(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 534 758**

(51) Int. CI.:

C12Q 1/68 (2006.01)
C12P 19/34 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **01.12.2010 E 10830938 (6)**

(97) Fecha y número de publicación de la concesión europea: **21.01.2015 EP 2366031**

(54) Título: **Métodos de secuenciación en diagnósticos prenatales**

(30) Prioridad:

19.01.2010 US 296358 P
01.07.2010 US 360837 P
26.10.2010 US 407017 P
26.10.2010 US 455849 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
28.04.2015

(73) Titular/es:

VERINATA HEALTH, INC. (100.0%)
800 Saginaw Drive
Redwood City CA 94063 , US

(72) Inventor/es:

RAVA, RICHARD P.;
CHINNAPPA, MANJULA;
COMSTOCK, DAVID A.;
HEILEK, GABRIELLE y
RHEES, BRIAN KENT

(74) Agente/Representante:

IZQUIERDO BLANCO, María Alicia

Métodos de secuenciación en diagnósticos prenatales

Descripción

1. CAMPO DE LA INVENCION

La invención es aplicable al campo de los diagnósticos prenatales y se refiere particularmente a métodos de secuenciación masivamente paralelos para determinar la presencia o ausencia de aneuploidías y/o fracción fetal.

2. ANTECEDENTES DE LA INVENCION

La detección y diagnóstico prenatal son una parte rutinaria del cuidado prenatal. Actualmente, el diagnóstico prenatal de condiciones genéticas o cromosómicas implica pruebas invasivas, como la amniocentesis o muestreo de vellosidades coriónicas (CVS), realizadas desde las 11 semanas de gestación y que implica un ~1% de riesgo de aborto involuntario. LA existencia de ADN libre de células circulante en la sangre materna (Lo et al., Lancet 350:485-487 [1997]) se está aprovechando para desarrollar procesos no invasivos que usan ácidos nucleicos fetales de una muestra de sangre periférica materna para determinar anomalías cromosómicas fetales (Fan HC y Quake SR Anal Chem 79:7576-7579 [2007]; Fan et al., Proc Natl Acad Sci 105:16266-16271 [2008] y Chu et al., Bionformatics, vol. 25, no. 10, pg. 1244-1250, que también divulgan métodos para preparas bibliotecas de secuenciación adecuadas y sus usos en métodos de secuenciación masivamente paralelos). Estos métodos ofrecen alternativas y fuente segura de material genético para el diagnóstico prenatal, y podrían anunciar el final de los procedimientos invasivos.

La secuenciación de ácidos nucleicos está evolucionando rápidamente como una técnica de diagnóstico en el laboratorio clínico. Las aplicaciones que implican secuenciación se ven en varias áreas, incluyendo pruebas de cáncer que abarcan pruebas genéticas para predisposición al cáncer y la evaluación de mutaciones genéticas en el cáncer; genéticas que abarcan pruebas de portadores y diagnóstico de enfermedades trasmitidas genéticamente; y microbiología que abarca genotipado viral y secuencias asociadas con la resistencia a los fármacos.

La llegada de tecnologías de secuenciación de próxima generación (NGS) que permiten la secuenciación de genomas completos en relativamente corto tiempo, ha proporcionado la oportunidad de comparar material genético originado de un cromosoma para que sea comparado con otro sin los riesgos asociados a los métodos de muestreo invasivos. Sin embargo, las limitaciones de los métodos existentes, que incluyen insuficiente sensibilidad derivada de los niveles limitados de ADN libre de células, y el sesgo de la secuenciación de la tecnología derivada de la naturaleza inherente de la información genómica, subyace la necesidad continuada para métodos no invasivos que proporcionen cualquiera o todos de especificidad, sensibilidad y aplicabilidad, para diagnosticas aneuploidías fetales en una variedad de entornos clínicos.

A medida que la secuenciación de ácidos nucleicos ha entrado en el ámbito clínico para las pruebas de cáncer, organizaciones como la NCCLS (National Council Of Clinical Laboratory Services) y la Association of Clinical Cytogenetics han proporcionado directrices para la estandarización de las pruebas basadas en secuenciación existentes que usan secuenciación basada en PCR, didesoxi-terminador, y extensión del cebador hechas en secuenciadores basados en gel o capilares (NCCLS: Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine MM9-A, Vol. 24 No. 40), secuenciación Sanger y QF-PCR (Association for Clinical Cytogenetics and Clinical Molecular Genetics Society, Practice Guidelines for Sanger Sequencing Analysis and Interpretation ratificada por el CMGS Executive Committee el 7 de agosto del 2009 disponible en la dirección web cm-gs.org/BPGs/pdfs%20current%20bpgs/Sequencingv2.pdf QF-PCR for the diagnosis of aneuploidy best practice guidelines (2007) v2.01). Las directrices se basan en pruebas de consenso de varios protocolos y entre otros aspiran a reducir la aparición de eventos adversos en el laboratorio clínico, por ejemplo, mezclas de muestras, mientras se conserva la calidad y fiabilidad de los ensayos. Como los laboratorios clínicos ya están experimentando con NIPS, los procedimientos de calidad para implementar las nuevas tecnologías de secuenciación se desarrollaran para preparar sistemas de administración de cuidado de la salud seguros y apropiados.

La presente invención proporciona métodos para preparar una biblioteca de secuenciación y su uso en los métodos de secuenciación de próxima generación que sean aplicables al menos a la práctica de diagnósticos prenatales no invasivos, y abarca procedimientos que aumentan la velocidad y calidad de los métodos a la vez que minimizan la pérdida de material, y reducen la probabilidad de errores de la muestra.

3. RESUMEN DE LA INVENCION

La invención se refiere a un nuevo protocolo para preparar bibliotecas de secuenciación que inesperadamente mejora la calidad del ADN de la biblioteca a la vez que agiliza el proceso de análisis de muestras para diagnósticos prenatales.

La invención proporciona un método para preparar una librería de secuenciación de una muestra materna

que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos, y en donde los pasos consecutivos excluyen purificar los productos reparados en el extremo antes del paso de la adición de colas de dA y excluyen purificar los productos de la adición de colas de dA antes del paso de ligado por adaptadores.

En una realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar una aneuploidía cromosómica fetal en una muestra de sangre materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla de moléculas de ácidos nucleicos fetales y maternos; en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos; (b) secuenciar al menos una porción de las moléculas de ácidos nucleicos, obteniendo de esta manera información de la secuencia para una pluralidad de moléculas de ácidos nucleicos fetales y maternos de una muestra de sangre materna; (c) usar la información de la secuencia para obtener una dosis de cromosoma par un cromosoma aneuploide; y (d) comparar la dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal.

En otra realización, se divulga en la presenta un uso de dicha biblioteca en un método para determinar una aneuploidía cromosómica fetal en una muestra de sangre materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla de moléculas de ácidos nucleicos fetales y maternos; en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos; (b) secuenciar al menos una porción de las moléculas de ácidos nucleicos, obteniendo de esta manera información de la secuencia para una pluralidad de moléculas de ácidos nucleicos fetales y maternos de una muestra de sangre materna; (c) usar la información de la secuencia para obtener una dosis de cromosoma para un cromosoma aneuploide; y (d) comparar la dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal. El método comprende además usar la información de la secuencia para identificar un número de etiquetas de la secuencia mapeada para al menos un cromosoma normalizador y para un cromosoma aneuploide; y usar el número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada identificadas para al menos un cromosoma normalizador para calcular una dosis de cromosoma para dicho cromosoma aneuploide como una proporción del número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada para el al menos un cromosoma normalizador. Opcionalmente, calcular la dosis de cromosoma comprende (i) calcular una proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide, relacionando el número de etiquetas de la secuencia mapeada identificadas para el cromosoma aneuploide en el paso con la longitud de dicho cromosoma aneuploide; (ii) calcular una proporción de densidad de etiqueta de secuencia para el al menos un cromosoma normalizador, relacionando el número de etiquetas de la secuencia mapeada identificadas para dicho al menos un cromosoma normalizador con la longitud del al menos un cromosoma normalizador; y (iii) usar las proporciones de densidad de etiqueta de secuencia calculadas en los pasos (i) y (ii) para calcular una dosis de cromosoma para el cromosoma aneuploide, en donde la dosis de cromosoma se calcula como la proporción de la proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide y la proporción de densidad de etiqueta de secuencia para el al menos un cromosoma normalizador.

En otra realización, se divulga en la presenta un uso de dicha biblioteca en un método para determinar una aneuploidía cromosómica fetal en una muestra de sangre materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla de moléculas de ácidos nucleicos fetales y maternos; en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos; (b) secuenciar al menos una porción de las moléculas de ácidos nucleicos, obteniendo de esta manera información de la secuencia para una pluralidad de moléculas de ácidos nucleicos fetales y maternos de una muestra de sangre materna; (c) usar la información de la secuencia para obtener una dosis de cromosoma para un cromosoma aneuploide; y (d) comparar la dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal. El método comprende además usar la información de la secuencia para identificar un número de etiquetas de la secuencia mapeada para al menos un cromosoma normalizador y para un cromosoma aneuploide; y usar el número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada identificadas para el al menos un cromosoma normalizador para calcular una dosis de cromosoma para dicho cromosoma aneuploide como una proporción del número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada identificadas para el al menos un cromosoma normalizador. El al menos un cromosoma normalizador es un cromosoma que tiene la variabilidad más pequeña y/o la diferenciabilidad más grande. Opcionalmente, calcular la dosis de cromosoma comprende (i) calcular una proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide, relacionando el número de etiquetas de la secuencia mapeada identificadas para el cromosoma aneuploide en el paso con la longitud de dicho cromosoma aneuploide; (ii) calcular una proporción de densidad de etiqueta de secuencia para el al menos un

cromosoma normalizador, relacionando el número de etiquetas de la secuencia mapeada identificadas para dicho al menos un cromosoma normalizador con la longitud del al menos un cromosoma normalizador; y (iii) usar las proporciones de densidad de etiqueta de secuencia calculadas en los pasos (i) y (ii) para calcular una dosis de cromosoma para el cromosoma aneuploide, en donde la dosis de cromosoma se calcula como la proporción de la proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide y la proporción de densidad de etiqueta de secuencia para el al menos un cromosoma normalizador.

En otra realización, se divulga en la presenta un uso de dicha biblioteca en un método para determinar una aneuploidía cromosómica fetal en una muestra de sangre materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla de moléculas de ácidos nucleicos fetales y maternos; en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos; (b) secuenciar al menos una porción de las moléculas de ácidos nucleicos, obteniendo de esta manera información de la secuencia para una pluralidad de moléculas de ácidos nucleicos fetales y maternos de una muestra de sangre materna; (c) usar la información de la secuencia para obtener una dosis de cromosoma para un cromosoma aneuploide; y (d) comparar la dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal. El método comprende además usar la información de la secuencia para identificar un número de etiquetas de la secuencia mapeada para al menos un cromosoma normalizador y para un cromosoma aneuploide; y usar el número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada identificadas para el al menos un cromosoma normalizador para calcular una dosis de cromosoma para dicho cromosoma aneuploide como una proporción del número de etiquetas de la secuencia mapeada identificadas para dicho cromosoma aneuploide y el número de etiquetas de la secuencia mapeada identificadas para el al menos un cromosoma normalizador. Opcionalmente, calcular la dosis de cromosoma comprende (i) calcular una proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide, relacionando el número de etiquetas de la secuencia mapeada identificadas para el cromosoma aneuploide en el paso con la longitud de dicho cromosoma aneuploide; (ii) calcular una proporción de densidad de etiqueta de secuencia para el al menos un cromosoma normalizador, relacionando el número de etiquetas de la secuencia mapeada identificadas para dicho al menos un cromosoma normalizador con la longitud del al menos un cromosoma normalizador; y (iii) usar las proporciones de densidad de etiqueta de secuencia calculadas en los pasos (i) y (ii) para calcular una dosis de cromosoma para el cromosoma aneuploide, en donde la dosis de cromosoma se calcula como la proporción de la proporción de densidad de etiqueta de secuencia para el cromosoma aneuploide y la proporción de densidad de etiqueta de secuencia para el al menos un cromosoma normalizador. En realizaciones, en las que el cromosoma aneuploide es el cromosoma 21, el al menos un cromosoma normalizador es seleccionado del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 11, cromosoma 12 y cromosoma 14. Alternativamente, el al menos un cromosoma normalizador para el cromosoma 21 es un grupo de cromosomas seleccionados del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 11, cromosoma 12 y cromosoma 14. En realizaciones en las que el cromosoma aneuploide es el cromosoma 118, el al menos un cromosoma normalizador es seleccionado del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 5, cromosoma 6, cromosoma 12 y cromosoma 14. Alternativamente, el al menos un cromosoma normalizador para el cromosoma 18 es un grupo de cromosomas seleccionados del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 5, cromosoma 6, cromosoma 12 y cromosoma 14. En realizaciones en las que el cromosoma aneuploide es el cromosoma 13, el al menso un cromosoma normalizador es seleccionado del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6 y cromosoma 8. Alternativamente, el al menos un cromosoma normalizador para el cromosoma 13 es un grupo de cromosomas seleccionados del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6 y cromosoma 8. En realizaciones, en las que el cromosoma aneuploide es el cromosoma X, el al menos un cromosoma normalizador es seleccionado del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6 y cromosoma 8. Alternativamente, el al menos un cromosoma normalizador para el cromosoma X es un grupo de cromosomas seleccionados del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6 y cromosoma 8.

