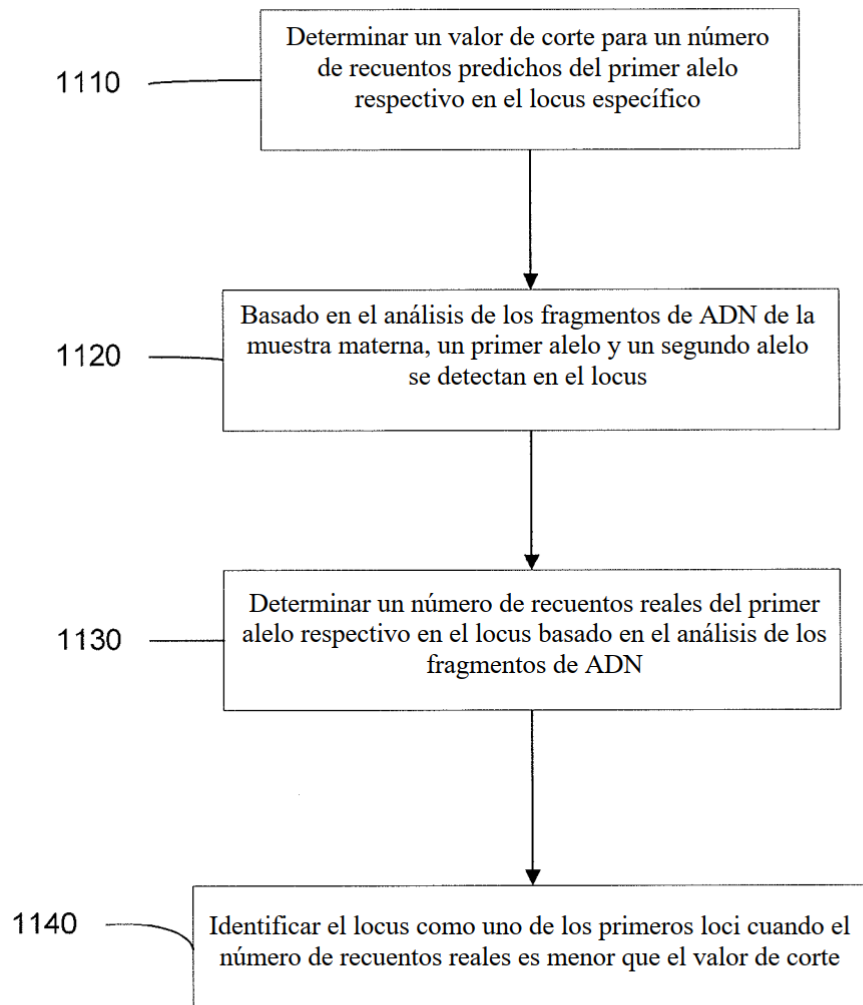


FIG. 11

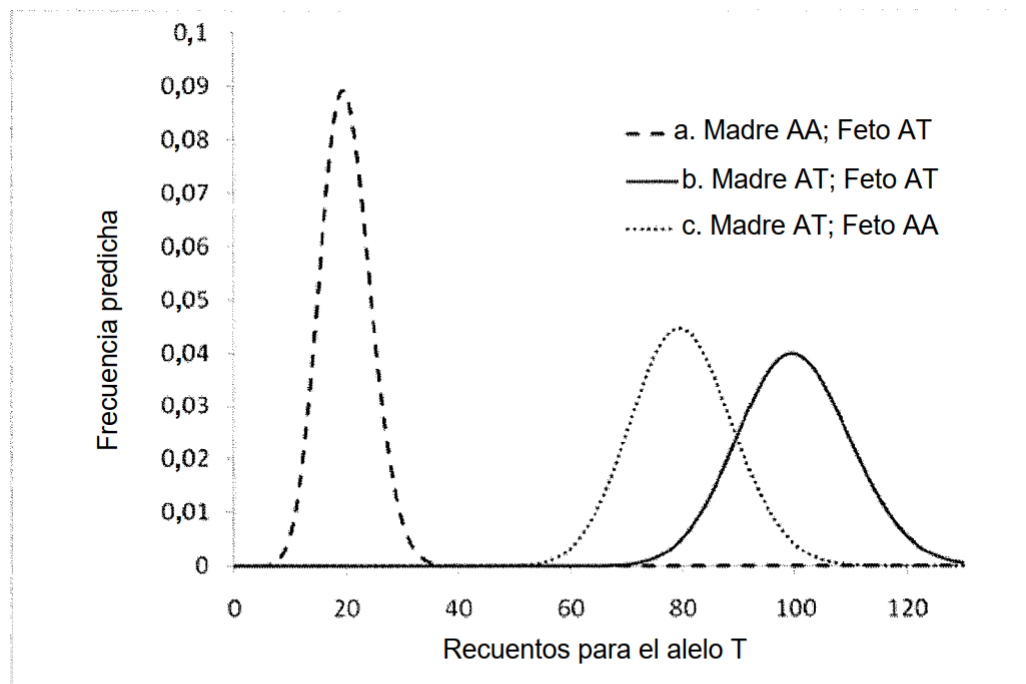


FIG. 12A

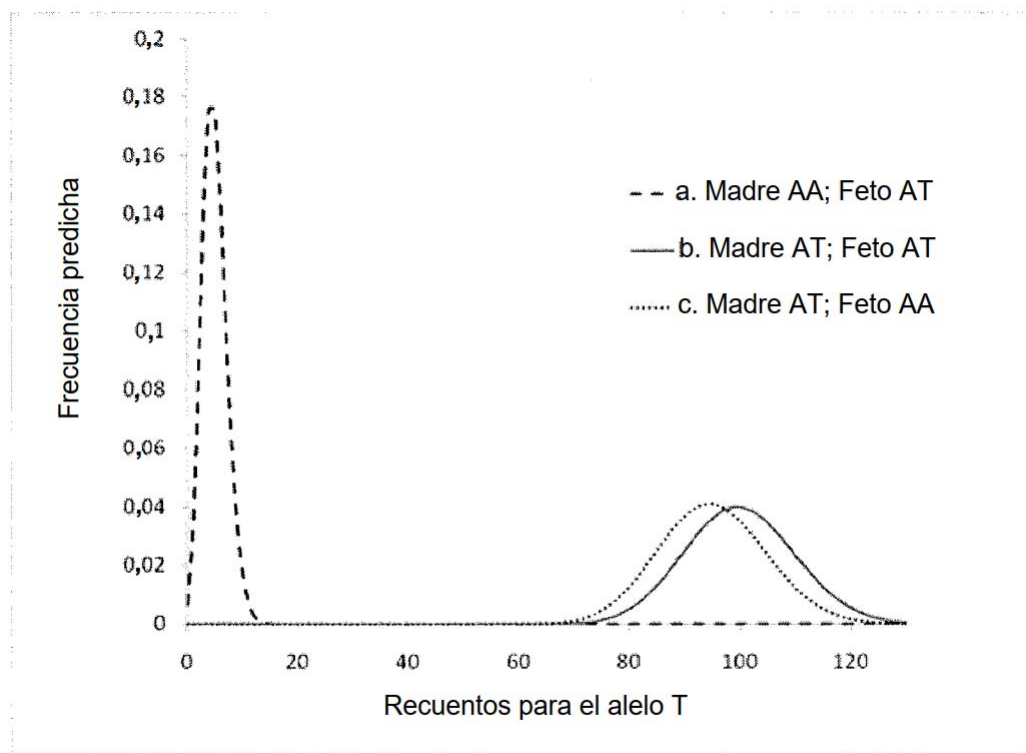


FIG. 12B

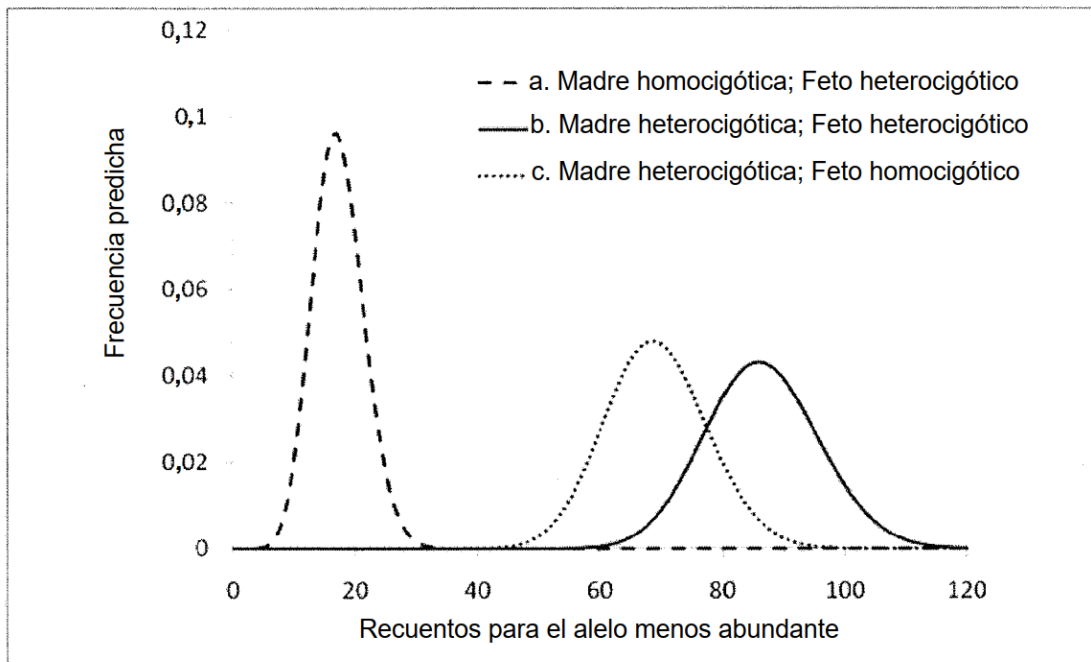


FIG. 13A