La muestra materna usada en las realizaciones del método para determinar una aneuploidía cromosómica fetal es un fluido biológico seleccionado de sangre, plasma, suero, orina y saliva. Preferiblemente, la muestra materna es una muestra de plasma. En algunas realizaciones, las moléculas de ácidos nucleicos comprendidas en la muestra materna son moléculas de ADN libre de células. En algunas realizaciones, los pasos consecutivos comprendidos en la preparación de la biblioteca de secuenciación se realizan en menos de una hora. Preferiblemente, los pasos consecutivos se realizan en ausencia de polietilenglicol. Más preferiblemente, los pasos consecutivos excluyen la purificación. La secuenciación de la biblioteca de secuenciación se consigue por métodos de secuenciación de próxima generación (NGS). En algunas realizaciones, la secuenciación comprende una amplificación. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por ligadura. En todavía otras realizaciones, la secuenciación es secuenciación de moléculas individuales.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la presencia o ausencia de una aneuploidía en una muestra materna que comprende una mezcla de moléculas de

ácidos nucleicos fetales y maternas, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla, en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos fetales y maternos; (b) secuenciar al menos una porción de la biblioteca de secuenciación, en donde la secuenciación comprende proporcionar una pluralidad de etiquetas de secuencia; y (c) en base a la secuenciación, determinar la presencia o ausencia de aneuploidía en la muestra.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la presencia o ausencia de una aneuploidía cromosómica o una parcial en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla, en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos fetales y maternos; (b) secuenciar al menos una porción de la biblioteca de secuenciación, en donde la secuenciación comprende proporcionar una pluralidad de etiquetas de secuencia; y (c) en base a la secuenciación, determinar la presencia o ausencia de la aneuploidía cromosómica o parcial en la muestra.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la presencia o ausencia de una aneuploidía cromosómica en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla, en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos fetales y maternos; (b) secuenciar al menos una porción de la biblioteca de secuenciación, en donde la secuenciación comprende proporcionar una pluralidad de etiquetas de secuencia; y (c) en base a la secuenciación, determinar la presencia o ausencia de aneuploidía cromosómica en la muestra. Las aneuploidías cromosómicas pueden determinarse de acuerdo con el método de incluir trisomía 8, trisomía 13, trisomía 15, trisomía 16, trisomía 18, trisomía 21, trisomía 22, monosomía X, y XXX.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la presencia o ausencia de una aneuploidía cromosómica o una parcial en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla, en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos fetales y maternos; (b) secuenciar al menos una porción de la biblioteca de secuenciación, en donde la secuenciación comprende proporcionar una pluralidad de etiquetas de secuencia; y (c) en base a la secuenciación, determinar la presencia o ausencia de una aneuploidía cromosómica o una parcial en la muestra que comprende calcular una dosis de cromosoma en base al número de dichas etiquetas de secuencia para un cromosoma de interés y para un cromosoma normalizador, y comparar dicha dosis con un valor límite.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la presencia o ausencia de una aneuploidía cromosómica en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) preparar una biblioteca de secuenciación de la mezcla, en donde preparar dicha biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichos ácidos nucleicos fetales y maternos; (b) secuenciar al menos una porción de la biblioteca de secuenciación, en donde la secuenciación comprende proporcionar una pluralidad de etiquetas de secuencia; y (c) en base a la secuenciación, determinar la presencia o ausencia de aneuploidía cromosómica en la muestra que comprende calcular una dosis de cromosoma en base al número de dichas etiquetas de secuencia para un cromosoma de interés y para un cromosoma normalizador, y comparar dicha dosis con un valor límite. Las aneuploidías cromosómicas se pueden determinar de acuerdo con el método de incluir trisomía 8, trisomía 13, trisomía 15, trisomía 16, trisomía 18, trisomía 21, trisomía 22, monosomía X, y XXX.

La muestra materna usada en las realizaciones del método para determinar la presencia o ausencia de una aneuploidía es un fluido biológico seleccionado de sangre, plasma, suero, orina y saliva. Preferiblemente, la muestra materna es una muestra de plasma. En algunas realizaciones, las moléculas de ácidos nucleicos comprendidas en la muestra materna son moléculas de ADN libre de células. En algunas realizaciones, los pasos consecutivos comprendidos en la preparación de la librería de secuenciación se realizan en menos de una hora. Preferiblemente, los pasos consecutivos se realizan en ausencia de polietilenglicol. Más preferiblemente, los pasos consecutivos excluyen la purificación. La secuenciación de la biblioteca de secuenciación se consigue por métodos de secuenciación de próxima generación (NGS). En algunas realizaciones, la secuenciación comprende una amplificación. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por ligadura. En todavía otras realizaciones, la secuenciación es secuenciación de moléculas individuales.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la

fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla; (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Opcionalmente, el método puede comprender adicionalmente determinar la presencia o ausencia de aneuploidía en la muestra materna.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla; (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Determinar la fracción comprende determinar el número de etiquetas de secuencia fetal y materna mapeadas a un genoma objetivo de referencia que comprende el al menos un ácido nucleico polimórfico. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido simple (SNP); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido simple (SNP); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Determinar la fracción comprende determinar el número de etiquetas de secuencia fetales y maternas mapeadas con un genoma objetivo de referencia que comprende el al menos un ácido nucleico polimórfico.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos una repetición corta en tándem (STR); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos una repetición corta en tándem (STR); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de

secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Determinar la fracción comprende determinar el número de etiquetas de secuencia fetales y maternas mapeadas con un genoma objetivo de referencia que comprende el al menos un ácido nucleico polimórfico. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

En otra realización, se divulga en la presente un uso de dicha biblioteca en un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternas, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido (SNP); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. En realizaciones en las que cada uno de la pluralidad de ácidos nucleicos objetivo polimórficos comprende al menso un polimorfismo de nucleótido simple (SNP), el SNP se selecciona de rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022. En realizaciones en las que cada uno de la pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido (SNP), el al menos un SNP es un SNP en tándem seleccionados de las parejas de SNP en tándem rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980-rs987981; rs4143392- rs4143391; rs1691324- rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 - rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

Se divulga en la presente un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos una repetición corta en tándem (STR); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. La al menos una STR se selecciona de CSF1PO, FGA, TH01, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, y D1GATA113. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

Se divulga en la presente un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido (SNP); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Determinar la fracción comprende determinar el número de etiquetas de secuencia fetales y maternas mapeadas con un genoma objetivo de referencia que comprende el al menos un ácido nucleico polimórfico. En realizaciones en las que el cada uno de la pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos un polimorfismo de nucleótido simples (SNP), el SNP se selecciona de rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022. En realizaciones en las

que cada uno de la pluralidad de los ácidos nucleicos objetivo polimórficos comprende al menso un polimorfismo de nucleótido (SNP), el menos un SNP es un SNP en tándem seleccionado de las parejas de SNP en tándem rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102- rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 -rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

Se divulga en la presente un método para determinar la fracción de moléculas de ácidos nucleicos fetales en una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende: (a) amplificar una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de la mezcla, en donde cada uno de dicha pluralidad de ácidos nucleicos objetivo polimórficos comprende al menos una repetición corta en tándem (STR); (b) preparar una biblioteca de secuenciación del producto amplificado obtenido en el paso (a) en donde preparar la biblioteca comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de dichas moléculas de ácidos nucleicos fetales y maternos; (c) secuenciar al menos una porción de la biblioteca de secuenciación; y (d) en base a dicha secuenciación, determinar la fracción de moléculas de ácidos nucleicos fetales. Determinar la fracción comprende determinar el número de etiquetas de secuencia fetales y maternas mapeadas con un genoma objetivo de referencia que comprende el al menos un ácido nucleico polimórfico. La al menos una STR se selecciona de CSF1PO, FGA, TH01, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, y D1GATA113. Opcionalmente, el método puede comprender además determinar la presencia o ausencia de aneuploidía en la muestra materna.

La muestra materna usada en las realizaciones del método para determinar la fracción de moléculas de ácidos nucleicos fetales, es un fluido biológico seleccionado de sangre, plasma, suero, orina y saliva. Preferiblemente, la muestra materna es una muestra de plasma. En algunas realizaciones, las moléculas de ácidos nucleicos comprendidas en la muestra materna son moléculas de ADN libre de células. En algunas realizaciones, los pasos consecutivos comprendidos en la preparación de la biblioteca de secuenciación se realizan en menos de una hora. Preferiblemente, los pasos consecutivos se realizan en ausencia de polietilenglicol. Más preferiblemente, los pasos consecutivos excluyen la purificación. La secuenciación de la biblioteca de secuenciación se consigue por métodos de secuenciación de próxima generación (NGS). En algunas realizaciones, la secuenciación comprende una amplificación. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por ligadura. En todavía otras realizaciones, la secuenciación es secuenciación de moléculas individuales.

4. BREVE DESCRIPCION DE LOS DIBUJOS

Las características nuevas de la invención se exponen con particularidad en las reivindicaciones añadidas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención con referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos acompañantes de los cuales:

La **Figura 1** es un diagrama de flujo de un método **100** para determinar la presencia o ausencia de una aneuploidía cromosómica en una muestra de ensayo que comprende una mezcla de ácidos nucleicos.

La **Figura 2** es un diagrama de flujo de un método **200** para determinar simultáneamente la presencia o ausencia de aneuploidía y la fracción fetal en una muestra de ensayo materna que comprende una mezcla de ácidos nucleicos fetales y maternos.

La **Figura 3** es un diagrama de flujo de un método **300** para determinar simultáneamente la presencia o ausencia de aneuploidía y la fracción fetal en una muestra de ensayo de plasma materno enriquecido por ácidos nucleicos polimórficos.

La **Figura 4** es un diagrama de flujo de un método **400** para determinar simultáneamente la presencia o ausencia de aneuploidía y la fracción fetal en una muestra de ensayo de ADN libre de células purificado materno que ha sido por ácidos nucleicos polimórficos.

La **Figura 5** es un diagrama de flujo de un método **500** para determinar simultáneamente la presencia o

ausencia de aneuploidía y la fracción fetal en una biblioteca de secuenciación construida de ácidos nucleicos fetales y maternos derivados de una muestra de ensayo materna y enriquecidos con ácidos nucleicos polimórficos.

La **Figura 6** es un diagrama de flujo de un método **600** para determinar la fracción fetal por secuenciación de una biblioteca de ácidos nucleicos objetivo polimórficos amplificados de una porción de una mezcla purificada de ácidos nucleicos fetales y maternos.

La **Figura 7** muestra electroferogramas de una biblioteca de secuenciación de ADN libre de células preparada de acuerdo con el protocolo abreviados descrito en el Ejemplo 2a (**A**), y el protocolo descrito en el Ejemplo 2b (**B**).

La **Figura 8** muestra en el eje Y la proporción del número de etiquetas de secuencia mapeadas para cada cromosoma (eje X) y el número total de etiquetas mapeadas para todos los cromosomas (1-22, X e Y) para la muestra M11281 cuando la biblioteca se preparó usando el protocolo abreviado del Ejemplo 2a (♦) y cuando se preparó de acuerdo con el protocolo de longitud completa del Ejemplo 2b (■). También se muestras las proporciones de etiquetas para la muestra M11297 obtenidas de la secuenciación de una biblioteca preparada de acuerdo con el protocolo abreviado del Ejemplo 2a (A) y de acuerdo con el protocolo de longitud completa del Ejemplo 2b (**X**).

La **Figura 9** muestra la distribución de la dosis de cromosoma para el cromosoma 21 determinada de la secuenciación de ADN libre de células extraído de un conjunto de 48 muestras de sangre obtenidas de sujetos humanos cada uno embarazado de un feto masculino o femenino. Las dosis de cromosoma 21 para las muestras de ensayo calificadas, es decir normales para el cromosoma 21 (O), y trisomía 21 se muestran (Δ) para los cromosomas 1-12 y X (**A**), y para los cromosomas 1-22 y X (**B**).

La **Figura 10** muestra la distribución de la dosis de cromosoma para el cromosoma 18 determinada de la secuenciación de ADN libre de células extraído de un conjunto de 48 muestras de sangre obtenidas de sujetos humanos cada uno embarazado de un feto masculino o femenino. Las dosis de cromosoma 18 para las muestras de ensayo calificadas, es decir normales para el cromosoma 18 (O), y trisomía 18 (Δ) se muestran para los cromosomas 1-12 y X (**A**), y para los cromosomas 1-22 y X (**B**).

La **Figura 11** muestra la distribución de la dosis de cromosoma para el cromosoma 13 determinada de la secuenciación de ADN libre de células extraído de un conjunto de 48 muestras de sangre obtenidas de sujetos humanos cada uno embarazado de un feto masculino o femenino. Las dosis de cromosoma 13 para las muestras de ensayo calificadas, es decir normales para el cromosoma 13 (O), y trisomía 13 (Δ) se muestran para los cromosomas 1-12 y X (**A**), y para los cromosomas 1-22 y X (**B**).

La **Figura 12** muestra la distribución de las dosis de cromosoma para el cromosoma X determinada de la secuenciación de ADN libre de células extraído de un conjunto de 48 muestras de sangre obtenidas de sujetos humanos cada uno embarazado de un feto masculino o femenino. Las dosis de cromosoma X para las muestras masculinas (46,XY; (O)), femeninas (46,XX(Δ)); monosomía X (45,X; (+)), y cariotipos complejos (Cplx (X)) se muestran para los cromosomas 1-12 y X (**A**) y para los cromosomas 1-22 y X (**B**).

La **Figura 13** muestra la distribución de las dosis de cromosoma para el cromosoma Y determinada de la secuenciación de ADN libre de células extraído de un conjunto de 48 muestras de sangre obtenidas de sujetos humanos cada uno embarazado de un feto masculino o femenino. Las dosis de cromosoma Y para las muestras masculinas (46,XY; (O)), femeninas (46,XX(Δ)); monosomía X (45,X; (+)), y cariotipos complejos (Cplx (X)) se muestran para los cromosomas 1-12 y X (**A**) y para los cromosomas 1-22 y X (**B**).

La **Figura 14** muestra el coeficiente de variación (CV) para los cromosomas 21 (■), 18 (●) y 13 (▲) que se determinó de las dosis mostradas en las **Figuras 9, 10** y **11**, respectivamente.

La **Figura 15** muestra el coeficiente de variación (CV) para los cromosomas X (■) e Y (●) que se determinó de las dosis mostradas en las **Figuras 12 y 13**, respectivamente.

La **Figura 16** muestra las dosis de secuencias (eje Y) para un segmento del cromosoma 11 (81000082-103000103bp) determinadas de la secuenciación de ADN libre de células extraído de un conjunto de 7 muestras calificadas (O) obtenidas y 1 muestra de ensayo (♦) de sujetos humanos embarazados. Se identificó una muestra de un sujeto que lleva un feto con una aneuploidía parcial del cromosoma 11 (♦)

La **Figura 17** muestra un gráfico de la proporción del número de etiquetas de secuencia mapeadas para cada cromosoma y el número total de etiquetas mapeadas para todos los cromosomas (1-22, X e Y) obtenidas de la secuenciación de una biblioteca de ADN libre de células no enriquecidas (●), y biblioteca de ADN libre de células enriquecidas con 5% (■) o 10% (♦) biblioteca de SNP multiplex amplificada.

La **Figura 18** muestra un diagrama de barras que muestra la identificación de secuencias polimórficas fetales y maternas (SNPs) usadas para determinar la fracción fetal en una muestra de ensayo. Se muestran el número total de lecturas de secuencia (eje Y) mapeadas para las secuencias SNP identificadas por números rs (eje X), y el nivel relativo de ácidos nucleicos fetales (*).

La **Figura 19** representa una realización del uso de la fracción fetal para determinar límites de corte para la detección de aneuploidía.

La **figura 20** ilustra la distribución de dosis de cromosoma normalizadas para el cromosoma 21 (A), cromosoma 18 (B), cromosoma 13 (C), cromosoma X (D) y cromosoma Y (E) en relación a la desviación estándar de la media (eje Y) para la dosis de cromosomas correspondiente en muestras no afectadas.

5. DESCRIPCION DETALLADA DE LA INVENCION

La invención se refiere a métodos para preparar una biblioteca de secuenciación y usos de la misma en métodos para determinar la presencia o ausencia de una aneuploidía, por ejemplo cromosómica o aneuploidía parcial, y/o fracción fetal en muestras maternas que comprenden ácidos nucleicos fetales y maternos por secuenciación masivamente paralela. El método comprende un nuevo protocolo para preparar bibliotecas de secuenciación que inesperadamente mejoran la calidad del ADN de la biblioteca a la vez que aceleran el proceso de análisis de muestras para diagnóstico prenatal. Los métodos también permiten determinar variaciones del número de copias (CNV) de cualquier secuencia de interés en una muestra de ensayo que comprende una mezcla de ácidos nucleicos que se sabe o son sospechosos de diferir en la cantidad de una i más secuencias de interés, y/o determinar la fracción de una de al menos dos poblaciones de ácidos nucleicos contribuidos a l muestra por diferentes genomas. Las secuencias de interés incluyen secuencias genómicas que varían de cientos de bases a decenas de megabases a cromosomas completos que se sabe o son sospechosos de estar asociados con una condición genética o una de enfermedad. Ejemplos de secuencias de interés incluyen cromosomas asociados con aneuploidías bien conocidas, por ejemplo trisomía 21, y segmentos de cromosomas que se multiplican en enfermedades como el cáncer, por ejemplo trisomía parcial 8 en leucemia mieloide aguda. El método comprende un enfoque estadístico que da cuenta de la variabilidad aguda derivada de procesos relacionados, variabilidad intercromosómica e inter-secuenciación. El método es aplicable para determinar el CNV de cualquier aneuploidía fetal y CNVs conocidos o sospechosos de estar asociados con una variedad de condiciones médicas.

A menos que se indique lo contrario, la práctica de la presente invención implica técnicas convencionales usadas comúnmente en biología molecular, microbiología, purificación de proteínas, diseño de proteínas, secuenciación de proteínas y ADN y campos de ADN recombinante, que están dentro de la técnica. Dichas técnicas son conocidas por los expertos en la técnica y se describen en numerosos textos estándar y trabajos de referencia.

Los intervalos numéricos incluyen los números que definen el intervalo. Se pretende que cada limitación numérica máxima dada a lo largo de esta especificación incluya cada limitación numérica inferior, como si dichas limitaciones numéricas inferiores estuvieran expresamente escritas en la presente. Cada limitación numérica mínima dad a lo largo de esta especificación incluirá cada limitación numérica más alta, como si dichas limitaciones numéricas más altas estuvieran escritas expresamente en la presente. Cada intervalo numérico dado a lo largo de esta especificación incluirá cada intervalo numérico más estrecho que caiga dentro de dicho intervalo numérico más amplio, como si dichos intervalos numéricos más estrechos estuvieran todos expresamente escritos en la presente.

Los encabezamientos proporcionados en la presente no son limitaciones de los varios aspectos o realizaciones de la invención que pueden ser tenidas con referencia a la Especificación como un todo. Por consiguiente, como se ha indicado anteriormente, los términos definidos inmediatamente a continuación se definen más completamente con referencia a la especificación como un todo.

A menos que se defina lo contrario en la presente, todos los términos técnicos y científicos usados en la presente tienen el mismo significado que el que entendería comúnmente alguien experto en la técnica a la que pertenece esta invención. Varios diccionarios científicos que incluyen los términos incluidos en la presente son bien conocidos y están disponibles para los expertos en la técnica. Aunque cualquier método y material similar o equivalente a los descritos en la presente encuentran uso en la práctica o ensayo de la presente invención, se describen algunos métodos y materiales preferidos. Por consiguiente, los términos definidos inmediatamente a continuación se describen más completamente con referencia a la Especificación como un todo. Se debe entender que esta invención no está limitada a la metodología, protocolos y reactivos particulares descritos, ya que estos pueden variar, dependiendo del contexto en el que se usen por los expertos en la técnica.

5.1 DEFINICIONES

Como se usa en la presente, los términos singulares "un", "uno" y "el" incluyen la referencia plural a menos que el contexto indique claramente lo contrario. A menos que se indique lo contrario, los ácidos nucleicos están escritos de izquierda a derecha en orientación 5' a 3' y las secuencias de aminoácidos están escritas de izquierda a derecha en orientación amino a carboxi, respectivamente.

El término "evaluar" en la presente se refiere a caracterizar el estado de una aneuploidía cromosómica por uno de los tres tipos de designaciones: "normal", "afectada" y "no designada". Por ejemplo, en presencia de trisomía la designación "normal" se determina por el valor de un parámetro, por ejemplo una dosis de cromosoma de ensayo que está por debajo de un límite definido por el usuario de fiabilidad, la designación "afectada" está determinada por un parámetro, por ejemplo una dosis de cromosoma de ensayo, que está por encima de un límite definido por el usuario de fiabilidad, y el resultado "no designada" se determina por un parámetro, por ejemplo una dosis de cromosoma de ensayo, que se encuentra entre los límites definidos por el usuario de fiabilidad para hacer una designación "normal" o una "afectada".

El término "variación del número de copias" en la presente se refiere a la variación en el número de copias de una secuencia de ácidos nucleicos que es de 1 kb o mayor presente en una muestra de ensayo en comparación con el número de copias de la secuencia de ácidos nucleicos presente en una muestra calificada. Una "variante del

número de copias" se refiere a la secuencia de 1kb o mayor del ácido nucleico en el que las diferencias del número de copias se encuentran por comparación de una secuencia de interés en la muestra de ensayo que se presenta en una muestra calificada. Las variantes/variaciones del número de copias incluyen deleciones, incluyendo microdeleciones, inserciones, incluyendo microinserciones, duplicaciones, multiplicaciones, inversiones, translocaciones y variantes multi-sitio complejas. El CNV abarca aneuploidías cromosómicas y aneuploidías parciales.

El término "aneuploidía" en la presente se refiere a un desequilibrio de material genético causado por una pérdida o ganancia de un cromosoma completo, o parte de un cromosoma.

El término "aneuploidía cromosómica" en la presente se refiere a un desequilibrio de material genético causado por una pérdida o ganancia de un cromosoma completo, e incluye aneuploidía de la línea germinal y aneuploidía en mosaico.

El término "aneuploidía parcial" en la presente se refiere a un desequilibrio de material genético causado por una pérdida o ganancia de parte de un cromosoma, por ejemplo, monosomía parcial y trisomía parcial, y abarca desequilibrios resultantes de translocaciones, deleciones e inserciones.

El término "pluralidad" se usa en la presente con referencia a un número de moléculas de ácidos nucleicos o etiquetas de secuencia que es suficiente para identificar diferencias significativas en variaciones del número de copias (por ejemplo dosis de cromosoma) en muestras de ensayo y muestras calificadas usando los métodos divulgados en la presente. En algunas realizaciones, se obtienen al menos alrededor de 3×10^6 etiquetas de secuencia, al menos alrededor de 5×10^6 etiquetas de secuencia, al menos alrededor de 8×10^6 etiquetas de secuencia, al menos alrededor de 10×10^6 etiquetas de secuencia, al menos alrededor de 15×10^6 etiquetas de secuencia, al menos alrededor de 20×10^6 etiquetas de secuencia, al menos alrededor de 30×10^6 etiquetas de secuencia, al menos alrededor de405×10^6 etiquetas de secuencia, al menos alrededor de 50×10^6 etiquetas de secuencia que comprenden lecturas entre de 20 y 40bp para cada muestra de ensayo.

El término "polinucleótido", "ácido nucleico" y "moléculas de ácidos nucleicos" se usan de manera intercambiable y se refieren a secuencias ligadas covalentemente de nucleótidos (es decir, ribonucleótidos para ARN y desoxiribonucleótidos para ADN) en las que la posición 3' de la pentosa de un nucleótido está unida por un grupo fosfodiéster a la posición 5' de la pentosa de la siguiente, incluye secuencias de cualquier forma de ácido nucleico incluyendo, pero no limitado a, moléculas de ARN, ADN y ADN libre de células. El término "polinucleótido" incluye, sin limitación, polinucleótidos de cadena sencilla y doble.

El término "porción" se usa en la presente en referencia a la cantidad de información de secuencia de moléculas de ácidos nucleicos fetales y maternos en una muestra biológica que en suma ascienden a menos de la información de la secuencia de <1 genoma humano.

El término "muestra de ensayo" en la presente se refiere a una muestra que comprende una mezcla de ácidos nucleicos que comprende al menos una secuencia de ácidos nucleicos cuyo número de copias se sospecha que ha sufrido variación. Los ácidos nucleicos presentes en la muestra de ensayo son referidos como "ácidos nucleicos de ensayo".

El término "muestra calificada" en la presente se refiere a una muestra que comprende una mezcla de ácidos nucleicos que están presentes en un número de copias conocido con los que los ácidos nucleicos en la muestra de ensayo se comparan, y es una muestra que es normal, es decir, no aneuploide, para la secuencia de interés, por ejemplo una muestra calificada usada para identificar un cromosoma normalizador para el cromosoma 21 es una muestra que no es una muestra de trisomía 21.

El término "acido nucleico calificado" se usa de manera intercambiable con "secuencia calificada" es una secuencia contra la que se compara la cantidad de una secuencia de ensayo o ácido nucleico de ensayo. Una secuencia calificada es una presente en una muestra biológica preferiblemente en una representación conocida, es decir, la cantidad de una secuencia calificada se conoce. Una "secuencia calificada de interés" es una secuencia calificada para la que la cantidad se conoce en una muestra calificada, y es una secuencia que está asociada con una diferencia en la representación de secuencia en un individuo con una condición médica.

El término "secuencia de interés" en la presente se refiere a una secuencia de ácidos nucleicos que está asociada con una diferencia en la representación de secuencia en individuos sanos frente a enfermos. Una secuencia de interés puede ser una secuencia en un cromosoma que está tergiversada, es decir, sobre- o sub-representado, en una enfermedad o condición genética. Una secuencia de interés puede también ser una porción de un cromosoma, o un cromosoma. Por ejemplo, una secuencia de interés puede ser un cromosoma que está sobre-representado en una condición de aneuploidía, o in gen que codifica un supresor tumoral que está sub-representado en un cáncer. Las secuencias de interés incluyen secuencias que estas sobre- o sub-representadas en la población total, o una subpoblación de células de un sujeto. Una "secuencia calificada de interés" es una secuencia de interés

en una muestra calificada. Una "secuencia de ensayo de interés" es una secuencia de interés en una muestra de ensayo.

El término "secuencia normalizadora" en la presente se refiere a una secuencia que muestra una variabilidad en el número de etiquetas de secuencia que están mapeadas para ella entre las muestras y ejecuciones de secuenciación que mejor se aproxima a la de la secuencia de interés para la que se usa como un parámetro normalizador, y que puede diferenciar mejor una muestra afectada de una o más muestras no afectadas. Un "cromosoma normalizador" es un ejemplo de "secuencia normalizadora".

El término "diferenciabilidad" en la presente se refiere a la característica de un cromosoma normalizador que permite distinguir una o más muestras no afectadas, es decir, normales de una o más muestras afectadas, es decir, aneuploides.

El término "dosis de secuencia" en la presente se refiere a un parámetro que relaciona la densidad de etiqueta de secuencia de una secuencia de interés con la densidad de etiqueta de una secuencia normalizadora. Un a "dosis de secuencia de ensayo" es un parámetro que relaciona la densidad de etiqueta de secuencia de una secuencia de interés, por ejemplo cromosoma 21, con la de una secuencia normalizadora, por ejemplo cromosoma 9, determinada en una muestra de ensayo. De manera similar, una "dosis de secuencia calificada" es un parámetro que relaciona la densidad de etiqueta de secuencia de una secuencia de interés con la de una secuencia normalizadora determinada en una muestra calificada.

El término "densidad de etiqueta de secuencia" en la presente se refiere al número de lecturas de secuencia que se mapean para una secuencia del genoma de referencia, por ejemplo la densidad de etiqueta de secuencia para el cromosoma 21 es el número de lecturas de secuencia generadas por el método de secuenciación que se mapean para el cromosoma 21 del genoma de referencia. El término "proporción de densidad de etiqueta de secuencia" en la presente se refiere a la proporción del número de etiquetas de secuencia que se mapean para un cromosoma del genoma de referencia, por ejemplo cromosoma 21, para la longitud del cromosoma 21 del genoma de referencia.

El término "parámetro" en la presente se refiere a un valor numérico que caracteriza un conjunto de datos cuantitativos y/o una relación numérica entre conjuntos de datos cuantitativos. Por ejemplo, una proporción (o función de una proporción) entre el número de etiquetas de secuencia mapeadas para un cromosoma y la longitud del cromosoma para el que las etiquetas se mapean, es un parámetro.

El término "valor límite" y "valor límite calificado" en la presente se refieren a cualquier número que se calcula usando un conjunto de datos de calificación y sirve como un límite de diagnóstico de una variación del número de copias, por ejemplo una aneuploidía, en un organismo. Si el límite se excede por los resultados obtenidos de la práctica de la invención, un sujeto puede ser diagnosticado con una variación del número de copias, por ejemplo trisomía 21.

El término "lectura" se refiere a una secuencia de ADN de longitud suficiente (por ejemplo, al menos alrededor de 30 bp) que se puede usar para identificar una secuencia o región más grande, por ejemplo, que puede ser alineada y asignada específicamente a un cromosoma o región genómica o gen.