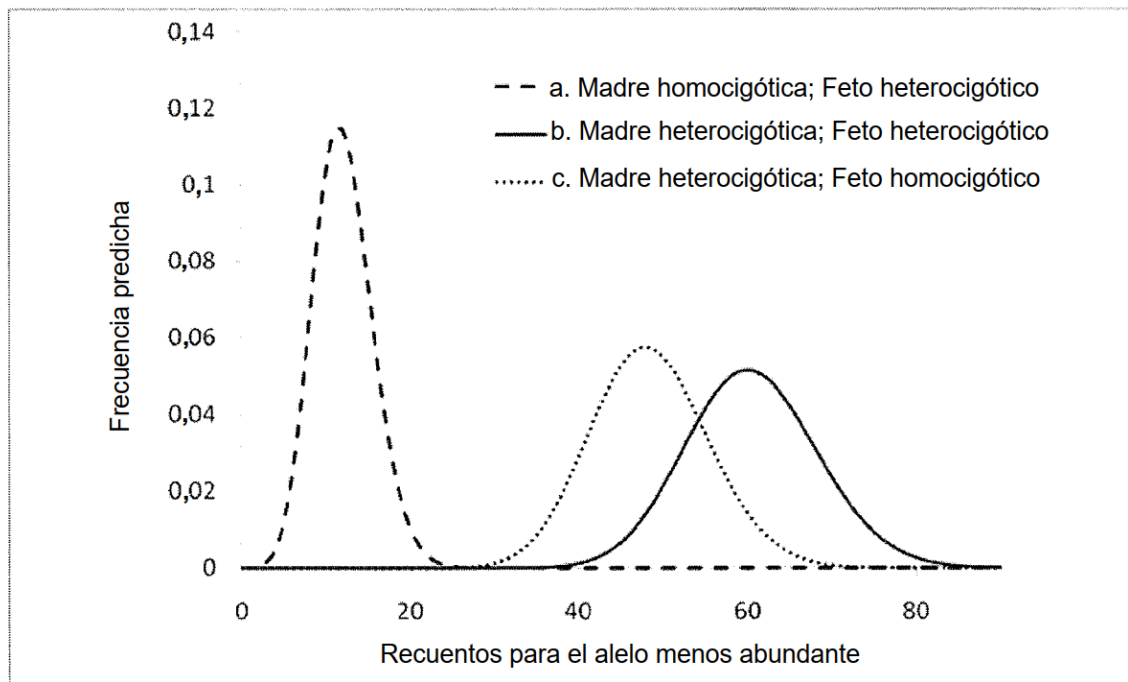


FIG. 13B

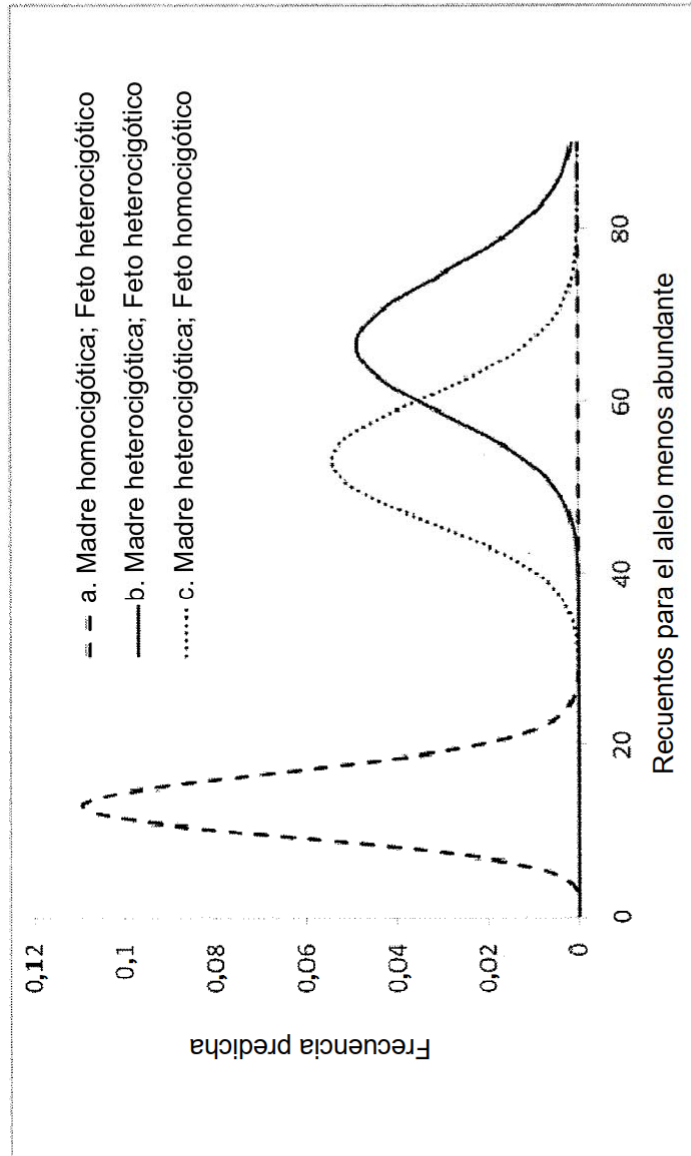


FIG. 14

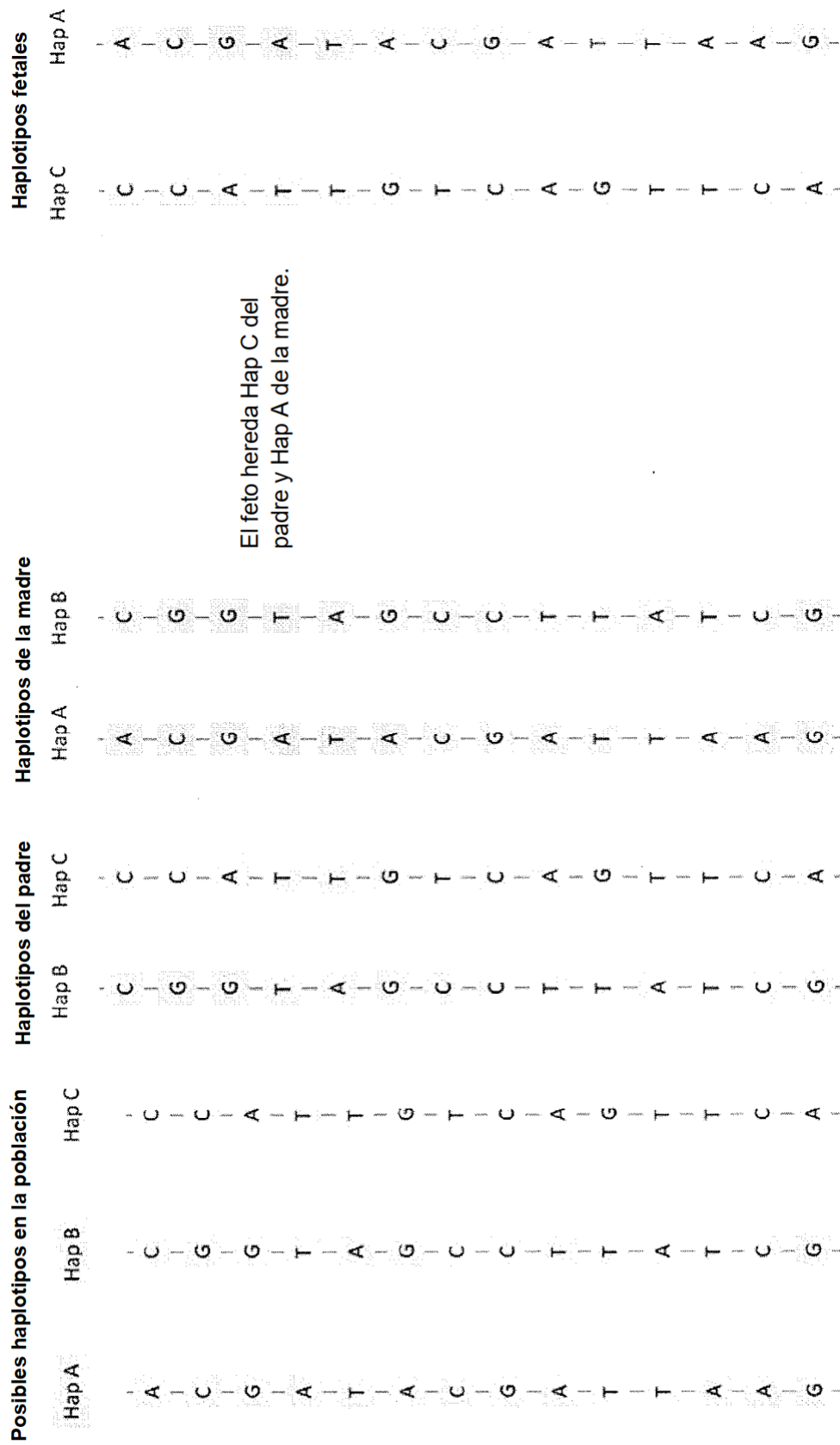


FIG. 15A

FIG. 15B

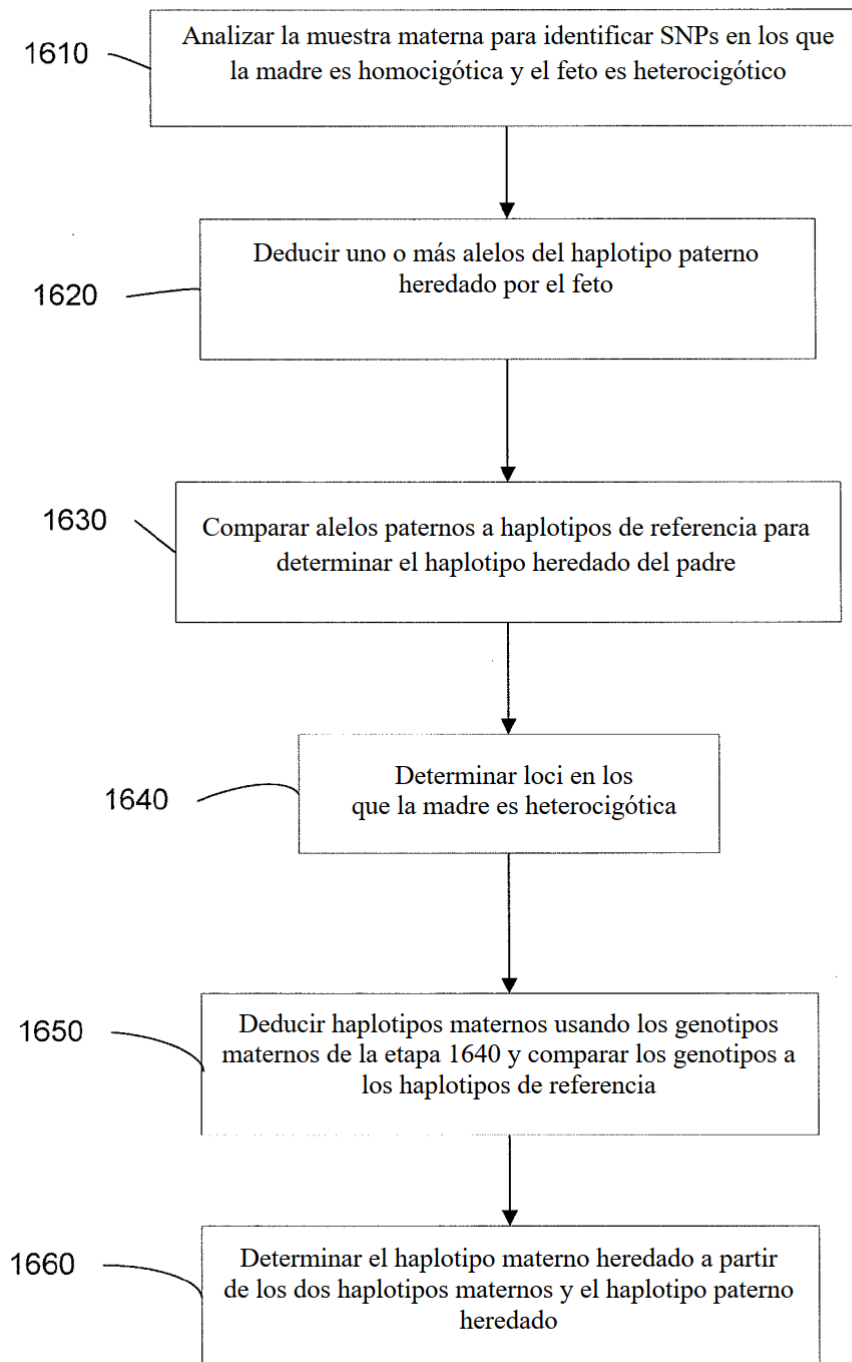
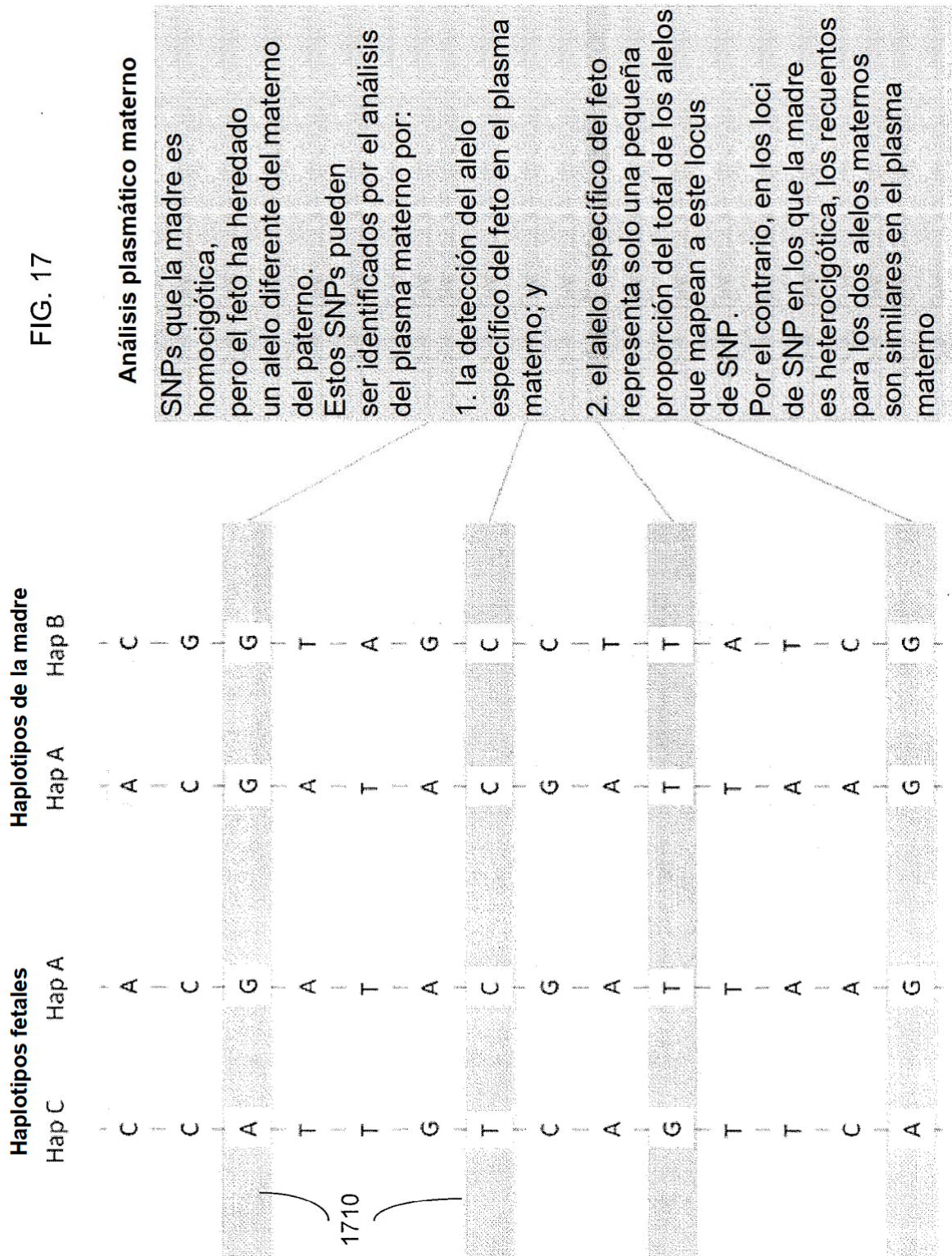