El término "etiqueta de secuencia" se usa en la presente de manera intercambiable con el término "etiqueta de secuencia mapeada" para referirse a una lectura de secuencia que ha sido asignada específicamente, es decir mapeada, a una secuencia más grande, por ejemplo un genoma de referencia, por alineamiento. Las etiquetas de secuencia mapeadas son mapeadas únicamente para un genoma de referencia, es decir son asignadas a una localización única para el genoma de referencia. Las etiquetas que pueden ser mapeadas para más de una localización en un genoma de referencia, es decir etiquetas que no mapean únicamente, no se incluyen en el análisis.

Como se usan en la presente, los términos "alineada", alineamiento" o "alinear" se refieren a una o más secuencias que se identifican como una equivalencia en términos del orden de sus moléculas de ácidos nucleicos para una secuencia conocida de un genoma de referencia. Dicha alineamiento se puede hacer manualmente o por un algoritmo informático, los ejemplos incluyendo el programa de ordenador Efficient Local Alignment of Nucleotide Data (ELAND) distribuido como parte de la línea Illumina Genomics Analysis. La equivalencia de una lectura de secuencia en la alineamiento puede ser un 100% de equivalencia de secuencia o menor del 100% (equivalencia no perfecta).

Como se usa en la presente, el término "genoma de referencia" se refiere a cualquier secuencia del genoma conocida particular, ya sea parcial o completa, de cualquier organismo o virus que pueda usarse para referenciar secuencias identificadas de un sujeto. Por ejemplo, un genoma de referencia usado para sujetos humanos así como muchos otros organismos se encuentra en el National Center for Biotechnology Information en www.ncbi.nlm.nih.gov. Un "genoma" se refiere a la información genética completa de un organismo o virus,

expresada en secuencias de ácidos nucleicos.

Los términos "genoma de secuencias objetivo artificial" y "genoma de referencia artificial" en la presente se refiere a una agrupación de secuencias conocidas que abarca alelos de sitios polimórficos conocidos. Por ejemplo, un "genoma de referencia SNP" es un genoma de secuencias objetivo artificial que comprende una agrupación de secuencias que abarca alelos de SNPs conocidas.

El término "secuencia clínicamente relevante" en la presente se refiere a una secuencia de ácidos nucleicos que se sabe o es sospechosa de estar asociada o implicada con una condición genético o enfermedad. Determinar la ausencia o presencia de una secuencia clínicamente relevante puede ser útil para determinar un diagnóstico o confirmar un diagnóstico de una condición médica, o proporcionar un pronóstico para el desarrollo de una enfermedad.

El término "derivado" cuando se usa en el contexto de un ácido nucleico o mezcla de ácidos nucleicos, en la presente se refiere al medio por el que se obtiene el ácido(s) nucleico de una fuente de la que es originario. Por ejemplo, en una realización, una mezcla de ácidos nucleicos que se deriva de dos genomas diferentes significa que los ácidos nucleicos, por ejemplo ADN libre de células, fueron liberados naturalmente por las células a través de procesos de origen natural como necrosis o apoptosis. En otra realización, una mezcla de ácidos nucleicos que se deriva de dos genomas diferentes significa que los ácidos nucleicos se extrajeron de dos tipos diferentes de células de un sujeto.

El término "muestra mixta" en la presente se refiere a una muestra que contiene una mezcla de ácidos nucleicos, que se derivan de genomas diferentes.

El término "muestra materna" en la presente se refiere a una muestra biológica obtenida de un sujeto embarazado, por ejemplo una mujer.

El término "muestra materna original" en la presente se refiere a una muestra biológica obtenida de un sujeto embarazado, por ejemplo una mujer, que sirve como la fuente de la que se elimina una porción para amplificar ácidos nucleicos objetivo polimórficos. La "muestra original" puede ser cualquier muestra obtenida de un sujeto embarazado, y las fracciones procesadas de la misma, por ejemplo una muestra de ADN libre de células extraída de una muestra de plasma materno.

El término "fluido biológico" en la presente se refiere a un líquido tomado de una fuente biológica e incluye, por ejemplo, sangre, suero, plasma, esputo, líquido de lavado, fluido cerebroespinal, orina, semen, sudor, lágrimas, saliva y similares. Como se usa en la presente, los términos "sangre", "plasma" y "suero" abarcan expresamente fracciones o porciones procesadas de los mismos. De manera similar, cuando una muestra se toma de una biopsia, hisopo, frotis, etc., la "muestra" abarca expresamente una fracción procesada o porción derivada de la biopsia, hisopo, frotis, etc.

Los términos "ácidos nucleicos maternos" y "ácidos nucleicos fetales" en la presente se refieren a los ácidos nucleicos de un sujeto femenino embarazado y los ácidos nucleicos del feto que es llevado por la mujer embarazada, respectivamente.

Como se usa en la presente, el término "correspondiente a" se refiere a una secuencia de ácidos nucleicos, por ejemplo un gen o un cromosoma, que está presente en el genoma de diferentes sujetos, y que no tiene necesariamente la misma secuencia en todos los genomas, pero sirve para proporcionar la identidad en lugar de la información genética de una secuencia de interés, por ejemplo un gen o un cromosoma.

Como se usa en la presente, el término "sustancialmente libre de células" abarca preparaciones de la muestra deseada de la que los componentes que están normalmente asociados con él se han eliminado. Por ejemplo, una muestra de plasma se vuelve esencialmente libre de células eliminando células sanguíneas, por ejemplo, glóbulos rojos, que están normalmente asociados con ella. En algunas realizaciones, se procesan muestras sustancialmente libres para eliminar células que contribuirían de otra manera al material genético deseado que se va a probar apara un CNV.

Como se usa en la presente, el término "fracción fetal" se refiere a la fracción de ácidos nucleicos fetales presentes en una muestra que comprende ácido nucleico fetal y materno.

Como se usa en la presente el término "cromosoma" se refiere al portador del gen que lleva la herencia de una célula viva que se deriva de cromatina y que comprende ADN y componentes de proteína (especialmente histonas). El sistema de numeración de cromosomas del genoma humano individuales internacionalmente reconocido convencional se usa en la presente.

Como se usa en la presente, el término "longitud de polinucleótido" se refiere al número absoluto de

moléculas de ácidos nucleicos (nucleótidos) en una secuencia o en una región de un genoma de referencia. El término "longitud de cromosoma" se refiere a la longitud conocida del cromosoma dada en pares de base, por ejemplo proporcionada en el montaje NCBI36/hg18 del cromosoma humano encontrado en la red mundial en genome.ucsc.edu/cgi-bin/hgTracks?hgsid=167155613 &chromInfoPage=

El término "sujeto" en la presente se refiere a un sujeto humano así como a un sujeto no humano como un mamífero, un invertebrado, un vertebrado, un hongo, una levadura, una bacteria y un virus. Aunque los ejemplos en la presente se refieren a humanos y el lenguaje está dirigido principalmente a preocupaciones humanas, el concepto de esta invención es aplicable a genomas de cualquier planta o animal, y es útil en los campos de medicina veterinaria, ciencias animales, laboratorios de investigación y demás.

El término "condición" en la presente se refiere a "condición médica" como un término amplio que incluye todas las enfermedades y trastornos, pero que también puede incluir lesiones y situaciones de salud normales, como el embarazo, que pueden afectar a la salud de una persona, beneficiarse de la asistencia médica, o tener implicaciones para tratamientos médicos.

El término "cromosoma aneuploide" en la presente se refiere a un cromosoma que está implicado en una aneuploidía.

El término "aneuploidía" en la presente se refiere a un desequilibrio de material genético causado por una pérdida o una ganancia de un cromosoma completo, o parte de un cromosoma.

Los términos "biblioteca" y "biblioteca de secuenciación" en la presente se refieren a una colección o pluralidad de moléculas plantilla que comparten secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'.

Los términos "terminación roma" y "reparación de extremos" se usan en la presente de manera intercambiable para referirse a un proceso enzimático que resulta en que ambas cadenas de una molécula de ADN de cadena doble termine en un par de bases, y no incluye purificar los productos de extremos romos del enzima de terminación roma.

El término "adición de colas de d-A" en la presente se refiere a un proceso enzimático que añade al menos una base de adenina al extremo 3' del ADN, y no incluye purificar el producto de la adición de colas de d-A de la enzima de la adición de colas de d-A.

El término "ligado por adaptadores" en la presente se refiere a un proceso que liga una secuencia adaptadora de ADN a fragmentos de ADN, y no incluye purificar los productos ligados de adaptadores de la enzima de ligamiento.

El término ´"recipiente de reacción" en la presente se refiere a un contenedor de cualquier forma, tamaño, capacidad o material que pueda usarse para procesar una muestra durante un procedimiento de laboratorio, por ejemplo investigación o clínico.

El término "pasos consecutivos" se usa en la presente en referencia a los pasos enzimáticos consecutivos de terminación roma, adición de colas de dA y ligado por adaptadores de ADN que no que no están interpuestos por pasos de purificación.

Como se usa en la presente, el término "purificado" se refiera al material (por ejemplo un polinucleótido aislado) que está relativamente en un estado puro, por ejemplo, al menos alrededor del 80% puro, al menos alrededor del 85% puro , al menos alrededor del 90% puro, al menos alrededor del 95% puro, al menos alrededor del 98% puro o incluso alrededor al menos alrededor del 99% puro.

Los términos "extraído", "recuperado", "aislado" y "separado" se refieren a un compuesto, proteína, célula, ácido nucleico o aminoácido que es eliminado de al menos un componente con el que está naturalmente asociado y encontrado en la naturaleza.

El término "SNPs en tándem" en la presente se refiere a dos o más SNPs que están presentes dentro de una secuencia de ácidos nucleicos objetivo polimórficos.

Los términos "ácido nucleico objetivo polimórfico", "secuencia polimórfica", "secuencia de ácidos nucleicos objetivo polimórfica" y "ácido nucleico polimórfico" se usan de manera intercambiable en la presente para referirse a una secuencia de ácidos nucleicos, por ejemplo una secuencia de ADN, que comprende uno o más sitios polimórficos.

El término "sitio polimórfico" en la presente se refiere a un polimorfismo de nucleótido simple (SNP) (SNP),

una deleción o inserción multi-base a pequeña escala, un Polimorfismo Multi-Nucleótido (MNP) o una Repetición Corta en Tándem (STR).

El término "pluralidad de ácidos nucleicos objetivo polimórficos" en la presente se refiere a un número de secuencias de ácidos nucleicos que comprende cada una al menos un sitio polimórfico de tal forma que al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40 o más sitios polimórficos diferentes se amplifican de los ácidos nucleicos objetivo polimórficos para identificar y/o cuantificar alelos fetales presentes en las muestras maternas que comprenden ácidos nucleicos fetales y maternos.

El término "enriquecer" en la presente se refiere al proceso de amplificar ácidos nucleicos objetivo polimórficos contenidos en una porción de una muestra materna, y combinar el producto amplificado con el resto de la muestra materna de la que se obtuvo la porción.

El término "densidad de etiqueta de secuencia" en la presente se refiere al número de lecturas de secuencia que se mapean para una secuencia del genoma de referencia, por ejemplo la densidad de etiqueta de secuencia para el cromosoma 21 es el número de lecturas de secuencia generadas por el método de secuenciación que se mapean para el cromosoma 21 del genoma de referencia. El término "proporción de densidad de etiqueta de secuencia" en la presente se refiere a la proporción del número de etiquetas de secuencia que se mapean para un cromosoma del genoma de referencia, por ejemplo cromosoma 21, para la longitud del cromosoma 21 del genoma de referencia.

Como se usa en la presente, el término "amplificación en fase sólida" como se usa en la presente se refiere a cualquier reacción de amplificación de ácidos nucleicos llevada a cabo en o en asociación con un soporte sólido de tal manera que todos o una porción de los productos amplificados se inmovilizan en el soporte sólido a medida que se forman. En particular, el término abarca la acción en cadena de la polimerasa en fase sólida (PCR en fase sólida) y al amplificación isotérmica en fase sólida que son reacciones análogas a la amplificación en fase de solución estándar, excepto en que uno o ambos de los cebadores de amplificación directo o inverso es/son inmovilizados en el soporte sólido. La PCR en fase sólida cubre sistemas como emulsiones, en donde un cebador se ancla a una microesfera el otro está en solución libre, y la formación de colonias en matrices de gel en fase sólida en donde un cebador está anclado en la superficie, y uno está en la solución libre. El término fase sólida, o superficie, se usa para insinuar o una matriz plana en donde los cebadores están unidos a una superficie plana, por ejemplo vidrio, sílice o portaobjetos de plástico o dispositivos celulares de flujo similares; microesferas, en donde uno o dos cebadores están unidos a las microesferas y las microesferas se amplifican; o una matriz de microesferas en una superficie después de que se han amplificado las microesferas.

Como se usa en la presente, el término "grupo de cromosomas" en la presente se refiere a un grupo de dos o más cromosomas.

Un "polimorfismo de nucleótido simple" (SNP) tiene lugar en un sitio polimórfico ocupado por un único nucleótido, que es el sitio de variación entre secuencias alélicas. El sitio es habitualmente precedido por y seguido por secuencias altamente conservadas del alelo (por ejemplo, secuencias que varían en menos de 1/100 ó 1/1000 miembros de las poblaciones). Un SNO surge habitualmente debido a la sustitución de un nucleótido por otro en el sitio polimórfico. Una transición es el reemplazo de una purina por otra purina o una pirimidina por otra pirimidina. Una transversión es el remplazo de una purina por una pirimidina o viceversa. Los SNPs pueden también surgir de una deleción de un nucleótido o una inserción de un nucleótido en relación a un alelo de referencia. Los polimorfismos de nucleótidos simples (SNPs) son posiciones en las que tienen lugar dos bases alternativas en frecuencia apreciable (>1%) en la población humana, y son el tipo más común de variación genética humana.

Como se usa en la presente, el término "repetición corta en tándem" o "STR" como se usa en la presente se refiere a una clase de polimorfismos que tienen lugar cuando un patrón de dos o más nucleótidos se repite y las secuencias repetidas son directamente adyacentes entre sí. El patrón puede variar en longitud de 2 a 10 pares de bases (bp) (por ejemplo (CATG)n en una región genómica) y es típicamente la región del intrón no codificante. Examinado varios loci de STR y contando cuantas repeticiones de una secuencia de STR específica hay en locus dado, es posible crear un perfil genético único de un individuo.