FIG. 16



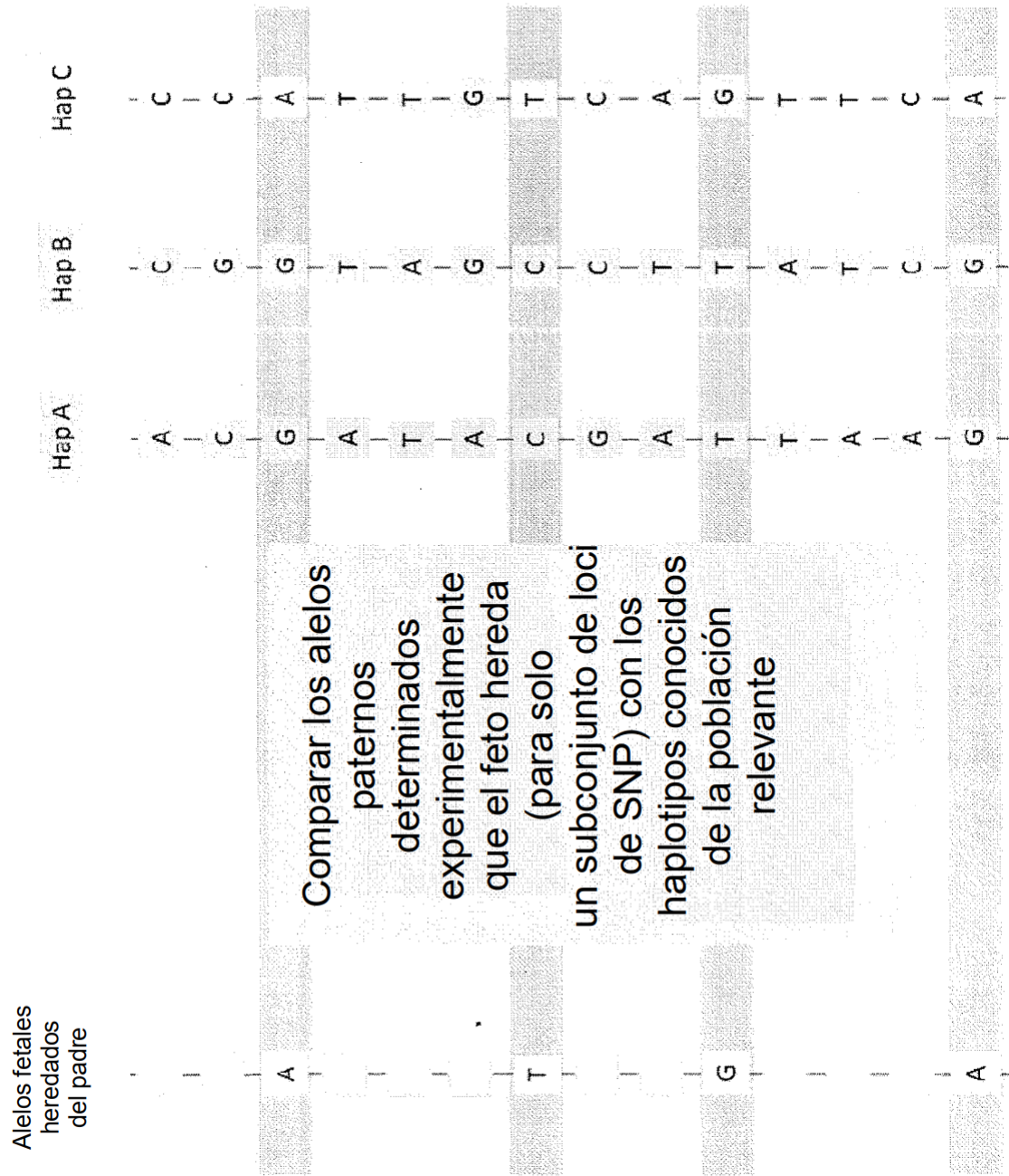


FIG. 18

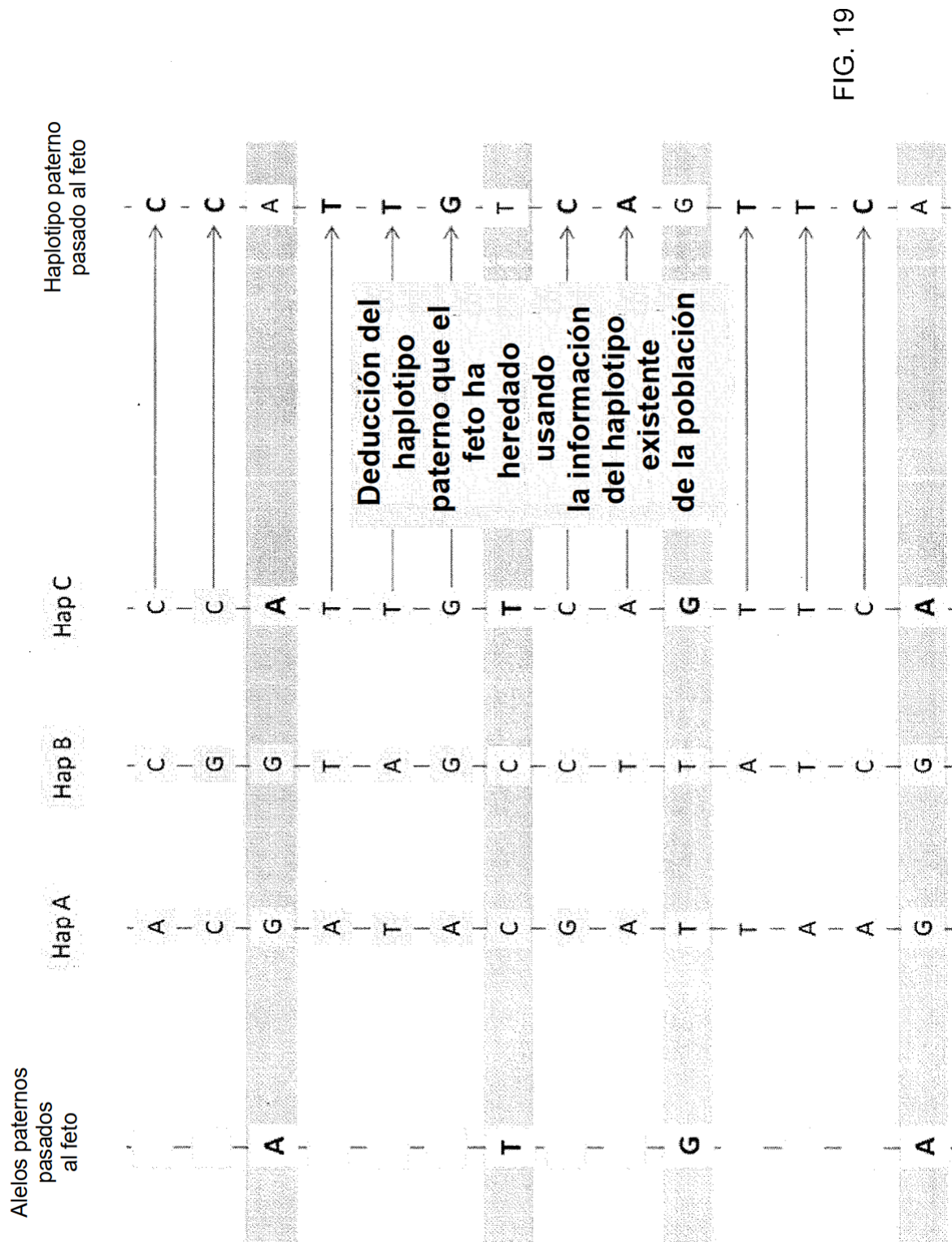
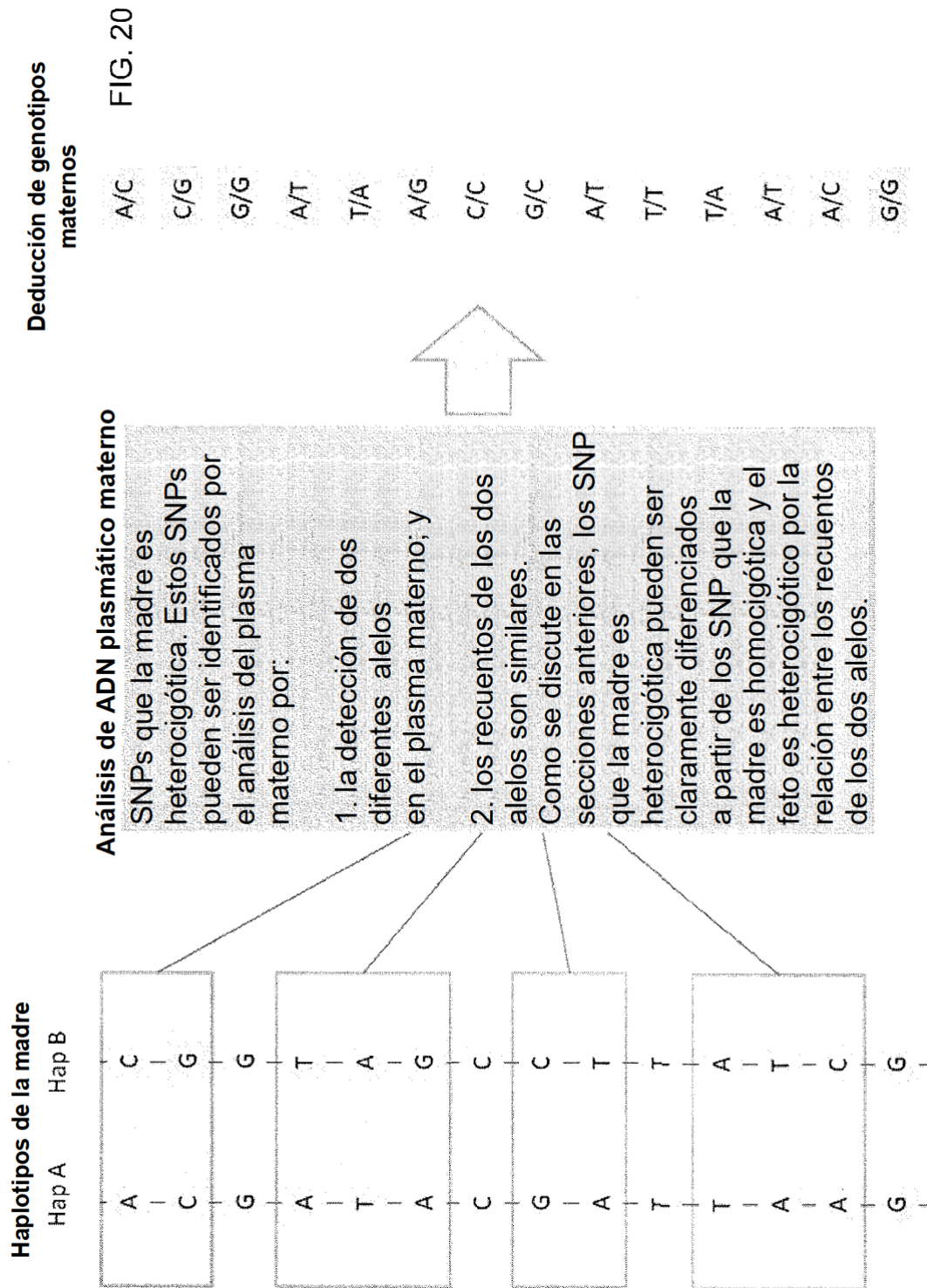