Como se usa en la presente, el término "miniSTR" en la presente se refiere a una repetición en tándem de cuatro o más pares de base que abarcan menso de alrededor de 300 pares de base, menos de alrededor de 250 pares de bases, menos de alrededor de 200 pares de bases, menos de alrededor de 150 pares de bases, menos de alrededor de 100 pares de bases, menos de alrededor de 50 pares de bases, o menos de alrededor de 25 pares de bases. Los "miniSTRs" son STRs que son amplificables de plantillas de ADN libre de células.

El término "SNPs en tándem" en la presente se refiere a dos o más SNPs que están presentes dentro de una secuencia de ácidos nucleicos objetivo polimórficos.

Como se usa en la presente, el término "Biblioteca enriquecida" en la presente se refiere a una biblioteca de

secuenciación que comprende secuencias de ácidos nucleicos objetivo polimórficos. Un ejemplo de una biblioteca enriquecida es una biblioteca de secuenciación que comprende secuencias de ADN libre de células de origen natural y secuencias de ácidos nucleicos objetivo amplificadas. Una "biblioteca no enriquecida" en la presente se refiere a una biblioteca de secuenciación que no comprende, es decir, una biblioteca generada de secuencias de ADN libre de células de origen natural. Una "biblioteca de ácidos nucleicos objetivo polimórficos" es una biblioteca generada de ácidos nucleicos objetivo amplificados.

Como se usa en la presente, el término "secuencias de ADN libre de células de origen natural" se refiere a fragmentos de ADN libre de células como están presentes en la muestra, y en contraste a fragmentos de ADN genómico que se obtienen por métodos de fragmentación descritos en la presente.

5.2 DESCRIPCION

La invención se refiere a métodos para generar una biblioteca de secuenciación de una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, y usos de la biblioteca en métodos para determinar la presencia o ausencia de una aneuploidía, por ejemplo cromosómica o aneuploidía parcial, y/o fracción fetal en muestras maternas que comprenden ácidos nucleicos fetales y maternos por secuenciación masivamente paralela. El método comprende un nuevo protocolo para preparar bibliotecas de secuenciación que inesperadamente mejoran la calidad del ADN de la biblioteca a l vez que aceleran el proceso de análisis de muestras para diagnóstico prenatal. Los métodos permiten determinar variaciones del número de copias (CNV) de cualquier secuencia de interés en una muestra de ensayo que comprende una mezcla de ácidos nucleicos que se sabe o se sospecha difieren en la cantidad de una o más secuencias de interés, y/o determinar la fracción de una de las dos poblaciones de ácidos nucleicos contribuidos a la muestra por genomas diferentes.

Métodos de Secuenciación

En una realización, el método descrito en la presente emplea tecnología de secuenciación de próxima generación (NGS) en la que se secuencian plantillas de ADN amplificadas clonalmente o moléculas de ADN individuales de una forma masivamente paralela dentro de una célula de flujo (por ejemplo como se describe en Volkerding et al. Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 11:31-46[2010]). Además de la información de la secuencia de alto rendimiento, la NGS proporciona información cuantitativa digital, en la que cada lectura de secuencia es una "etiqueta de secuencia" que se puede contar que representa una plantilla de ADN clonal individual o una molécula de ADN individual. Esta cuantificación permite al NGS expandir el concepto de PCR digital de contar moléculas de ADN libre de células (Fan et al., Proc Natl Acad Sci U S A 105:16266-16271 [2008]; Chiu et al., Proc Natl Acad Sci U S A 2008;105:20458-20463 [2008]). Las tecnologías de secuenciación de NGS incluyen pirosecuenciación, secuenciación por síntesis con terminadores de colorante reversibles, secuenciación por ligadura de sonda de oligonucleótidos y secuenciación en tiempo real.

Algunas de las tecnologías de secuenciación están disponibles comercialmente, como la plataforma de secuenciación por hibridación de Affymetrix Inc. (Sunnyvale, CA) y las plataformas de secuenciación por síntesis de Life Sciences (Bradford, CT), Illumina/Solexa (Hayward, CA) y Helicos Biosciences (Cambridge, MA), y la plataforma de ligadura por secuenciación de Applied Biosystems (Foster City, CA), como se describe a continuación. Además de la secuenciación de moléculas individuales realizada usando la secuenciación por síntesis de Helicos Biosciences, se abarcan otras tecnologías de secuenciación de moléculas individuales por el método de la invención e incluyen la tecnología SMRT™ de Pacific Biosciences, la tecnología Ion Torrent™, y la secuenciación por nanoporos que está siendo desarrollada por, por ejemplo Oxford Nanopore Technologies.

Aunque el método Sanger automatizado se considera como una tecnología de "primera generación", la secuenciación Sanger que incluye secuenciación Sanger automatizada, también se puede emplear por el método de la invención. Los métodos de secuenciación adicionales que comprenden el uso de desarrollar tecnologías de formación de imágenes de ácidos nucleicos, por ejemplo microscopía de fuerza atómica (AFM) o microscopía electrónica de transmisión (TEM), también se abarcan por el método de la invención. Las tecnologías de secuenciación ejemplares se describen a continuación.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es la Helicos True Single Molecule Sequencing (tSMS) (por ejemplo, como se describe en Harris T.D. et al., Science 320:106-109 [2008]). En la técnica tSMS, una muestra de ADN se escinde en cadenas de aproximadamente 100 a 200 nucleótidos, y se añade una secuencia polyA al extremo 3' de cada cadena de ADN. Cada cadena es etiquetada por la adición de un nucleótido de adenosina etiquetado fluorescentemente. Las cadenas de ADN son después hibridadas a una célula de flujo, que contiene millones de sitios de captura de oligo-T que se inmovilizan en la superficie de la célula de flujo. Las plantillas pueden estar a una densidad de alrededor de 100 millones de plantillas/cm^2. La célula de flujo se carga entonces en un instrumento, por ejemplo, secuenciador HeliScope™, y un laser ilumina la superficie de la célula de flujo, revelando la posición de cada plantilla. Una cámara CCD puede mapear la posición de las plantillas en la superficie de la célula de flujo. La etiqueta fluorescente de la plantilla es después escindida y se hace desaparecer. La reacción de secuenciación empieza introduciendo una polimerasas de

ADN y un nucleótido etiquetado por fluorescencia. El ácido nucleico oligo-T sirve como un cebador. La polimerasa incorpora los nucleótidos etiquetados al cebador de una manera dirigida a la plantilla. La polimerasa y los nucleótidos no incorporados se eliminan. Las plantillas que tienen incorporación directa del nucleótido etiquetado por fluorescencia se distinguen por formación de imágenes de la superficie de la célula de flujo. Después de la formación de imágenes, un paso de escindido elimina la etiqueta fluorescente, y el proceso se repite con otros nucleótidos etiquetados por fluorescencia hasta que se consigue la longitud de lectura deseada. La información de la secuencia se recoge con cada paso de adición de nucleótidos.

En una realización, la tecnología de secuenciación que se usa en el método de la invención es la secuenciación 454 (Roche) por ejemplo como se describe en Margulies, M. et al. Nature 437:376-380 [2005]). La secuenciación 454 supone dos pasos. En el primer paso, el ADN es cortado en fragmentos de aproximadamente 300-800 pares de base, y los fragmentos son terminados romos. Los adaptadores de oligonucleótidos son entonces ligados a los extremos de los fragmentos. Los adaptadores sirven como cebadores para la amplificación y secuenciación de los fragmentos. Los fragmentos pueden ser unidos a microesferas de captura de ADN, por ejemplo microesferas recubiertas de estreptavidina usando, por ejemplo Adaptador B, que contiene etiqueta de biotina 5'. Los fragmentos unidos a las microesferas se amplifican por PCR con gotitas de una emulsión de aceite-agua. El resultado son copias múltiples de fragmentos de ADN clonalmente amplificados en cada microesfera. En el segundo paso, las microesferas se capturan en pocillos (de tamaño de pico-litros). La pirosecuenciación se realiza en cada fragmento de ADN en paralelo. La adición de uno o más nucleótidos genera una señal de luz que se registra por una cámara CCD en un instrumento de secuenciación. La fuerza de la señal es proporcional al número de nucleótidos incorporado. La polisecuenciación hace uso de pirofosfato (PPi) que es liberado en el momento de la adición de nucleótidos. El PPi se convierte a ATP por ATP sulfurilasa en presencia de adenosina 5' fosfosulfato. La luciferasa usa ATP para convertir luciferina en oxiluciferina, y esta reacción genera luz que es distinguida y analizada.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es la tecnología SOLiD™ (Applied Biosystems). En la secuenciación por ligadura SOLiD™, el ADN genómico es cortado en fragmentos, y los adaptadores se unen a los extremos 5' y 3' de los fragmentos para generar una biblioteca de fragmentos. Alternativamente, se pueden introducir adaptadores internos ligando adaptadores a los extremos 5' y 3' de los fragmentos, circularizando los fragmentos, digiriendo el fragmento circularizado para generar un adaptador interno, y uniendo los adaptadores a los extremos 5' y 3' de los fragmentos resultantes para generar una biblioteca de compañeros emparejadas. Después, se preparan las poblaciones de microesferas clonales en microreactores que contienen microesferas, cebadores, plantilla y componentes de PCR. Después del PCR, las plantillas se desnaturalizan y las microesferas se enriquecen para separar las microesferas con plantillas extendidas. Las plantillas en las microesferas seleccionadas se someten a una modificación 3' que permite el enlace con un portaobjetos de vidrio. La secuencia se puede determinar por hibridación secuencial y ligadura de oligonucleótidos parcialmente aleatorios con una base central determinada (o pares de bases) que se identifica por un fluoróforo específico. Después de que se registra un color, el oligonucleótido ligado es escindido y eliminado y el proceso es después repetido.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es la tecnología de secuenciación de molécula individual, en tiempo real (SMRT™) de Pacific Biosciences. En la secuenciación SMRT, la incorporación continua de nucleótidos marcados con colorante se forma en imágenes durante la síntesis del ADN. Las moléculas de polimerasa de ADN individual se unen a la superficie inferior de los identificadores de longitud de onda de modo cero individuales (identificadores ZMW) que obtienen información de la secuencia a la vez que los nucleótidos fosfoligados están siendo incorporados en la cadena del cebador creciente. Un ZMW es una estructura de confinamiento que permite la observación de incorporación de un nucleótido individual por ADN polimerasa contra el fondo de los nucleótidos fluorescentes que se difunden rápidamente dentro y fuera del ZMW (en microsegundos). Toma varios milisegundos el incorporar un nucleótido en una cadena creciente. Durante este tiempo, la etiqueta fluorescente se excita y produce una señal fluorescente, y la etiqueta fluorescente se escinde. La identificación de la fluorescencia correspondiente del colorante indica que base se incorporó. El proceso se repite.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es secuenciación por nanoporos (por ejemplo, como se describe en Soni GV y Meller A. Clin Chem 53: 1996-2001 [2007]). Las técnicas de análisis de ADN de secuenciación por nanoporos se están desarrollando industrialmente por una variedad de compañías, incluyendo Oxford Nanopore Technologies)Oxford, Reino Unido). La secuenciación por nanoporos es una tecnología de secuenciación de una sola molécula por la que una única molécula de ADN se secuencia directamente a medida que pasa a través de un nanoporo. Un nanoporo es un agujero pequeño, del orden de 1 nanómetro de diámetro. La inmersión de un nanoporo en un fluido conductor y la aplicación de un potencial (voltaje) a través de él resulta en una ligera corriente eléctrica debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño y forma del nanoporo. A medida que una molécula de ADN pasa a través de un nanoporo, cada nucleótido en la molécula de ADN obstruye el nanoporo a un grado diferente, cambiando la magnitud de la corriente a través del nanoporo en grados diferentes. Así, este cambio en la corriente a medida que la molécula de ADN pasa a través del nanoporo representa una lectura de la secuencia de ADN.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es la matriz de transistor de efecto de campo sensible a sustancias químicas (chemFET) (por ejemplo, como se describe en la Publicación de Solicitud de Patente U.S. Nº 20090026082). En un ejemplo de la técnica, las moléculas de ADN pueden ser colocadas en cámaras de reacción, y las moléculas de la plantilla pueden ser hibridadas a un cebador de secuenciación enlazado a una polimerasa. La incorporación de uno o más trifosfatos en una nueva cadena de ácidos nucleicos en el extremo 3' del cebador de secuenciación puede distinguirse por un cambio en la corriente por un chemFET. Una matriz puede tener múltiples sensores de chemFET. En otro ejemplo, los ácidos nucleicos individuales pueden estar unidos a microesferas, y los ácidos nucleicos pueden amplificarse en la microesfera, y las microesferas individuales pueden transferirse a cámaras de reacción individuales en una matriz de chemFET, con cada cámara teniendo un sensor de chemFET, y los ácidos nucleicos se pueden secuenciar.

En una realización, la tecnología de secuenciación de ADN que se usa en el método de la invención es el método de Halcyon Molecular que usa microscopía electrónica de transmisión (TEM). El método denominado Nano Transferencia Rápida de Colocación de Moléculas Individuales (IMPRNT), comprende utilizar formación de imágenes por microscopía electrónica de transmisión de resolución de átomos individuales de ADN de alto peso molecular (150kb o mayor) etiquetado selectivamente con marcadores de átomos pesados y organizar estas moléculas en películas ultra finas en matrices paralelas ultra-densas (3nm de cadena a cadena) con separación base a base consistente. El microscopio electrónico se usa para formar imágenes de las moléculas en las películas para determinar la posición de marcadores de átomos pesados y para extraer información de la secuencias de bases del ADN. El método se describe adicionalmente en la publicación de patente de PCT WO2009/046445. El método permite secuenciar completamente genomas humanos en menos de diez minutos.