FIG. 19



Deducción de los haplotipos maternos a partir de los genotipos maternos

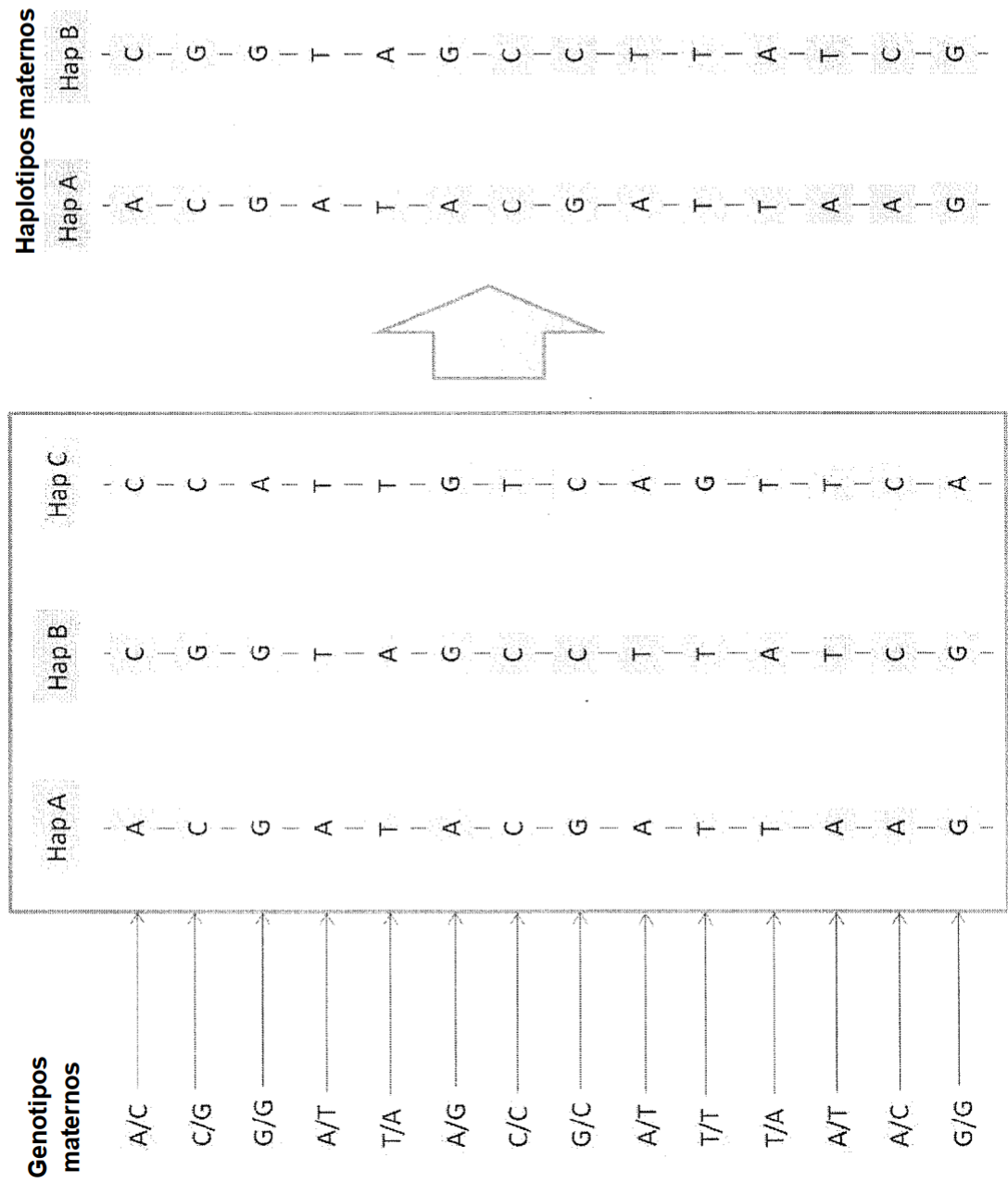


FIG. 21

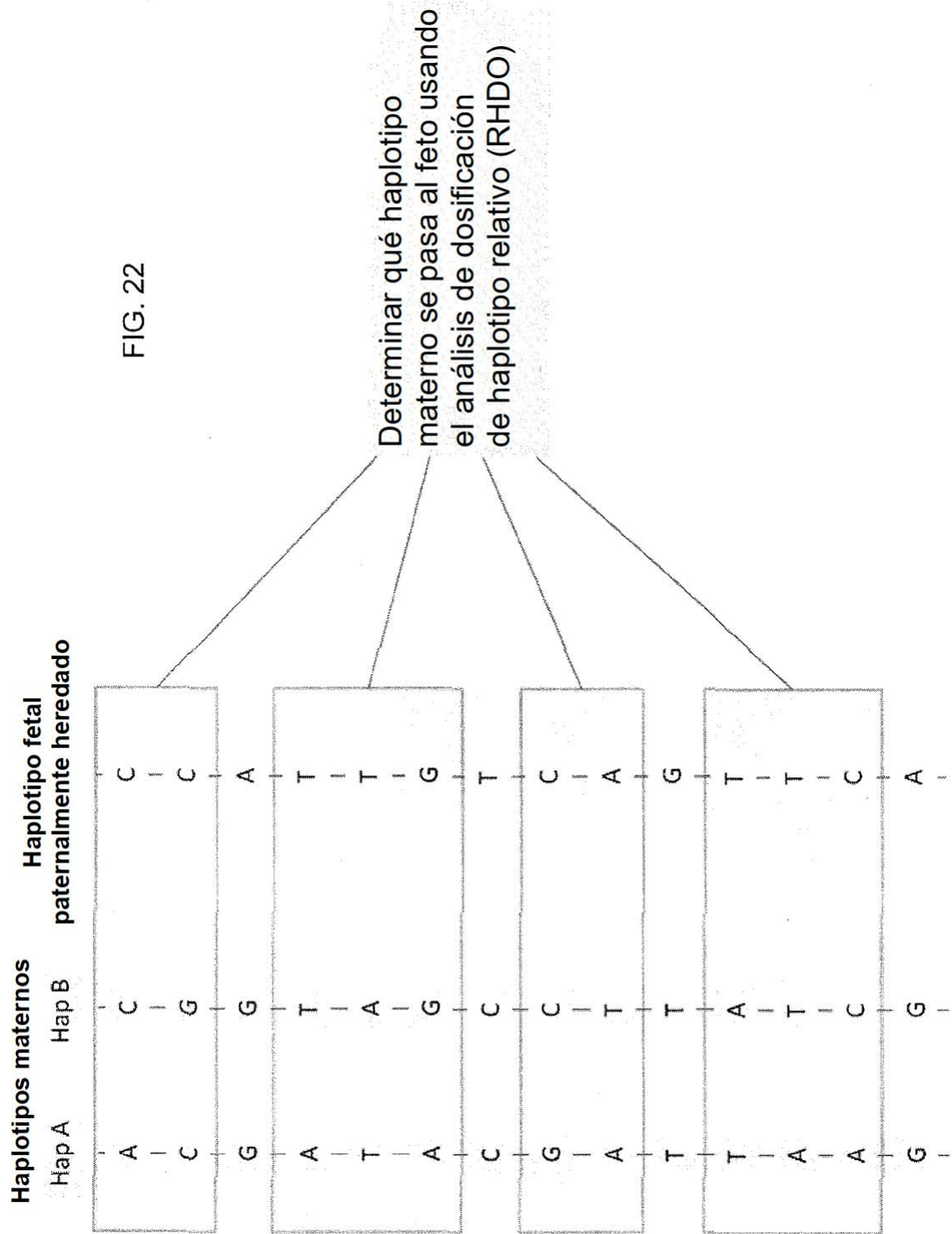
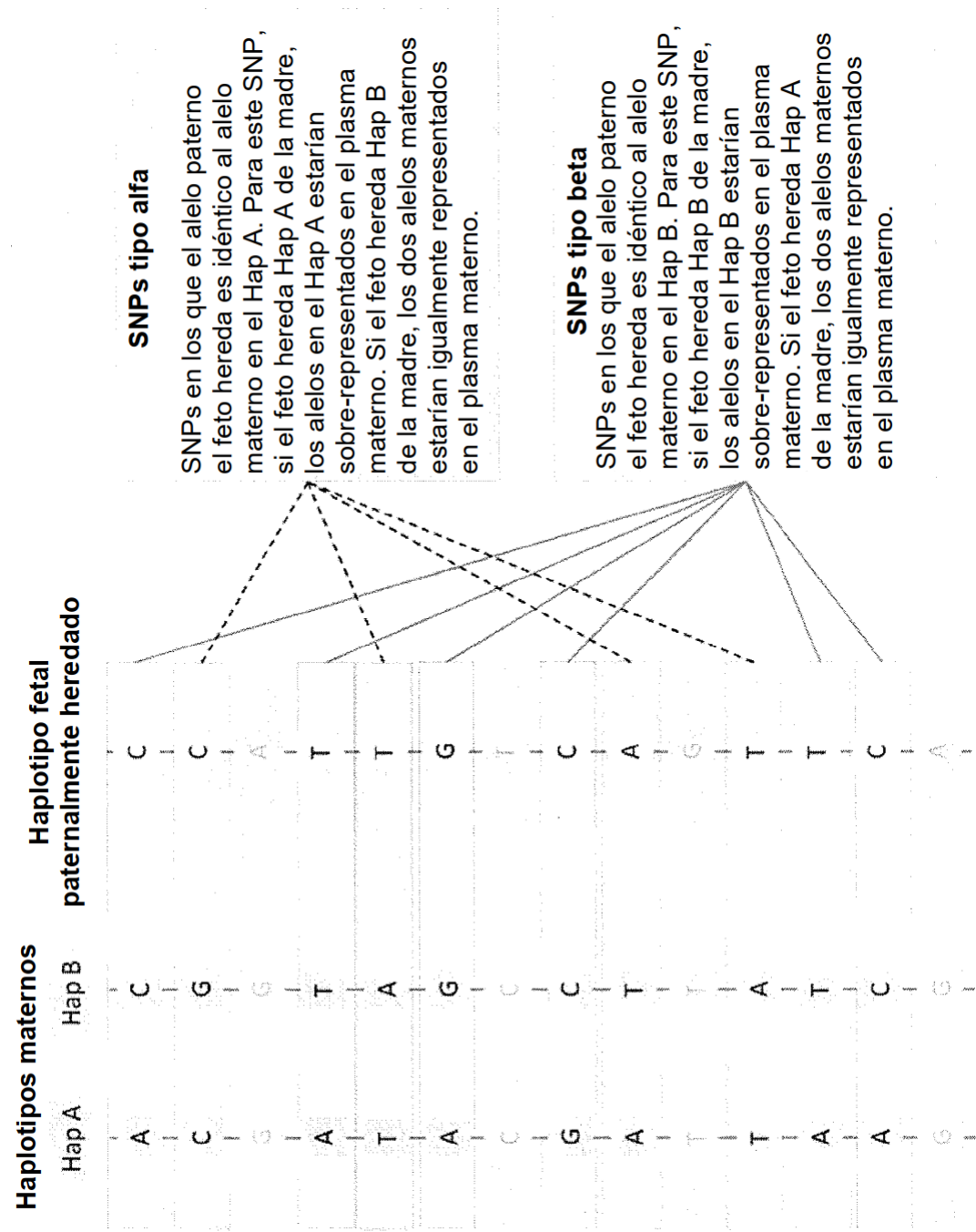


FIG. 23



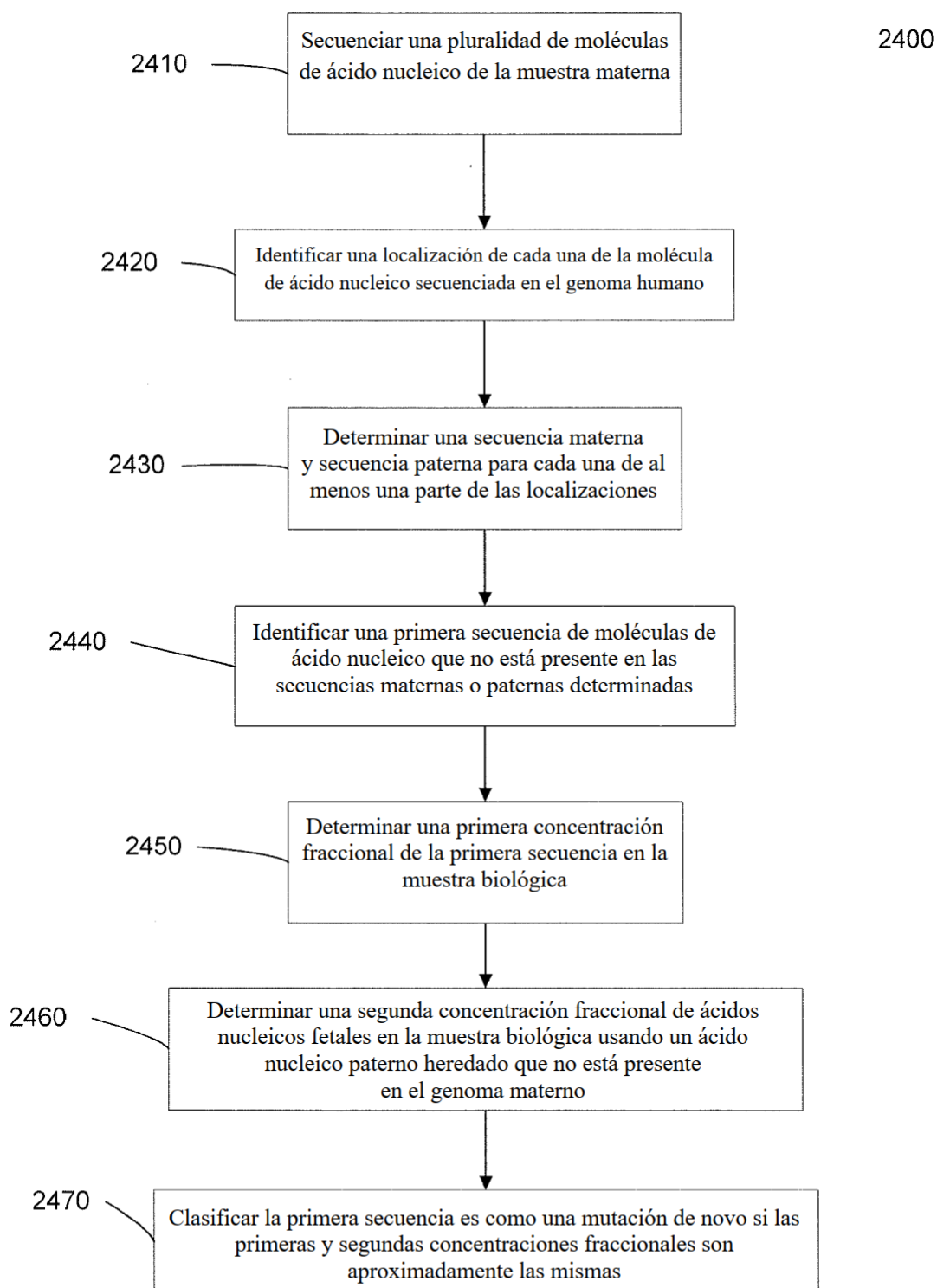


FIG. 24

Combinación de genotipo	No. de SNPs	Porcentaje (%)
Feto heterocigótico	220.296	24,57
Madre homocigótica; padre heterocigótico	131.037	14,61
Madre homocigótica; feto heterocigótico	111.775	12,46
Madre homocigótica; padre heterocigótico; feto heterocigótico	65.875	7,35
Madre homocigótica; padre heterocigótico; feto homocigótico	65.162	7,27
Madre homocigótica; padre heterocigótico; pero con diferentes alelos	45.900	5,12
Total	896.756	100

FIG. 25A

FIG. 25B

Celda de flujo	Lectura en bruto total (millones)	Lectura alineada total (millones)	Tasa de alineamiento, %
1	265	200	75,29
2	274	206	75,22
3	268	204	75,96
4	254	193	76,10
5	258	196	76,02
6	280	215	76,80
7	272	207	75,97
8	253	192	75,80
9	260	195	75,17
10	266	208	78,11
11	245	188	76,68
12	249	190	76,22
13	250	189	75,32
14	257	192	74,79
15	250	187	75,14
16	243	187	76,93
17	270	200	74,25
18	258	197	76,48
19	258	193	74,98
20	253	193	76,31

**FIG. 26. Concentración fetal de ADN
estimada a partir de 20 FCs**

Crom.	Padre hetero, Madre homo, Feto hetero (%)	Padre hetero, Madre homo, difer. alelos (%)
Crom1	11,83	11,92
Crom2	12,15	11,99
Crom3	11,80	11,92
Crom4	11,89	11,84
Crom5	12,01	12,10
Crom6	12,00	11,80
Crom7	12,14	11,75
Crom8	12,04	11,95
Crom9	11,70	11,82
Crom10	11,57	11,75
Crom11	11,88	11,90
Crom12	12,09	11,80
Crom13	12,39	11,94
Crom14	11,90	11,84
Crom15	11,75	11,61
Crom16	11,93	11,36
Crom17	11,77	11,49
Crom18	11,99	11,74
Crom19	11,77	12,02
Crom20	11,74	11,56
Crom21	11,74	11,37
Crom22	11,81	11,34
CromX		12,03

Crom: cromosoma
homo: homocigótico
hetero: heterocigótico

FIG. 27A.

Proporción de SNPs cubierta por lecturas específicas fetales
(madre y padre homocigótico pero con diferentes alelos)

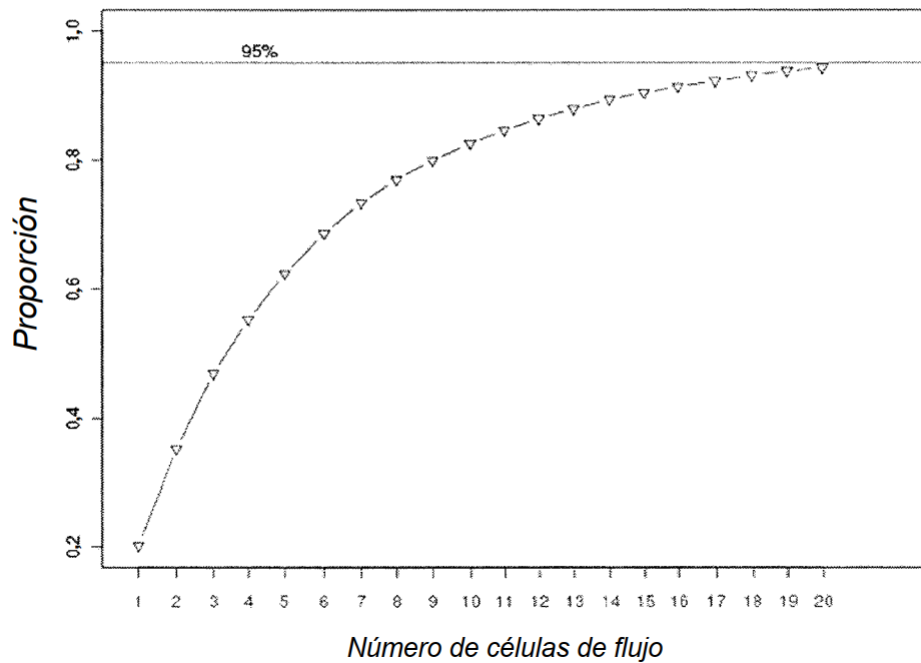


FIG. 27B.

Proporción de SNPs cubierta por lecturas específicas fetales
(madre y padre homocigótico pero con diferentes alelos)

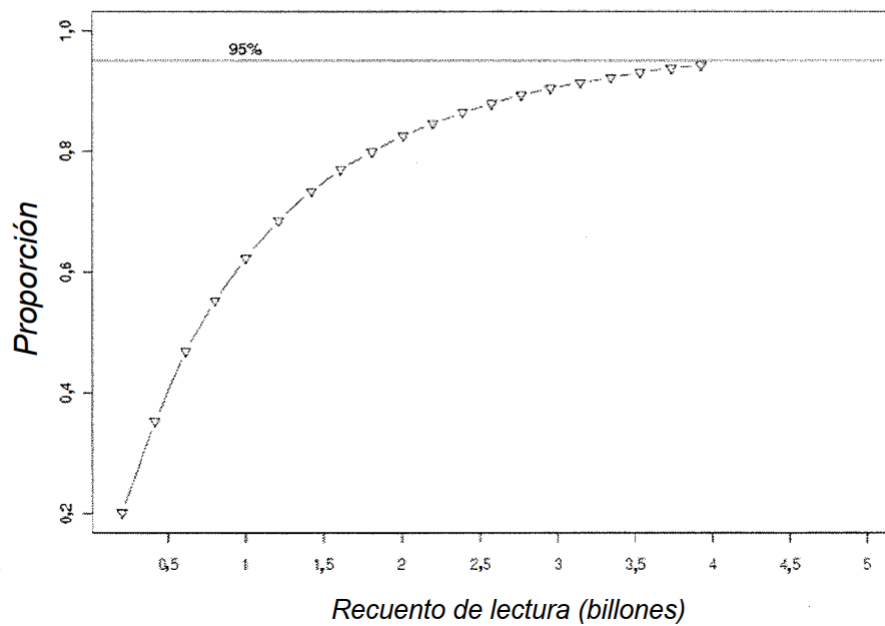


Fig. 28A. Correlación entre la cobertura de alelos parentalmente heredados y recuentos de lectura (madre homocigótica; padre heterocigótico; feto heterocigótico)

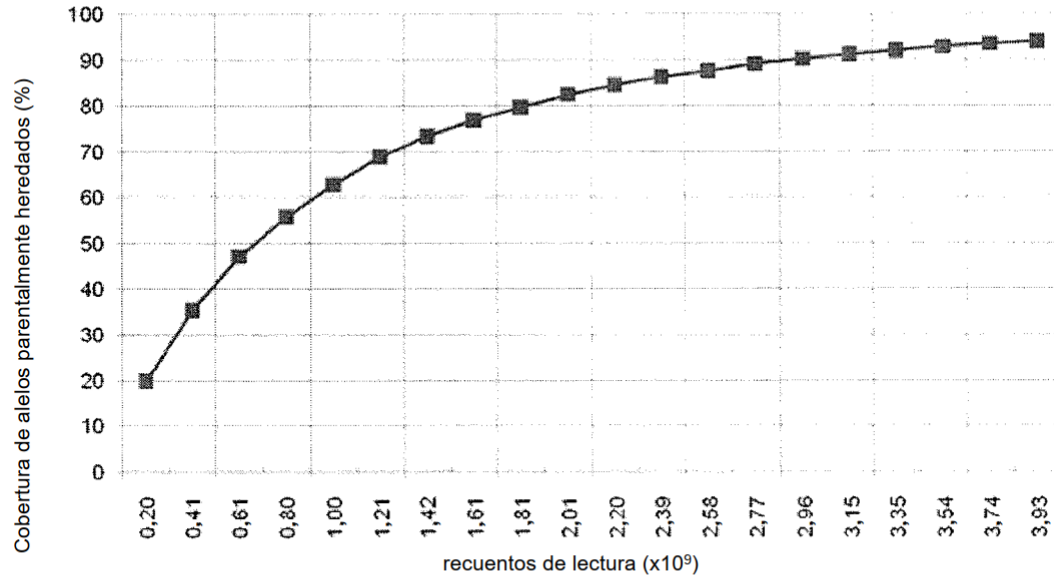


Fig. 28B. Correlación entre la cobertura de alelos parentalmente heredados y número de células de flujo secuenciadas (madre homocigótica; padre heterocigótico; feto heterocigótico)

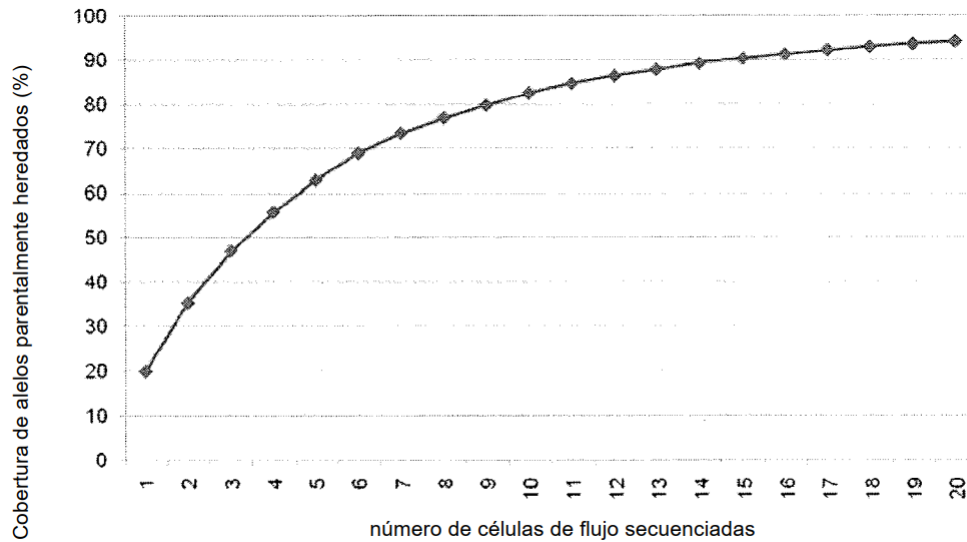


FIG. 29A. Correlación entre la tasa de falsos-positivos y el número de células de flujo secuenciadas

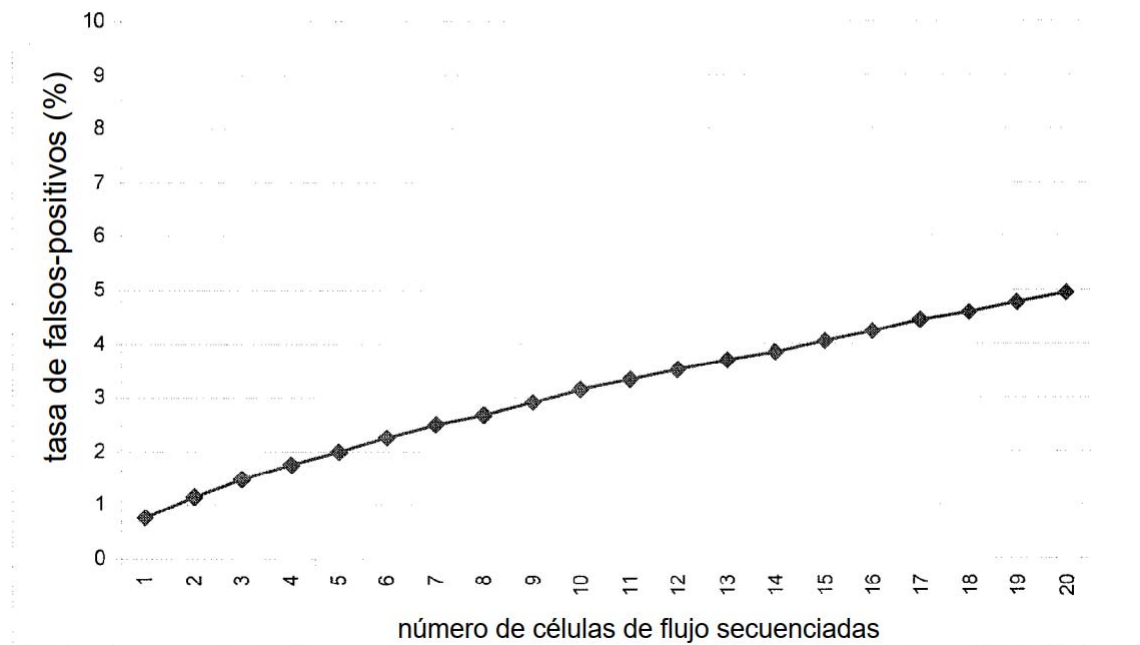
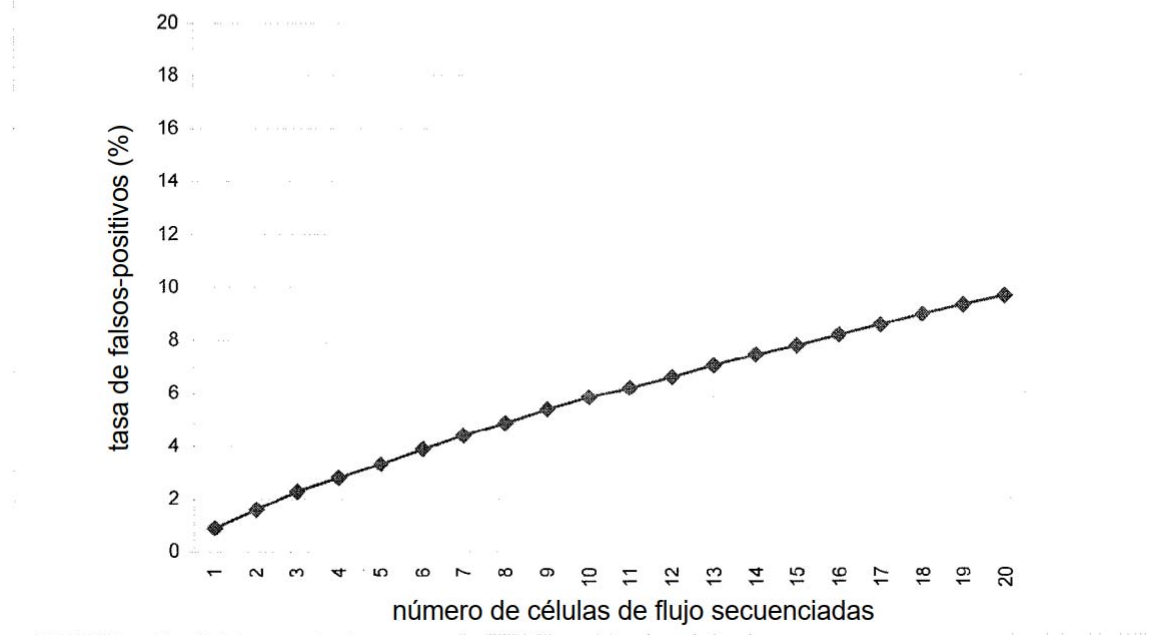


FIG. 29B. Correlación entre la tasa de falsos-positivos y el número de células de flujo secuenciadas (madre homocigótica; padre heterocigótico; con los mismos alelos)