En una realización, la tecnología de secuenciación de ADN es la secuenciación de moléculas individuales Ion Torrent, que empareja tecnología de semiconductores con una química de secuenciación simple para traducir directamente información codificada químicamente (A, C, G, T) en información digital (0, 1) en un chip semiconductor. En la naturaleza, cuando se incorpora un nucleótido en una cadena de ADN por una polimerasa, se libera un ion de hidrógeno como un subproducto. La Ion Torrent usa una matriz de alta densidad de pocillos de micro-mecanizado para realizar este proceso bioquímico de una manera masivamente paralela. Cada pocillo mantiene una molécula de ADN diferente. Por debajo de los pocillos hay una capa sensible a los iones y por debajo de esa un sensor de iones. Cuando se añade un nucleótido, por ejemplo uno C, a una plantilla de ADN y es después incorporado en una cadena de AN, se liberará un ion de hidrógeno. La carga de ese ion cambiará el pH de la solución, que puede ser identificado por el sensor de iones del Ion Torrent. El secuenciador - esencialmente el medidor de pH en estado sólido más pequeño del mundo, designa la base, yendo directamente de la información química a la información digital. El secuenciador Ion personal Genome Machine (PGM™) inunda después secuencialmente el chip con un nucleótido después de otro. Si el siguiente nucleótido que inunda el chip no es una equivalencia, no se registrará cambio de voltaje y no se designará una base. Di hay dos bases idénticas en la cadena de ADN, el voltaje será doble, y el chip registrará dos bases idénticas designadas. La identificación directa permite el registro de la incorporación de nucleótidos en segundos.

Otros métodos de secuenciación incluyen PCR digital y secuenciación por hibridación. La reacción en cadena de polimerasa digital (PCR digital o dPCR) se puede usar para identificar directamente y cuantificar ácidos nucleicos en una muestra. La PCR digital puede realizarse en una emulsión. Los ácidos nucleicos individuales se separan, por ejemplo, en un dispositivo de cámara de microfluidos, y cada ácido nucleico es amplificado individualmente por PCR. Los ácidos nucleicos pueden ser separados de tal manera que haya una media de aproximadamente 0,5 ácidos nucleicos/pocillo, o no más de un ácido nucleico/pocillo. Se pueden usar diferentes sondas para distinguir alelos fetales y alelos maternos. Los alelos se pueden enumerar para determinar el número de copias. En la secuenciación por hibridación, la hibridación comprende poner en contacto la pluralidad de secuencias de polinucleótidos con una pluralidad de sondas de polinucleótidos, en donde cada una de la pluralidad de sondas de polinucleótidos puede estar atada opcionalmente a un sustrato. El sustrato puede ser una superficie plana que comprende una matriz de secuencias de nucleótidos conocidas. El patrón de hibridación a la matriz puede usarse para determinar las secuencias de polinucleótidos presentes en la muestra. En otras realizaciones, cada sonda está atada a una microesfera, por ejemplo, una microesfera magnética o similar. La hibridación a las microesferas puede identificarse y usarse para identificar la pluralidad de secuencias de polinucleótidos dentro de la muestra.

En una realización, el método emplea secuenciación masivamente paralela de millones de fragmentos de ADN usando la secuenciación por síntesis de Illumina y química de secuenciación basada en terminadores reversible (por ejemplo como se describe en Bentley et al., Nature 6:53-59 [2009]). El ADN de la plantilla puede ser ADN genómico, por ejemplo ADN libre de células. En algunas realizaciones, se usa el ADN genómico de células aisladas como la plantilla, y se fragmenta en longitudes de varios cientos de pares de bases. En otras realizaciones, se usa ADN libre de células como la plantilla, y no se requiere fragmentación ya que el ADN libre de células existe como fragmentos cortos. Por ejemplo el ADN libre de células fetal circula en el torrente sanguíneo como fragmentos de <300 bp, y se ha estimado que el ADN libre de células materno circula como fragmentos de entre alrededor de 0,5 y 1 Kb (Li et al., Clin Chem, 50: 1002-1011 [2004]). La tecnología de secuenciación Illumina se basa en la unión

de ADN genómico fragmentado con una superficie plana, ópticamente trasparente en la que los anclajes de oligonucleótidos se enlazan. El ADN de la plantilla se repara en los extremos para generar extremos romos 5'-fosforilados, y la actividad de polimerasa del fragmento Klenow se usa para añadir una única base A al extremo 3' de los fragmentos de ADN fosforilados romos. Esta adición prepara los fragmentos de ADN para la ligadura con adaptadores de oligonucleótidos, que tienen un saliente de una única base T en sus extremos 30 para aumentar la eficiencia de la ligadura. Los oligonucleótidos adaptadores son complementarios a los anclajes de las células de flujo. Bajo condiciones de dilución limitativas se añade ADN de plantilla modificado por adaptadores, de una sola cadena a la célula de flujo y se inmovilizan por hibridación a los anclajes. Los fragmentos de ADN unidos se extienden y se amplifican por puente para crear una célula de flujo de secuenciación de ultra-alta densidad con cientos de millones de clústeres, cada uno conteniendo ~1000 copias de la misma plantilla. En una realización, el ADN genómico fragmentado aleatoriamente, por ejemplo ADN libre de células, se amplifica usando PCR antes de que se someta a amplificación de clústeres. Alternativamente, se usa una preparación de biblioteca genómica libre de amplificación, y el ADN genómico fragmentado aleatoriamente, por ejemplo ADN libre de células se enriquece usando sólo la amplificación de clústeres (Kozarewa et al., Nature Methods 6:291-295 [2009]). Las plantillas se secuencian usando una tecnología de secuenciación por síntesis de ADN de cuatro colores robusta que emplea terminadores reversibles con colorantes fluorescentes extraíbles. La identificación por fluorescencia de alta sensibilidad se consigue usando excitación por laser y ópticas de reflexión internas totales. Las lecturas de secuencias cortas de alrededor de 20-40 bp, por ejemplo 36 bp, se alinean contra un genoma de referencia enmascarado de repetición y las diferencias genéticas se designan usando software de línea de análisis de datos especialmente desarrollado. Después de la terminación de la primera lectura, las plantillas pueden generarse in situ para permitir una segunda lectura del extremo opuesto de los fragmentos. Así, se usa cualquier secuenciación de extremos emparejados o de único extremo de los fragmentos de ADN de acuerdo con el método. Se realiza la secuenciación parcial de fragmentos de ADN presentes en la muestra, y se cuentan las etiquetas de secuencia que comprenden lecturas de longitud predeterminada, por ejemplo 36 bp, que son mapeadas para un genoma de referencia conocido.

La longitud de la lectura de secuencia está asociada con la tecnología de secuenciación particular. Los métodos NGS proporcionan lecturas de secuencias que varían de tamaño de decenas a cientos de pares de bases. En algunas realizaciones del método descrito en la presente las lecturas de secuencia son de alrededor de 20bp, de alrededor de 25bp, de alrededor de 30bp, de alrededor de 35bp, de alrededor de 40bp, de alrededor de 45bp, de alrededor de 50bp, de alrededor de 55bp, de alrededor de 60bp, de alrededor de 65bp, de alrededor de 70bp, de alrededor de 75bp, de alrededor de 80bp, de alrededor de 85bp, de alrededor de 90bp, de alrededor de 95bp, de alrededor de 100bp, de alrededor de 110bp, de alrededor de 120bp, de alrededor de 130, de alrededor de 140bp, de alrededor de 150bp, de alrededor de 200bp, de alrededor de 250bp, de alrededor de 300bp, de alrededor de 350bp, de alrededor de 400bp, de alrededor de 450bp, o de alrededor de 500bp. Se espera que los avances tecnológicos permitan lecturas de extremos individuales de más de 500 bp permitiendo lecturas de más de alrededor de 1000 bp cuando se generan lecturas de extremos emparejados. En una realización, las lecturas de secuencia son 36 bp. Otros métodos de secuenciación que se pueden empelar por el método de la invención incluyen métodos de secuenciación de moléculas individuales que pueden secuenciar moléculas de ácidos nucleicos de >5000 bp. La cantidad masiva de salida de secuencia se transfiere por una línea de análisis que transforma salida de formación de imagen primaria del secuenciador en cuerdas de bases. Un paquete de algoritmos integrados realiza los pasos de transformación de datos primarios del núcleo: análisis de imágenes, puntuación de intensidad, designación de bases y alineamiento.

En una realización, se realiza la secuenciación parcial de fragmentos de ADN presentes en la muestra y se cuentan las etiquetas de secuencia que comprenden lecturas de longitud predeterminada, por ejemplo 36 bp, que mapean para un genoma de referencia conocido. Sólo se cuentan como etiquetas de secuencia las lecturas de secuencia que alinean únicamente con el genoma de referencia. En una realización, el genoma de referencia es la secuencia del genoma de referencia humano NCBI36/ hg 18, que está disponible en la red mundial en genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgsid=166260105). Otras fuentes de información de secuencias públicas incluyen GenBank, dbEST, dbSTS, EMBL (el Laboratorio de Biología Molecular Europeo), y el DDBJ (el Banco de datos de ADN de Japón). En otra realización el genoma de referencia comprende la secuencia del genoma de referencia humano NCBI36/hg18 y un genoma de secuencias objetivo artificiales, que incluye secuencias objetivo polimórficas, por ejemplo un genoma SNP que comprende las SEQ ID NOs: 1-56. En todavía otra realización, el genoma de referencia es un genoma de secuencia objetivo artificial que comprende secuencias objetivo polimórficas, por ejemplo secuencias SNP de las SEQ ID NOs: 1-56.

El mapeado de las etiquetas de secuencia se consigue comparando la secuencia de la etiqueta con la secuencia del genoma de referencia para determinar el origen cromosómico del molécula de ácidos nucleicos secuenciada (por ejemplo ADN libre de células), y la información de la secuencia genética específica no se necesita. Hay disponibles una variedad de algoritmos informáticos para alinear secuencias, incluyendo sin limitación BLAST (Altschul et al., 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), FASTA (Person & Lipman, 1988), BOWTIE (Langmead et al., Genome Biology 10:R25.1-R25.10 [2009]), o ELAND (Illumina, Inc., San Diego, CA, USA). En una realización, un extremo de las copias expandidas clonalmente de las moléculas de ADN libre de células del plasma se secuencia y procesa por análisis de alineamiento bioinformático para el Illumina Genome Analyzer que usa el

software Efficient Large-Scale Alignment of Nucleotide Databases (ELAND). El análisis de la información de secuenciación para la determinación de aneuploidía puede permitir un pequeño grado de discrepancia (0-2 discrepancias por etiqueta de secuencia) para dar cuenta de los polimorfismos menores que pueden existir entre el genoma de referencia y los genomas en la muestra mezclada. El análisis de la información de secuenciación para la determinación de la fracción fetal puede permitir un pequeño grado de discrepancia dependiendo de la secuencia polimórfica. Por ejemplo, se puede permitir un pequeño grado de discrepancia si la secuencia polimórfica es una STR. En casos donde la secuencia polimórfica es un SNP, todas las secuencias que coinciden exactamente con cualquiera de los dos alelos en el sitio del SNP se cuentan primero y se filtran de las lecturas restantes, para las que se puede permitir un pequeño grado de discrepancia.

Preparación de la Biblioteca de Secuenciación

Los secuenciadores de ADN de próxima generación, como el 454-FLX (Roche; en la página web 454.com), el SOLiD™ 3 (Applied Biosystems; en la página web solid.appliedbiosystems.com), y el Genome Analyzer (Illumina; http://www.illumina.com/pages.ilmn?ID=204) han transformado el horizonte de la genética a través de su capacidad de producir cientos de megabases de información de secuencias en una única ejecución.

Los métodos de secuenciación requieren la preparación de bibliotecas de secuenciación. La preparación de bibliotecas de secuenciación implica la producción de una colección aleatoria de fragmentos de ADN modificados con adaptadores, que son fáciles de secuenciar. Las bibliotecas de secuenciación de polinucleótidos se pueden preparar de ADN o ARN, incluyendo equivalentes, análogos de o ADN o ADNc, que es ADN complementario o copia producido de una plantilla de ARN, por ejemplo por la acción de la transcriptasa inversa. Los polinucleótidos pueden originarse en forma de ADN de doble cadena (ADNds) (por ejemplo fragmentos de ADN genómicos, PCR y productos de amplificación) o polinucleótidos que se han originado en forma de cadena simple, como ADN o ARN, y se han convertido a forma de ADNds. A modo de ejemplo, las moléculas de ARN, pueden copiarse en ADNc de doble cadena adecuados para su uso en la preparación de una biblioteca de secuenciación. La secuencia precisa de las moléculas de polinucleótidos primarios es generalmente no material para el método de preparación de la biblioteca, y puede ser conocida o desconocida. En una realización, las moléculas de polinucleótidos son moléculas de ADN. Más particularmente, las moléculas de polinucleótidos representan el complemento genético completo de un organismo, y son moléculas de ADN genómico, por ejemplo moléculas de ADN libre de células, que incluyen tanto la secuencia intrón como la exón (secuencia de codificación), así como las secuencias reguladoras no codificantes como las secuencias promotoras y potenciadoras. Todavía más particularmente, las moléculas de polinucleótidos primarios son moléculas de ADN genómico humano, por ejemplo moléculas de ADN libre de células presentes en sangre periférica de un sujeto embarazado. La preparación de bibliotecas de secuenciación para algunas plataformas de secuenciación de NGS requieren que los polinucleótidos sean de un rango específico de tamaños de fragmento, por ejemplo 0-1200 bp. Por lo tanto, se puede requerir la fragmentación de polinucleótidos, por ejemplo ADN genómico. El ADN libre de células existe como fragmentos de <300 pares de bases. Por lo tanto, la fragmentación del ADN libre de células no es necesaria para generar una biblioteca de secuenciación usando muestras de ADN libre de células. La fragmentación de moléculas de polinucleótidos por medios mecánicos, por ejemplo, nebulización, sonicación e hidrocorte, resulta en fragmentos con una mezcla heterogénea de extremos romos y 3'- y 5'- salientes. Si los polinucleótidos son fragmentados a la fuerza o existen de forma natural como fragmentos, se convierten a ADN de extremo romo que tiene 5-fosfatos y 3'-hidroxilo.