**FIG. 30: Profundidad de cobertura de SNPs con diferentes
números de células de flujo secuenciadas**

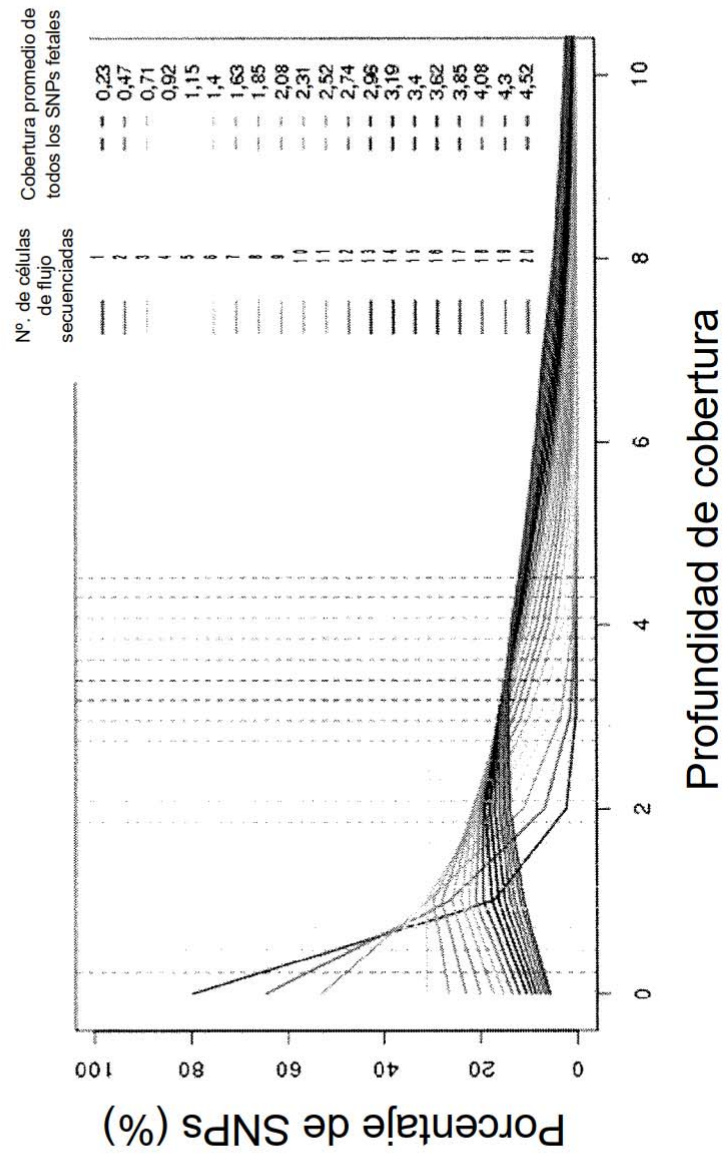


FIG. 31: Precisión de la clasificación Tipo A (10 células de flujo)

Crom.	Clasificación correcta		Clasificación incorrecta		Incorrecto para ambos análisis de entrelazamiento		Total
	No.	Porcentaje	No.	Porcentaje	No.	Porcentaje	
Crom. 1	224	99,1%	2	0,9%	0	0,0%	226
Crom. 2	215	96,8%	7	3,2%	0	0,0%	222
Crom. 3	176	97,2%	5	2,8%	0	0,0%	181
Crom. 4	162	98,8%	2	1,2%	0	0,0%	164
Crom. 5	139	99,3%	1	0,7%	0	0,0%	140
Crom. 6	170	99,4%	1	0,6%	0	0,0%	171
Crom. 7	134	99,3%	1	0,7%	0	0,0%	135
Crom. 8	154	98,7%	2	1,3%	0	0,0%	156
Crom. 9	134	99,3%	1	0,7%	0	0,0%	135
Crom. 10	154	99,4%	1	0,6%	0	0,0%	155
Crom. 11	151	98,7%	2	1,3%	0	0,0%	153
Crom. 12	136	99,3%	1	0,7%	0	0,0%	137
Crom. 13	95	100,0%	0	0,0%	0	0,0%	95
Crom. 14	91	100,0%	0	0,0%	0	0,0%	91
Crom. 15	76	100,0%	0	0,0%	0	0,0%	76
Crom. 16	95	99,0%	1	1,0%	0	0,0%	96
Crom. 17	82	100,0%	0	0,0%	0	0,0%	82
Crom. 18	67	100,0%	0	0,0%	0	0,0%	67
Crom. 19	46	100,0%	0	0,0%	0	0,0%	46
Crom. 20	79	97,5%	2	2,5%	0	0,0%	81
Crom. 21	45	100,0%	0	0,0%	0	0,0%	45
Crom. 22	32	100,0%	0	0,0%	0	0,0%	32
Crom. X	103	96,3%	4	3,7%	0	0,0%	107
Total	2760	98,8%	33	1,2%	0	0,0%	2793

FIG. 32: Precisión de la clasificación Tipo B (10 células de flujo)

Crom.	Clasificación correcta		Clasificación incorrecta		Incorrecto para ambos análisis de entrelazamiento		Total
	No.	Porcentaje	No.	Porcentaje	No.	Porcentaje	
Crom. 1	210	97,2%	6	2,8%	0	0,0%	216
Crom. 2	230	97,9%	5	2,1%	0	0,0%	235
Crom. 3	163	99,4%	1	0,6%	0	0,0%	164
Crom. 4	127	100,0%	0	0,0%	0	0,0%	127
Crom. 5	182	98,9%	2	1,1%	0	0,0%	184
Crom. 6	155	96,9%	5	3,1%	0	0,0%	160
Crom. 7	115	97,5%	3	2,5%	0	0,0%	118
Crom. 8	160	98,8%	2	1,2%	0	0,0%	162
Crom. 9	123	98,4%	2	1,6%	0	0,0%	125
Crom. 10	130	97,0%	4	3,0%	0	0,0%	134
Crom. 11	129	97,7%	3	2,3%	0	0,0%	132
Crom. 12	118	100,0%	0	0,0%	0	0,0%	118
Crom. 13	90	97,8%	2	2,2%	0	0,0%	92
Crom. 14	83	97,6%	2	2,4%	0	0,0%	85
Crom. 15	75	98,7%	1	1,3%	0	0,0%	76
Crom. 16	80	94,1%	5	5,9%	0	0,0%	85
Crom. 17	50	100,0%	0	0,0%	0	0,0%	50
Crom. 18	71	97,3%	2	2,7%	0	0,0%	73
Crom. 19	48	100,0%	0	0,0%	0	0,0%	48
Crom. 20	80	100,0%	0	0,0%	0	0,0%	80
Crom. 21	27	100,0%	0	0,0%	0	0,0%	27
Crom. 22	44	97,8%	1	2,2%	0	0,0%	45
Crom. X	118	97,5%	3	2,5%	0	0,0%	121
Total	2608	98,2%	49	1,8%	0	0,0%	2657

FIG. 33: Precisión del análisis SPRT Tipo A

Crom.	Clasificación de un solo bloque			Algoritmo de dos bloques consecutivos		
	Correcta (%)	Incorrecta (%)	Total	Correcta (%)	Incorrecta (%)	Total
1	314 (99,1%)	3 (0,9%)	317	309 (100%)	0 (0,0%)	309
2	305 (99,3%)	2 (0,7%)	307	301 (100%)	0 (0,0%)	301
3	240 (99,6%)	1 (0,4%)	241	237 (100%)	0 (0,0%)	237
4	218 (99,5%)	1 (0,5%)	219	215 (100%)	0 (0,0%)	215
5	203 (100%)	0 (0,0%)	203	201 (100%)	0 (0,0%)	201
6	240 (98,0%)	5 (2,0%)	245	237 (98,8%)	3 (1,3%)	240
7	182 (98,9%)	2 (1,1%)	184	179 (100%)	0 (0,0%)	179
8	217 (99,5%)	1 (0,5%)	218	214 (100%)	0 (0,0%)	214
9	177 (99,4%)	1 (0,6%)	178	174 (100%)	0 (0,0%)	174
10	217 (99,1%)	2 (0,9%)	219	213 (100%)	0 (0,0%)	213
11	205 (100%)	0 (0,0%)	205	203 (100%)	0 (0,0%)	203
12	193 (99,5%)	1 (0,5%)	194	190 (100%)	0 (0,0%)	190
13	133 (100%)	0 (0,0%)	133	132 (100%)	0 (0,0%)	132
14	125 (100%)	0 (0,0%)	125	124 (100%)	0 (0,0%)	124
15	103 (100%)	0 (0,0%)	103	102 (100%)	0 (0,0%)	102
16	133 (98,5%)	2 (1,5%)	135	129 (100%)	0 (0,0%)	129
17	104 (100%)	0 (0,0%)	104	102 (100%)	0 (0,0%)	102
18	95 (97,9%)	2 (2,1%)	97	91 (100%)	0 (0,0%)	91
19	67 (100%)	0 (0,0%)	67	65 (100%)	0 (0,0%)	65
20	112 (100%)	0 (0,0%)	112	110 (100%)	0 (0,0%)	110
21	60 (100%)	0 (0,0%)	60	59 (100%)	0 (0,0%)	59
22	39 (100%)	0 (0,0%)	39	38 (100%)	0 (0,0%)	38
X	156 (98,7%)	2 (1,3%)	158	152 (100%)	0 (0,0%)	152
Total	3838 (99,4%)	25 (0,6%)	3863	3777 (99,9%)	3 (0,1%)	3780

FIG. 34: Precisión del análisis SPRT Tipo B

Crom.	Clasificación de un solo bloque			Algoritmo de dos bloques consecutivos		
	Correcta (%)	Incorrecta (%)	Total	Correcta (%)	Incorrecta (%)	Total
1	267 (98,2%)	5 (1,8%)	272	262 (100%)	0 (0,0%)	262
2	283 (99,0%)	3 (1,0%)	286	278 (100%)	0 (0,0%)	278
3	222 (98,2%)	4 (1,8%)	226	216 (100%)	0 (0,0%)	216
4	175 (98,9%)	2 (1,1%)	177	171 (100%)	0 (0,0%)	171
5	255 (99,6%)	1 (0,4%)	256	252 (100%)	0 (0,0%)	252
6	195 (99,0%)	2 (1,0%)	197	192 (100%)	0 (0,0%)	192
7	155 (97,5%)	4 (2,5%)	159	150 (100%)	0 (0,0%)	150
8	211 (100%)	0 (0,0%)	211	209 (100%)	0 (0,0%)	209
9	162 (99,4%)	1 (0,6%)	163	160 (100%)	0 (0,0%)	160
10	172 (96,6%)	6 (3,4%)	178	167 (98,2%)	3 (1,8%)	170
11	171 (99,4%)	1 (0,6%)	172	168 (100%)	0 (0,0%)	168
12	156 (99,4%)	1 (0,6%)	157	153 (100%)	0 (0,0%)	153
13	119 (98,3%)	2 (1,7%)	121	117 (99,2%)	1 (0,8%)	118
14	116 (98,3%)	2 (1,7%)	118	115 (99,1%)	1 (0,9%)	116
15	101 (100%)	0 (0,0%)	101	100 (100%)	0 (0,0%)	100
16	115 (94,3%)	7 (5,7%)	122	107 (99,1%)	1 (0,9%)	108
17	67 (98,5%)	1 (1,5%)	68	65 (100%)	0 (0,0%)	65
18	92 (100%)	0 (0,0%)	92	90 (100%)	0 (0,0%)	90
19	57 (100%)	0 (0,0%)	57	55 (100%)	0 (0,0%)	55
20	95 (100%)	0 (0,0%)	95	93 (100%)	0 (0,0%)	93
21	39 (100%)	0 (0,0%)	39	38 (100%)	0 (0,0%)	38
22	59 (100%)	0 (0,0%)	59	58 (100%)	0 (0,0%)	58
X	142 (99,3%)	1 (0,7%)	143	139 (100%)	0 (0,0%)	139
Total	3426 (98,8%)	43 (1,2%)	3469	3355 (99,8%)	6 (0,2%)	3361

FIG. 35A

Lecturas con la mutación CD41/42 paternalmente heredada

Recuento = 10

CGGGTATTGTCGTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGAC
CGGGTATTGTCGTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGAC
GGGTATTGTCGTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGACC
GGTATTGTCGTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGACCC
CGTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGACCCAGGTTCCC
GTAGTCCTCACCTGCTAGGGGTTTCCTGAGTTGGAGACCCAGGTTCCCA
CTGCTAGGGGTTTCCTGAGTTGGAGACCCAGGTTCCCATCTGGTGGTCG
TGTCTAGGGGTTTCCTGAGTTGGAGACCCAGGTTCCCATCTGGTGGTCGT
TTCTGAGTGGGAGACCCAGGTTCCCATCTGGTGGTCGTCGGATTCCCAC
GTGGGAGACCCAGGTTCCCATCTGGTGGTCGTCGGATTCCCACCTTTTA

FIG. 35B

[illegible]

Lecturas con la secuencia tipo salvaje en CD41/42

Recuento = 62

Por tanto, % mutante = $10/72 = 0,1389$

FIG. 36: Análisis Tipo A en el grupo de β -globina sobre el cromosoma 11

Genotipo										Recuento Recuento					
Crom.	dbSNPId	Loc	Padre	Madre	Feto	Mutación nt-28 de beta-globina de talasemia				Recuento Recuento acumul.				Clasificación	
						Hap III	Hap IV	HapIII	HapIV	Hap III	Hap IV	Hap III	Hap IV		
Crom. 11		5204905													
crom.11	rs10742583	5205217	TT	CT	TT	T	C	22	26	22	26	22	26		No clasificado
crom.11	rs10742584	5205346	TT	CT	TT	T	C	21	12	43	38	43	38		No clasificado
crom.11	rs1003586	5205946	GG	AG	GG	G	A	24	11	67	49	67	49		No clasificado
crom.11	rs6578588	5208827	GG	AG	GG	G	A	43	38	110	87	110	87		No clasificado
crom.11	rs3759076	5213765	CC	CT	CC	C	T	29	0	139	87	139	87		No clasificado
crom.11	rs10837643	5214614	AA	AT	AA	A	T	23	11	162	98	162	98		No clasificado
crom.11	rs4283007	5215066	TT	AT	TT	T	A	53	36	215	134	215	134		Hap III

FIG. 37: Análisis Tipo B en el grupo de β -globina sobre el

Crom.	dbSNPId	Loc	Padre	Madre	Feto	Hap III	Hap IV	Recuento			Clasificación	
								HapIII	HapIV	acumul.		
Crom. 11		5204905	Mutación nt-28 de beta-globina de talasemia									
crom.11	rs12275403	5286913	CC	CG	CG	G	C	19	23	19	23	No clasificado
crom.11	rs7926089	5289019	CC	CT	CT	T	C	13	21	32	44	No clasificado
crom.11	rs12365420	5289330	GG	GT	GT	T	G	5	12	37	56	No clasificado
crom.11	rs4910747	5294746	CC	AC	AC	A	C	27	36	64	92	No clasificado
crom.11	rs6578604	5299185	AA	AG	AG	G	A	15	18	79	110	No clasificado
crom.11	rs10837814	5301704	GG	AG	AG	A	G	39	45	118	155	No clasificado
crom.11	rs4910550	5302436	GG	AG	AG	A	G	15	22	133	177	No clasificado
crom.11	rs2723381	5304451	TT	CT	CT	C	T	38	34	171	211	No clasificado
crom.11	rs6578606	5306671	GG	CG	CG	C	G	40	31	211	242	No clasificado
crom.11	rs3824950	5585548	CC	CT	CT	T	C	59	66	270	308	No clasificado
crom.11	rs10838441	5601579	GG	CG	CG	C	G	20	10	290	318	No clasificado
crom.11	rs11038628	5645516	CC	CT	CT	T	C	65	5	355	323	Hap III

Coordenados genómicos del SNP	Alelos maternos en		Recuentos para alelos en		Recuentos acumulativos para alelos en		SPRT Clasificación
	Hap I	Hap II	Hap I	Hap II	Hap I	Hap II	
3251050	A	G	22	15	22	15	No clasificado
3251284	G	A	14	10	36	25	No clasificado
3251791	A	G	18	11	54	36	Hap I
9865342	A	G	5	3	5	3	No clasificado
9872792	A	G	7	5	12	8	No clasificado
9874947	G	C	1	3	13	11	No clasificado
15242479	G	A	17	5	30	16	Hap I
15710521	G	A	1	2	1	2	No clasificado
15710672	A	G	13	13	14	15	No clasificado
21895385	C	T	15	3	29	18	No clasificado
22149641	T	C	5	3	34	21	Hap I
22201527	A	C	11	10	11	10	No clasificado
23806924	C	T	20	9	31	19	No clasificado
31134605	A	G	2	1	33	20	Hap I
32413115	C	T	6	8	6	8	No clasificado
32473184	G	A	8	9	14	17	No clasificado
35730716	A	G	17	14	31	31	No clasificado
35903375	C	T	7	9	38	40	No clasificado
35903407	G	A	6	2	44	42	No clasificado
35917535	T	G	6	7	50	49	No clasificado
43475980	T	G	7	16	57	65	No clasificado
47351305	A	G	7	6	64	71	No clasificado
48561783	A	G	5	1	69	72	No clasificado
50676020	C	T	19	15	88	87	No clasificado
53601143	A	G	8	10	96	97	No clasificado
54854802	T	C	9	3	105	100	No clasificado
74921254	C	A	3	3	108	103	No clasificado
78102680	A	C	5	3	113	106	No clasificado
78313644	C	T	40	24	153	130	Hap I

FIG. 38A

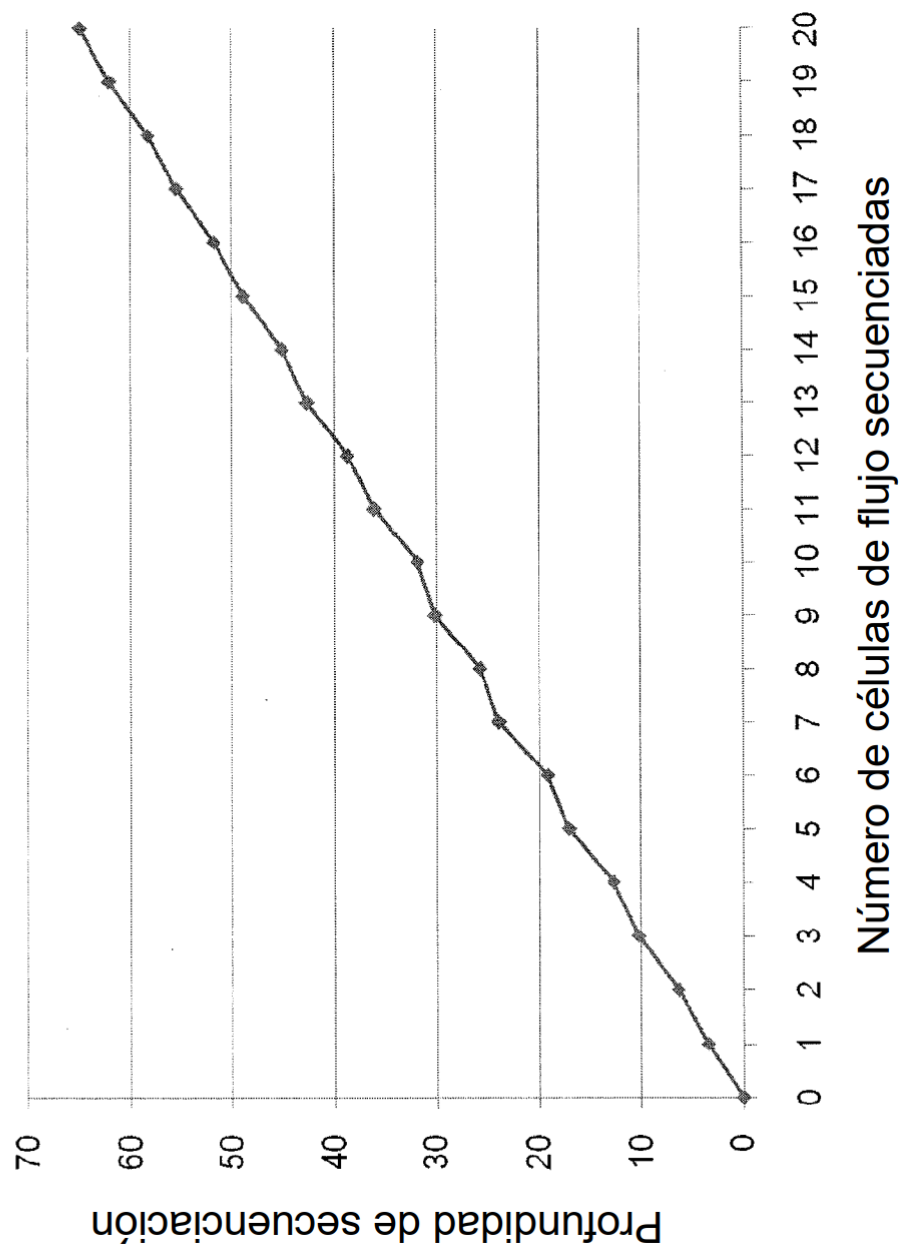
Coordenados genómicos del SNP	Alelos maternos en		Recuentos para alelos en		Recuentos acumulativos para alelos en		SPRT Clasificación
	Hap I	Hap II	Hap I	Hap II	Hap I	Hap II	
84236666	A	G	3	0	3	0	No clasificado
84249796	A	G	29	12	32	12	Hap I
100273993	T	C	16	10	16	10	No clasificado
100494847	A	G	13	5	29	15	Hap I
106283058	G	A	5	8	5	8	No clasificado
106283111	G	A	13	12	18	20	No clasificado
107283614	G	A	10	4	28	24	No clasificado
107335295	T	G	5	1	33	25	No clasificado
114785655	C	T	11	7	44	32	No clasificado
117584169	G	A	8	2	52	34	Hap I
118471031	T	C	3	5	3	5	No clasificado
118488464	C	T	5	15	8	20	No clasificado
119474561	A	T	5	4	13	24	No clasificado
123022716	C	A	12	12	30	36	No clasificado
128506711	C	G	4	3	34	39	No clasificado
130047404	C	T	4	1	38	40	No clasificado
130243499	C	T	13	19	51	59	No clasificado
134814366	C	T	12	8	63	67	No clasificado
134818744	G	A	27	18	90	85	No clasificado
134821671	C	T	5	9	95	94	No clasificado
138507378	A	T	7	3	102	97	No clasificado
138695073	A	G	10	3	112	100	No clasificado
144713436	C	T	10	16	122	116	No clasificado
149577161	A	G	11	5	133	121	No clasificado
150888835	C	A	11	11	144	132	No clasificado
152381349	T	C	8	3	152	135	Hap I
152694738	C	T	1	0	1	0	No clasificado
152848587	G	A	4	6	5	6	No clasificado

FIG. 38B

Caso	Nº total de SNPs informativos	Profundidad de secuenciación promedio para cada SNP (veces)	Concentración de la fracción de ADN fetal (%)	Nº total de clasificaciones de SPRT	Nº de clasificaciones correctas	Nº de clasificaciones incorrectas
PW226	57	17,6	19,4	9	9	0
PW263	62	15,8	18,4	4	4	0
PW316	50	18,5	17,4	2	2	0
PW370	54	18,9	11,5	1	1	0
PW421	46	19,6	15,3	2	2	0

FIG. 39

FIG. 40: Profundidad de secuenciación frente al número de células de flujo secuenciadas



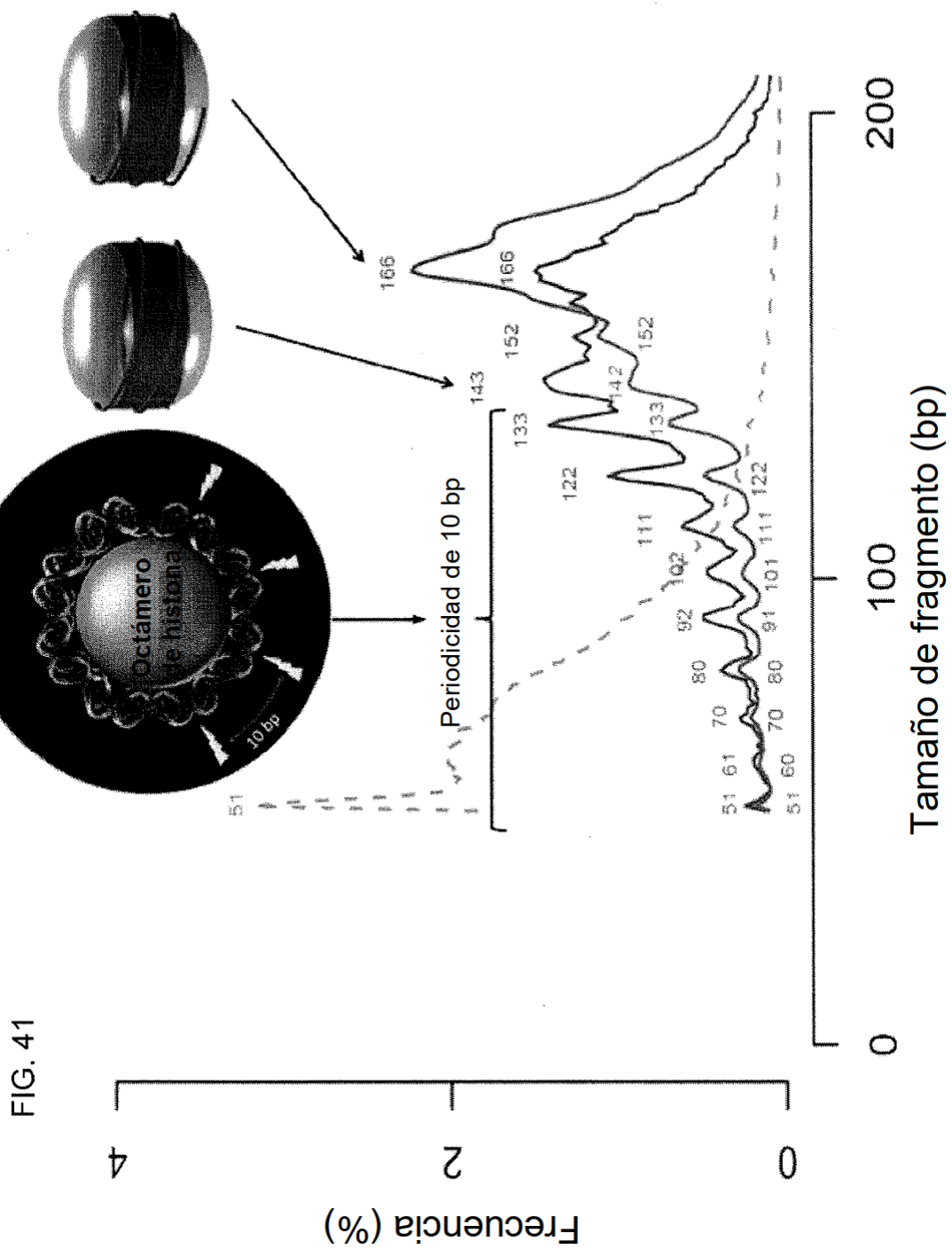
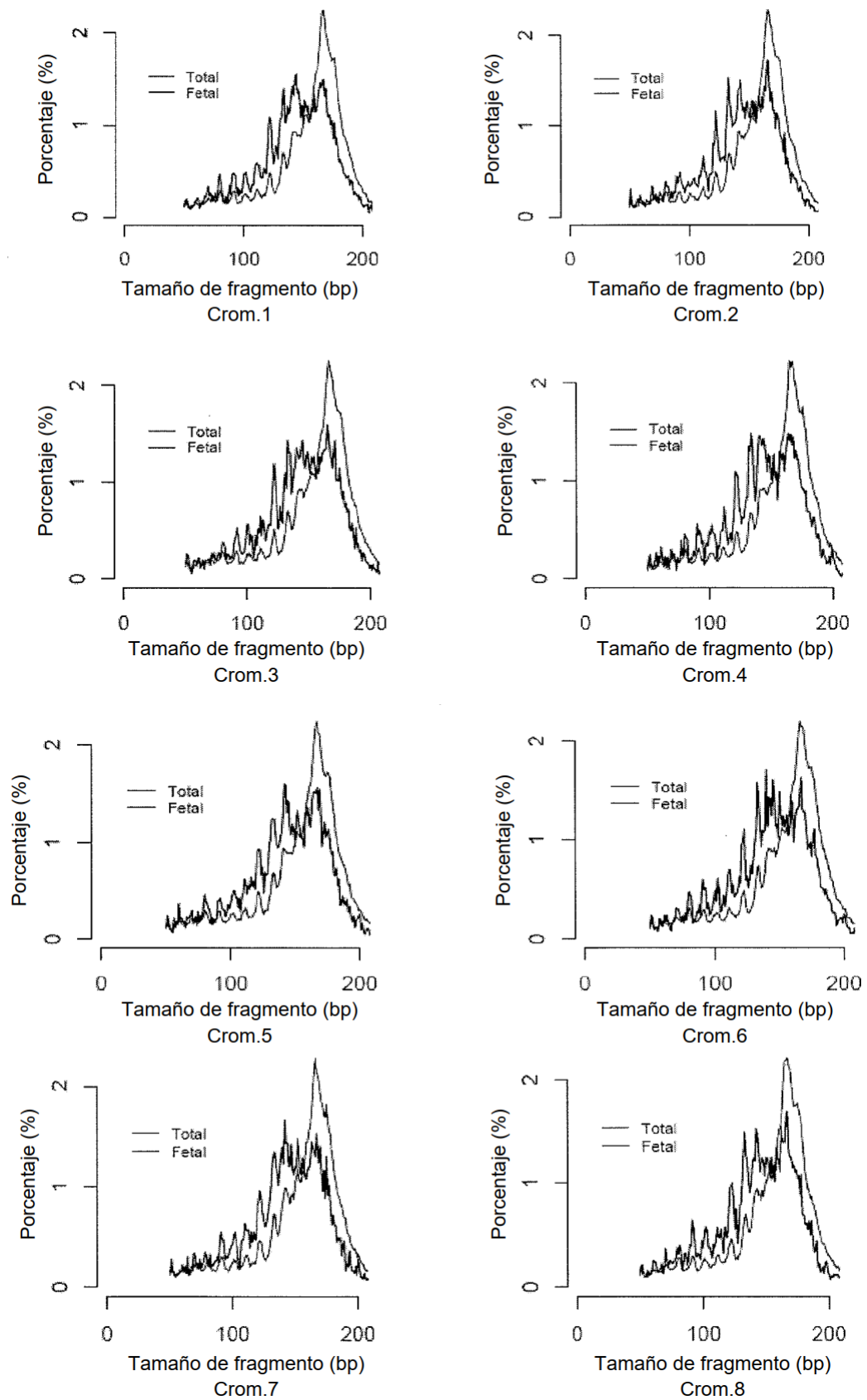