Típicamente, los extremos de los fragmentos se reparan en los extremos, es decir extremos romos usando métodos o kits conocidos en la técnica. Los fragmentos de extremos romos pueden ser fosforilados por tratamiento enzimático, por ejemplo usando quinasa de polinucleótidos. En algunas realizaciones, un único desoxinucleótido, por ejemplo desoxiadenosina (A) se añade a los extremos 3' de los polinucleótidos, por ejemplo, por la actividad de ciertos tipos de ADN polimerasa como la polimerasa Taq o polimerasa Klenow exo minus. Los productos de adicción de colas de DA son compatibles con el saliente 'T' presente en el término 3' de cada región dúplex de los adaptadores a los que son ligados en un paso posterior. La adicción de colas de dA evita la auto-ligadura de ambos de los polinucleótidos de extremo romo de tal manera que hay un desplazamiento hacia la formación de las secuencias ligadas por adaptadores. Los polinucleótidos de la adición de colas de dA se ligan a secuencias de polinucleótidos adaptadores de doble cadena. El mismo adaptador puede ser usado para ambos extremos del polinucleótido, o se pueden utilizar dos conjuntos de adaptadores. Los métodos de ligadura se conocen en la técnica y utilizan enzimas de ligasa como ADN ligasa para enlazar covalentemente el adaptados al polinucleótido de la adición de colas de d-A. El adaptador puede contener un resto de 5'-fosfato para facilitar la ligación al 3'-OH objetivo. El polinucleótido de la adicción de colas de dA contiene un resto de 5'fosfato, ya sea residual de un proceso de corte, o añadido usando un paso de tratamiento enzimático, y ha sido reparado en los extremos, y opcionalmente extendido por una base o bases salientes, para dar un 3'-OH adecuado para la ligadura. Los productos de la reacción de ligadura se purifican para eliminar adaptadores no ligados, adaptadores que pueden haberse ligado entre sí y para seleccionar un intervalo de tamaño de plantillas de generación de clústeres, que pueden estar precedidos por una amplificación, por ejemplo amplificación por PCR. La purificación de los productos de la ligadura puede obtenerse por métodos que incluyen electroforesis en gel e inmovilización reversible en fase sólida (SPRI).

Los protocolos estándar, por ejemplo protocolos para secuenciación que usan, por ejemplo, la plataforma

Illumina, instruyen a los usuarios para purificar los productos reparados en los extremos antes de la adición de colas de dA, y para purificar los productos de la adición de colas de Da antes de los pasos de ligadura de adaptadores de la preparación de la biblioteca. La purificación de los productos reparados en los extremos y los productos de la adición de colas de dA elimina enzimas, tampones, sales y similares para proporcionar condiciones de reacción favorables para el paso enzimático posterior. En una realización los pasos de reparación de extremos, adición de colas de dA y ligado por adaptadores excluyen los pasos de purificación. Así, en una realización, el método de la invención abarca preparar una biblioteca de secuenciación que comprende los pasos consecutivos de reparación de extremos, adición de colas de dA y ligado por adaptadores. En realizaciones para preparar las bibliotecas de secuenciación que no requieren el paso de adición de colas de dA, por ejemplo protocolos para secuenciación que usan las plataformas Roche 454 y SOLID™, los pasos de reparación de extremos y ligado por adaptadores excluyen el paso de purificación de los productos reparados en los extremos antes del ligado por adaptadores.

En un siguiente paso de una realización del método, se prepara una reacción de amplificación. El paso de amplificación introduce a las moléculas de la plantilla ligadas por adaptadores las secuencias de oligonucleótidos requeridas para la hibridación a la célula de flujo. Los contenidos de una reacción de amplificación se conocen por alguien experto en la técnica e incluyen sustratos apropiados (como dNTPs), enzimas (por ejemplo una ADN polimerasa) y componentes de tampón requeridos para la reacción de amplificación. Opcionalmente, la amplificación de polinucleótidos ligados por adaptadores se puede omitir. Generalmente las reacciones de amplificación requieren al menos dos cebadores de amplificación, es decir oligonucleótidos cebadores, que pueden ser idénticos, e incluyen una "porción específica del adaptador", capaz de hibridar a una secuencia que enlaza con el cebador en la molécula de polinucleótidos a ser amplificada (o el complemento de la misma se la plantilla se ve como una cadena única) durante el paso de hibridación. Una vez formada, la biblioteca de plantillas preparada de acuerdo con los métodos descritos anteriormente se puede usar para amplificación de ácidos nucleicos en fase sólida. El término "amplificación en fase sólida" como se usa en la presente se refiere a una reacción de amplificación de ácidos nucleicos llevada a cabo en o en asociación con un soporte sólido de tal manera que todos o una porción de los productos amplificados se inmovilizan en el soporte sólido a medida que se forman. En particular, el término abarca reacción en cadena de polimerasa en fase sólida (PCR en fase sólida) y amplificación isotérmica en fase sólida que son reacciones análogas a la amplificación en fase de solución estándar, excepto que uno o ambos de los cebadores directo e inverso es/están inmovilizados en el soporte sólido. La PCR en fase sólida cubre sistemas como emulsiones, en donde un cebador se ancla a una microesfera y el otro está en la solución libre, y la formación de colonias en matrices en gel en fase sólida en donde un cebador está anclado a la superficie, y uno está en la solución libre. Después de la amplificación, las bibliotecas de secuenciación pueden ser analizadas por electroforesis capilar de microfluidos para asegurar que la biblioteca está libre de dímeros de adaptadores o ADN de cadena única. La biblioteca de moléculas de polinucleótidos de la plantilla es particularmente adecuada para su uso en métodos de secuenciación en fase sólida. Además de proporcionar plantillas para la secuenciación en fase sólida y PCR en fase sólida, las plantillas de la biblioteca proporcionan plantillas para la amplificación del genoma completo.

En una realización, la biblioteca de polinucleótidos ligados por adaptadores es sometida a secuenciación masivamente paralela, que incluye técnicas para secuenciar millones de fragmentos de ácidos nucleicos, por ejemplo usando la unión de ADN genómico fragmentado aleatoriamente a una superficie plana, ópticamente transparente y amplificación en fase sólida para crear una célula de flujo de secuenciación de alta densidad con millones de clústeres. Las matrices agrupadas se pueden preparar usando o un proceso de termociclado, como se describe en la patente WO9844151, o un proceso por el que la temperatura se mantiene como una constante, y los ciclos de extensión y desnaturalización se realizan usando cambios de reactivos. El método Solexa/Illumina referido en la presente se basa en la unión de ADN genómico aleatoriamente fragmentado a una superficie plana, ópticamente transparente. Los fragmentos de ADN unidos se extienden y se amplifican por puente para crear una célula de flujo de secuenciación de ultra-alta densidad con millones de clústeres cada uno conteniendo miles de copias de la misma plantilla (WO 00/18957 y WO 98/44151). Las plantillas de clústeres se secuencian usando una tecnología de secuenciación por síntesis de ADN de cuatro colores robusta que emplea terminadores reversibles con colorantes fluorescentes extraíbles. Alternativamente, la biblioteca puede ser amplificada en microesferas en donde cada microesfera contiene un cebador de amplificación directo e inverso.

La secuenciación de las bibliotecas amplificadas puede llevarse a cabo usando cualquier técnica de secuenciación adecuada como se describe en la presente. En una realización, la secuenciación es secuenciación masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles. En otras realizaciones, la secuenciación es secuenciación masivamente paralela usando secuenciación por ligadura. En otras realizaciones, la secuenciación es secuenciación de moléculas individuales.

Determinación de Aneuploidía

La precisión requerida para determinar correctamente si hay presencia o ausencia de una aneuploidía en una muestra se determina en parte por la variación del número de etiquetas de secuencia que mapean para el genoma de referencia entre las muestras de una ejecución de secuenciación (variabilidad inter-cromosómica), y la variación del número de etiquetas de secuencia que mapean para el genoma de referencia en diferentes ejecuciones de secuenciación (variabilidad inter-secuenciación). Por ejemplo, las variaciones pueden ser particularmente

pronunciadas para etiquetas que mapean para secuencias de referencia ricas en GC o pobres en GC. En una realización, el método usa información de secuenciación para calcular la dosis de cromosoma, que cuenta intrínsecamente para la variabilidad acumulada derivada de la variabilidad intercromosómica, inter-secuenciación y dependiente de la plataforma. Las dosis de cromosomas se determinan de la información de la secuenciación, es decir, el número de etiquetas de secuencia, para la secuencia de interés, por ejemplo cromosoma 21, y el número de etiquetas de secuencia para una secuencia normalizadora. La identificación de una secuencia normalizadora se realiza en un conjunto de muestras calificadas que se sabe no contienen una aneuploidía de la secuencia de interés. El diagrama de flujo proporcionado en la Figura 1 muestra una realización del método **100** por el que se identifican las secuencias normalizadoras, por ejemplo cromosomas normalizadores, y se determina la presencia o ausencia de una aneuploidía.

En el paso **110**, se obtiene un conjunto de muestras maternas calificadas para identificar secuencias normalizadoras calificadas, por ejemplo cromosomas normalizadores, y para proporcionar valores de varianza para su uso en determinar estadísticamente la identificación significativa de una aneuploidía en muestras de ensayo. En el paso **110**, se obtienen una pluralidad de muestras calificadas biológicas de una pluralidad de sujetos que se sabe comprenden células que tienen un número de copias normal para cualquier secuencia de interés, por ejemplo un cromosoma de interés como un cromosoma asociado con una aneuploidía. En una realización, las muestras calificadas se obtienen de madres embarazadas con un feto que se ha sido confirmado usando métodos citogenéticos que tiene un número de copias normal de cromosomas en relación al cromosoma de interés. Las muestras biológicas calificadas pueden ser muestras de fluidos biológicos, por ejemplo muestras de plasma, o cualquier muestra como se ha descrito anteriormente que contenga una mezcla de moléculas de ADN libre de células fetales y maternas. La muestra es una muestra materna que se obtiene de una hembra embarazada, por ejemplo una mujer embarazada. Cualquier muestra biológica materna se puede usar como fuente de ácidos nucleicos fetales y maternos que estén contenidos en células o que sean de "células libres". En algunas realizaciones, es ventajoso obtener una muestra materna que comprenda ácidos nucleicos libre de células, por ejemplo ADN libre de células. Preferiblemente, la muestra biológica materna es una muestra de fluido biológico. Un fluido biológico incluye, como ejemplos no limitativos, muestras de sangre, plasma, suero, sudor, lágrimas, esputo, orina, esputo, flujo del oídos, linfa, saliva, fluido cefalorraquídeo, estragos, suspensión de médula ósea, flujo vaginal, lavado transcervical, fluido cerebral, ascitis, leche, secreciones de los tractos respiratorio, intestinal y genitourinario, fluido amniótico y leucoforesis. En algunas realizaciones, la muestra de fluido biológico es una muestra que se puede obtener fácilmente por procedimientos no invasivos, por ejemplo, sangre, plasma, suero, sudor, lágrimas, esputo, orina, esputos, fluido del oído y saliva. En algunas realizaciones, la muestra biológica es una muestra de sangre periférica, o fracciones del plasma y/o suero de la misma. En otras realizaciones la muestra es una mezcla de dos o más muestras biológicas, por ejemplo, una muestra biológica puede comprender dos o más de una muestra de fluido biológico. Como se usa en la presente, los términos "sangre", "plasma" y "suero" abarcan expresamente fracciones o porciones procesadas de los mismos. En algunas realizaciones, la muestra biológica se procesa para obtener una fracción de la muestra, por ejemplo plasma, que contiene la mezcla de ácidos nucleicos fetales y maternos. En algunas realizaciones, la mezcla de ácidos nucleicos fetales y maternos se procesa adicionalmente de la fracción de la muestra, por ejemplo plasma, para obtener una muestra que comprende una mezcla purificada de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células. Los ácidos nucleicos libres de células, incluyendo del ADN libre de células, se pueden obtener por varios métodos conocidos en la técnica a partir de muestras biológicas que incluyen, pero no están limitadas a, plasma, suero y orina (Fan et al., Proc Natl Acad Sci 105:16266-16271 [2008]; Koide et al., Prenatal Diagnosis 25:604-607 [2005]; Chen et al., Nature Med. 2: 1033-1035 [1996]; Lo et al., Lancet 350: 485-487 [1997). Para separar el ADN de células libres de las células, se pueden usar métodos de fraccionamiento, centrifugación (por ejemplo centrifugación en gradiente de densidad), precipitación específica del ADN, o clasificación de células de alto rendimiento y/o separación. Hay disponibles kits comercialmente disponibles para la separación manual o automática de ADN libre de células (Roche Diagnostics, Indianapolis, IN, Qiagen,Valencia, CA, Macherey-Nagel, Duren, DE). En algunas situaciones, puede ser ventajoso fragmentar las moléculas de ácidos nucleicos en la muestra de ácido nucleico. La fragmentación puede ser aleatoria, o puede ser específica, como se consigue, por ejemplo, usando digestión con endonucleasas de restricción. Los métodos para la fragmentación aleatoria son bien conocidos en la técnica, e incluyen, por ejemplo, digestión con ADNasa limitada, tratamiento alcalino y corte físico. En una realización, los ácidos nucleicos de la muestra se obtiene de ADN libre de células, que no está sometido a fragmentación. En otras realizaciones, los ácidos nucleicos de la muestra se obtienen como ADN genómico, que está sometido a fragmentación en fragmentos de aproximadamente 500 o más pares de bases, y a los que se pueden aplicar fácilmente métodos de NGS. Una biblioteca de secuenciación se prepara de ADN fragmentado de forma natural o fragmentado a la fuerza. En una realización, la preparación de la biblioteca de secuenciación comprende los pasos consecutivos de reparación de extremos, adición de colas de dA, y ligado por adaptadores de los fragmentos de ADN. En otra realización, la preparación de la biblioteca de secuenciación comprende los pasos consecutivos de reparación de extremos y ligado por adaptadores de los fragmentos de ADN.