FIG. 42A



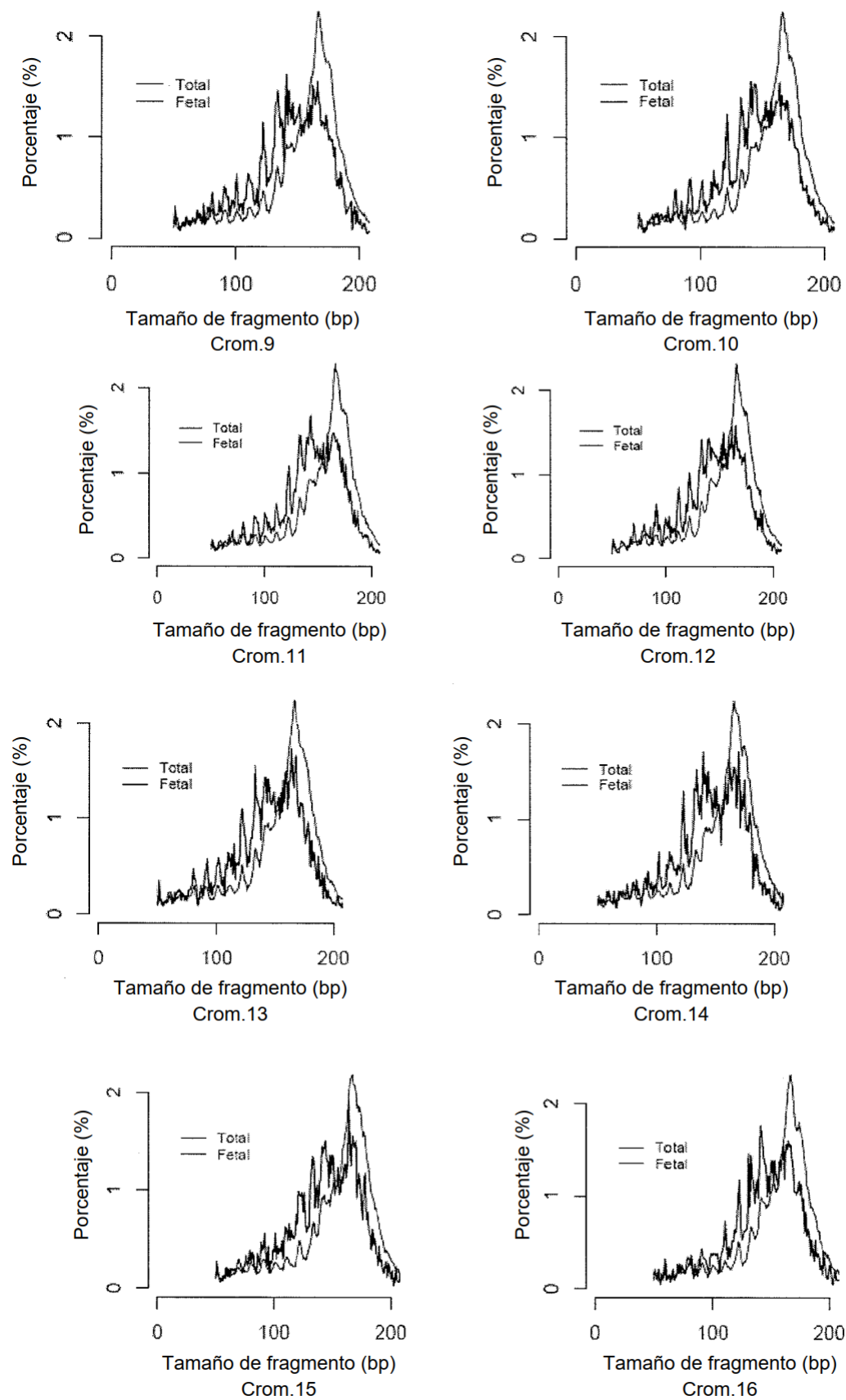


FIG. 42B

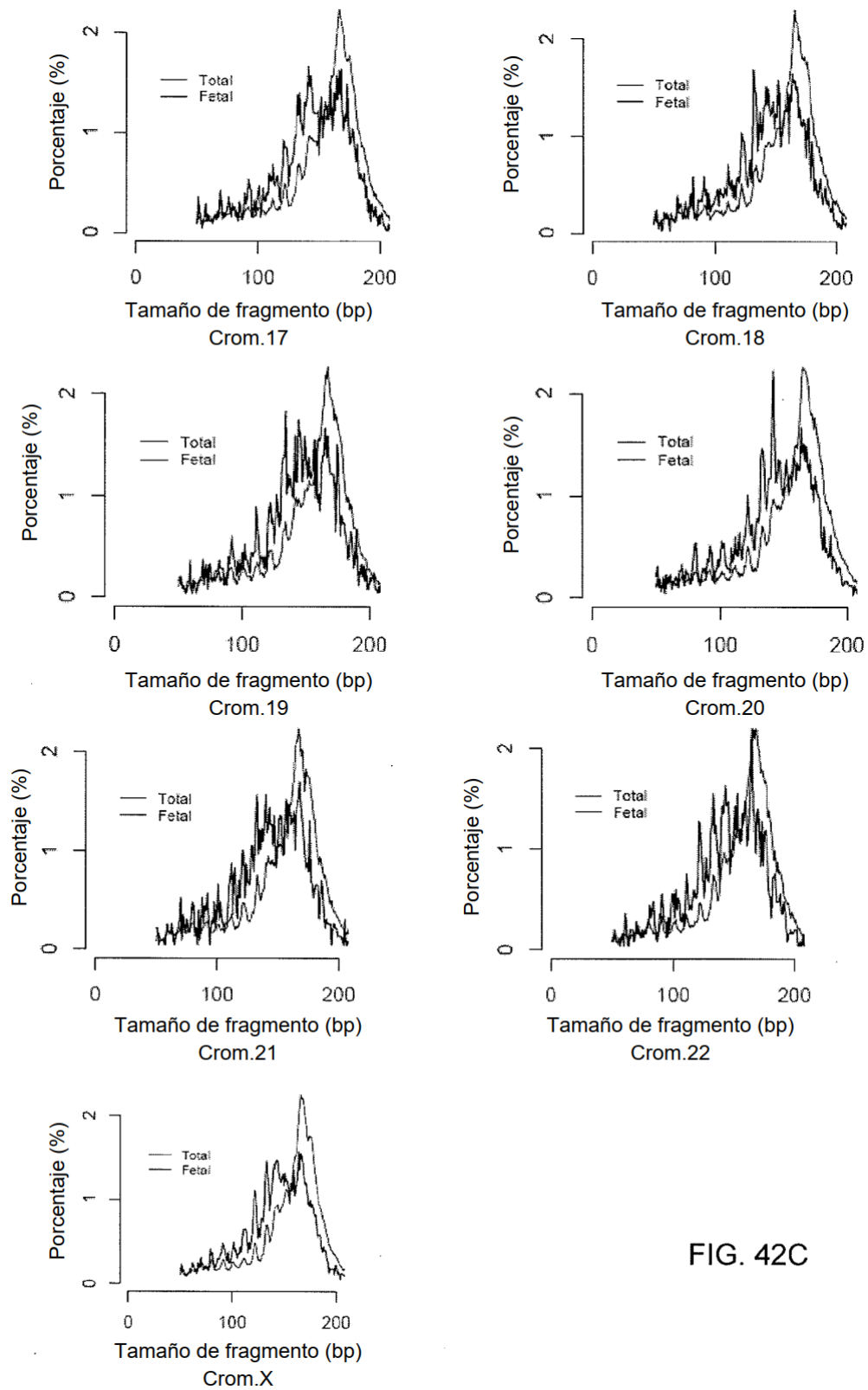


FIG. 42C

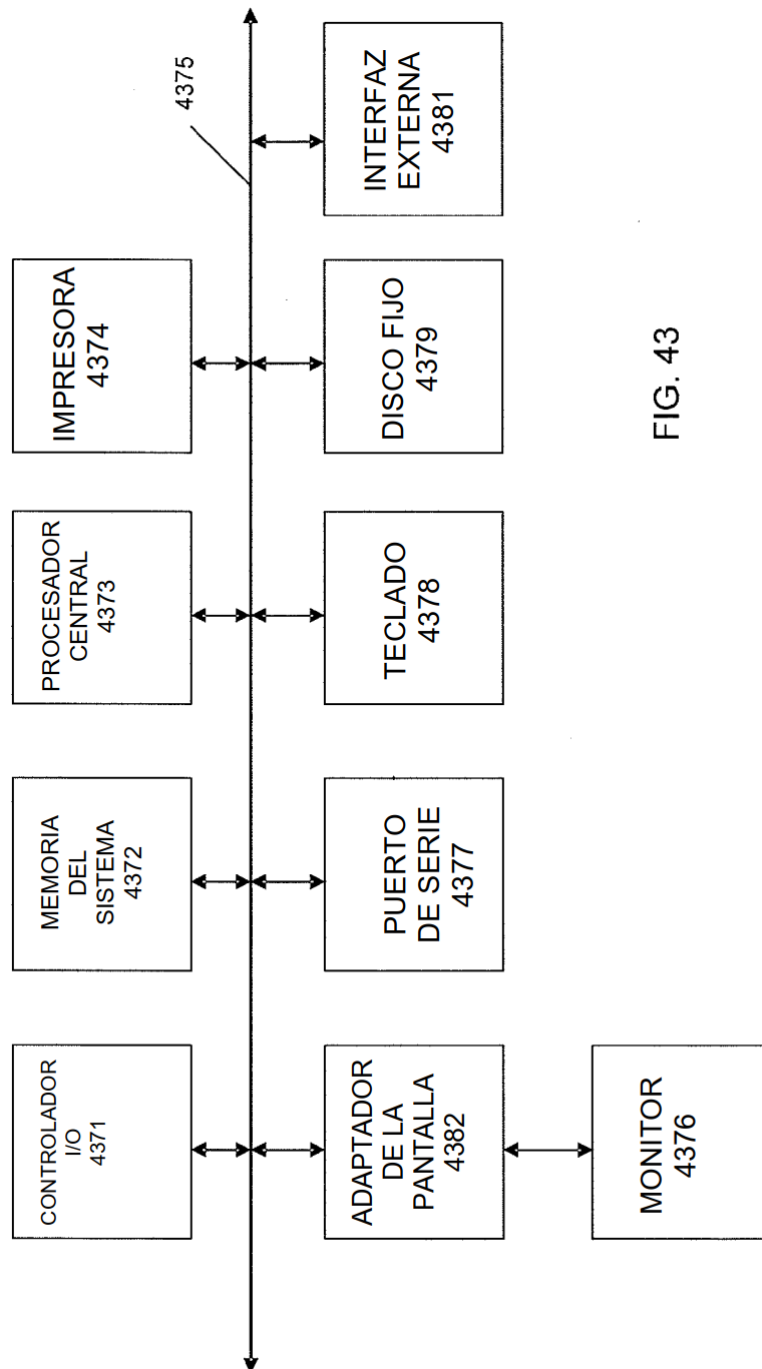


FIG. 43