En el paso **120**, se secuencia al menos una porción de cada uno de todos los ácidos nucleicos calificados contenidos en las muestras maternas calificadas. Antes de la secuenciación, la mezcla de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células purificado, se modifica para preparar una biblioteca de secuenciación para generar lecturas de secuencia de entre 20 y 40 bp, por ejemplo 36 bp, que están alineadas a un genoma de

referencia, por ejemplo hg18. En algunas realizaciones, las lecturas de secuencia comprenden alrededor de 20bp, alrededor de 25bp, alrededor de 30bp, alrededor de 35bp, alrededor de 40bp, alrededor de 45bp, alrededor de 50bp, alrededor de 55bp, alrededor de 60bp, alrededor de 65bp, alrededor de 70bp, alrededor de 75bp, alrededor de 80bp, alrededor de 85bp, alrededor de90bp, alrededor de 95bp, alrededor de 100bp, alrededor de 110bp, alrededor de 120bp, alrededor de 130, alrededor de 140bp, alrededor de 150bp, alrededor de 200bp, alrededor de 250bp, alrededor de 300bp, alrededor de 350bp, alrededor de 400bp, alrededor de 450bp, o alrededor de 500bp. Se espera que los avances tecnológicos permitan lecturas de extremos individuales de más de 500bp permitiendo lecturas de más de alrededor de 1000bp cuando se generan lecturas de extremos emparejados. En una realización, las lecturas de secuencia comprenden 36bp. Las lecturas de secuencia están alineadas a un genoma de referencia humano, y las lecturas que se mapean únicamente para el genoma de referencia humana se cuentan como etiquetas de secuencia. En una realización, se obtienen al menos alrededor de 3×10^6 etiquetas de secuencia calificadas, al menos alrededor de 5×10^6 etiquetas de secuencia calificadas, al menos alrededor de 8×10^6 etiquetas de secuencia calificadas, al menos alrededor de 10×10^6 etiquetas de secuencia calificadas, al menos alrededor de 15×10^6 etiquetas de secuencia calificadas, al menos alrededor de 20×10^6 etiquetas de secuencia calificadas, al menos alrededor de 30×10^6 etiquetas de secuencia calificadas, al menos alrededor de 40×10^6 etiquetas de secuencia calificadas, o al menos alrededor de 50×10^6 etiquetas de secuencia calificadas, que comprende lecturas de entre 20 y 40bp se obtienen de lecturas que mapean únicamente para un genoma de referencia.

En el paso **130**, todas las etiquetas obtenidas de la secuenciación de ácidos nucleicos en las muestras maternas calificadas se cuentan para determinar una densidad de etiqueta de secuencia calificada. En una realización la densidad de etiqueta de secuencia se determina como el número de etiquetas de secuencia calificadas mapeadas para la secuencia de interés en el genoma de referencia. En otra realización, la densidad de etiqueta de secuencia calificad se determina como el número de etiquetas de secuencia calificadas mapeadas para una secuencia de interés normalizada a la longitud de la secuencia calificada de interés para la que se mapean. Las densidades de etiqueta de secuencia que se determinan como una proporción de la densidad de etiqueta en relación a la longitud de la secuencia de interés se refieren en al presente como proporciones de densidad de etiqueta. La normalización a la longitud de la secuencia de interés no se requiere, y puede ser incluida como un paso para reducir el número de dígitos en un número para simplificarla para la interpretación humana. Como todas las etiquetas de secuencia calificadas se mapean y cuentan en cada una de las muestras calificadas, la densidad de etiqueta de secuencia para una secuencia de interés, por ejemplo cromosoma de interés, se determina en las muestras calificadas, como lo son las densidades de etiquetas de secuencia para las secuencias adicionales de las que se identifican posteriormente las secuencias normalizadoras, por ejemplo cromosomas. En una realización, la secuencia de interés es un cromosoma que está asociado con una aneuploidía cromosómica, por ejemplo cromosoma 21, y la secuencia normalizadora calificada es un cromosoma que no está asociado con una aneuploidía cromosómica y cuya variación en la densidad de etiqueta de secuencia se aproxima mejor a la del cromosoma 21. Por ejemplo, una secuencia normalizadora calificada es una secuencia que tiene la variabilidad más pequeña. En algunas realizaciones, la secuencia normalizadora es una secuencia que distingue mejor una o más muestras calificadas de una o más muestras afectadas, es decir, la secuencia normalizadora es una secuencia que tiene la diferenciabilidad más alta. El nivel de diferenciabilidad puede determinarse como una diferencia estadística entre las dosis de cromosomas en una población de muestras calificadas y las dosis de cromosomas en una o más muestras de ensayo. En otra realización, la secuencia de interés es un segmento de un cromosoma asociado con una aneuploidía parcial, por ejemplo una deleción o inserción cromosómica, o translocación cromosómica desequilibrada, y la secuencia normalizadora es un segmento cromosómico que no está asociado con la aneuploidía parcial y cuya variación en la densidad de etiqueta de secuencia mejor se aproxima a la del segmento de cromosoma asociado con la aneuploidía parcial

En el paso **140**, en base a las densidades de etiquetas calificadas calculadas, se determina una dosis de secuencia calificada para una secuencia de interés como la proporción de la densidad de etiqueta de secuencia para la secuencia de interés y la densidad de etiqueta de secuencia calificada para secuencias adicionales de las que se identifican posteriormente secuencias normalizadoras. En una realización, las dosis para el cromosoma de interés, por ejemplo, cromosoma 21, se determinan como una proporción de la densidad de etiqueta de secuencia del cromosoma 21 y la densidad de etiqueta de secuencia para cada uno de los cromosomas restantes, es decir cromosomas 1-20, cromosoma 22, cromosoma X y cromosoma Y (ver Ejemplos 3-5, y Figuras 9-15).

En el paso **145**, una secuencia normalizadora, por ejemplo un cromosoma normalizador, se identifica para una secuencia de interés, por ejemplo cromosoma 21, en una muestra calificada basada en las dosis de secuencias calculadas. El método identifica secuencias que tienen inherentemente características similares y que son propensas a variaciones similares entre muestras y ejecuciones de secuenciación, y que son útiles para determinar dosis de secuencias en muestras de ensayo. En algunas realizaciones, la secuencia normalizadora es una que diferencia mejor una muestra afectada, por ejemplo una muestra aneuploide, de una o más de las muestras calificadas. En otras realizaciones, una secuencia normalizadora es una secuencia que muestra una variabilidad en el número de etiquetas de secuencia que se mapean para ella entra muestras y ejecuciones de secuenciación que mejor se aproximan a la de la secuencia de interés para la que se usa como un parámetro normalizador y/o que puede diferenciar mejor una muestra afectada de una o más muestras no afectadas.

En algunas realizaciones, se identifica más de una muestra normalizadora. Por ejemplo, la variación, por ejemplo coeficiente de variación, en la dosis de cromosoma para el cromosoma de interés 21 es menor cuando se usa la densidad de etiqueta de secuencia del cromosoma 14. En otras realizaciones, se identifican dos, tres, cuatro, cinco, seis, siete, ocho o más secuencias normalizadoras para su uso al determinar una dosis de secuencia para una secuencia de interés en una muestra de ensayo.

En una realización, la secuencia normalizadora para el cromosoma 21 es seleccionada del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, cromosoma 14, cromosoma 15, cromosoma 16, y cromosoma 17. Preferiblemente, la secuencia normalizadora para el cromosoma 21 se selecciona del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 11, cromosoma 12, y cromosoma 14. Alternativamente, la secuencia normalizadora para el cromosoma 21 es un grupo de cromosomas seleccionados del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, cromosoma 14, cromosoma 15, cromosoma 16, y cromosoma 17. En otras realizaciones, la secuencia normalizadora para el cromosoma 21 es un grupo de cromosomas seleccionados del cromosoma 9, cromosoma 1, cromosoma 2, cromosoma 11, cromosoma 12, y cromosoma 14.

En una realización, la secuencia normalizadora para el cromosoma 18 se selecciona del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, y cromosoma 14. Preferiblemente, la secuencia normalizadora para el cromosoma 18 se selecciona del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 5, cromosoma 6, cromosoma 12, y cromosoma 14. Alternativamente, la secuencia normalizadora para el cromosoma 18 es un grupo de cromosomas seleccionados del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, y cromosoma 14. En otras realizaciones, la secuencia normalizadora para el cromosoma 18 es un grupo de cromosomas seleccionados del cromosoma 8, cromosoma 2, cromosoma 3, cromosoma 5, cromosoma 6, cromosoma 12, y cromosoma 14.

En una realización, la secuencia normalizadora para el cromosoma X se selecciona del cromosoma 1, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, cromosoma 14, cromosoma 15, y cromosoma 16. Preferiblemente, la secuencia normalizadora para el cromosoma X se selecciona del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, y cromosoma 8. Alternativamente, la secuencia normalizadora para el cromosoma X es un grupo de cromosomas seleccionados del cromosoma 1, cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 13, cromosoma 14, cromosoma 15, y cromosoma 16. En otras realizaciones, la secuencia normalizadora para el cromosoma X es un grupo de cromosomas seleccionados del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, y cromosoma 8.

En una realización, la secuencia normalizadora para el cromosoma 13 es un cromosoma seleccionado del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 14, cromosoma 18, y cromosoma 21. Preferiblemente, la secuencia normalizadora para el cromosoma 13 se selecciona del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, y cromosoma 8. En otra realización, la secuencia normalizadora para el cromosoma 13 es un grupo de cromosomas seleccionados del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, cromosoma 7, cromosoma 8, cromosoma 9, cromosoma 10, cromosoma 11, cromosoma 12, cromosoma 14, cromosoma 18, y cromosoma 21. En otras realizaciones, la secuencia normalizadora para el cromosoma 13 es un grupo de cromosomas seleccionado del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma 5, cromosoma 6, y cromosoma 8.

La variación en la dosis de cromosoma para el cromosoma Y es mayor de 30 independientemente de que cromosoma normalizador se use para determinar la dosis de cromosoma Y. Por lo tanto, cualquier cromosoma, o un grupo de dos o más cromosomas seleccionados de los cromosomas 1-22 y el cromosoma X se puede usar como la secuencia normalizadora para el cromosoma Y. En una realización, el al menos un cromosoma normalizadora es un grupo de cromosomas que consiste de los cromosomas 1-22, y el cromosoma X. En otra realización, el al menos un cromosoma normalizador es un grupo de cromosomas seleccionados del cromosoma 2, cromosoma 3, cromosoma 4, cromosoma ,5 y cromosoma 6.

En base a la identificación de la secuencia(s) normalizadora en muestras calificadas, se determina una dosis de cromosoma para una secuencia de interés en una muestra de ensayo que comprende una mezclad e ácidos nucleicos derivados de genomas que difieren en una o más secuencias de interés.

En el paso **115**, se obtiene una muestra de ensayo, por ejemplo muestra de plasma, que comprende ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células, de un sujeto embarazado, por ejemplo una mujer embarazada, para la que se necesita determinar la presencia o ausencia de una aneuploidía fetal.

Se prepara una biblioteca de secuenciación como se describe para el paso 120, y en el paso **125**, al menos una porción de los ácidos nucleicos de ensayo en la muestra de ensayo se secuencia para generar millones de lecturas de secuencia que comprenden entre 20 y 500 bp, por ejemplo 36bp. Como en el paso **120**, las lecturas generadas de la secuenciación de ácidas nucleicos en la muestra de ensayo se mapean únicamente para un genoma de referencia humano y se cuentan. Como se describe en el paso **120**, se obtienen al menos alrededor de 3 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 5 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 8 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 10 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 15 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 20 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 30 x 10^6 etiquetas de secuencia calificadas, al menos alrededor de 40 x 10^6 etiquetas de secuencia calificadas, o al menos alrededor de 50 x 10^6 etiquetas de secuencia calificadas que comprenden entre 20 y 40bp de lecturas que mapean únicamente para el genoma de referencia humano.

En el paso **135**, todas las etiquetas obtenidas de la secuenciación de ácidos nucleicos en las muestras de ensayo se cuentan para determinar la densidad de etiqueta de secuencia. En una realización, el número de etiquetas de secuencia de ensayo mapeadas para una secuencia de interés se normaliza a la longitud conocida de una secuencia de interés a la que están mapeadas para proporcionar una densidad de etiqueta de secuencia de ensayo. Como se describe para las muestras calificadas, no se requiere la normalización a la longitud conocida de una secuencia de interés, y puede incluirse como un paso para reducir al número de dígitos en un número para simplificarla para la interpretación humana. A medida que todas las etiquetas de secuencia de ensayo mapeadas se cuentan en la muestra de ensayo, se determina en las muestras de ensayo la densidad de etiqueta de secuencia para una secuencia de interés, por ejemplo una secuencia clínicamente relevante como el cromosoma 21, como los son las densidades de etiqueta de secuencia para secuencias adicionales que corresponden con la menos una secuencia normalizadora identificada en las muestras calificadas.

En el paso **150**, en base a la identidad de al menos una secuencia normalizadora en las muestras calificadas, se determina una dosis de secuencia de ensayo para una secuencia de interés en la muestra de ensayo. La dosis de secuencia, por ejemplo dosis de cromosoma, para una secuencia de interés en una muestra de ensayo es una proporción de la densidad de etiqueta de secuencia determinada para la secuencia de interés en la muestra de ensayo y la densidad de etiqueta de secuencia de al menos una secuencia normalizadora determinada en la muestra de ensayo, en donde la secuencia normalizadora en la muestra de ensayo se corresponde con la secuencia normalizadora identificada en las muestras calificadas para la secuencia particular de interés. Por ejemplo, si se determina que la secuencia normalizadora para el cromosoma 21 en las muestras calificadas es el cromosoma 14, entonces la dosis de secuencia de ensayo para el cromosoma 21 (secuencia de interés) se determina como la proporción de la densidad de etiqueta de secuencia para el cromosoma 21 en y la densidad de etiqueta de secuencia para el cromosoma 14 cada una determinada en la muestra de ensayo. De manera similar, se determinan las dosis de cromosomas para los cromosomas 13, 18, X, Y y otros cromosomas asociados con las aneuploidías cromosómicas. Como se ha descrito anteriormente, una secuencia de interés puede ser parte de un cromosoma, por ejemplo un segmento de cromosoma. Por consiguiente, la dosis para un segmento de cromosoma puede determinarse como la proporción de la densidad de etiqueta de secuencia determinada para el segmento en la muestra de ensayo y la densidad de etiqueta de secuencia para el segmento de cromosoma normalizador en la muestra de ensayo, en donde el segmento normalizador en la muestra de ensayo corresponde con el segmento normalizador identificado en las muestras calificadas para el segmento particular de interés.

En el paso 155, los valores límite se derivan de los valores de la desviación estándar establecidos por una pluralidad de dosis de secuencia calificadas. La clasificación precisa depende de las diferencias entre las distribuciones de probabilidad para las diferentes clases, es decir tipo de aneuploidía. Preferiblemente, los límites se eligen de la distribución empírica para cada tipo de aneuploidía, por ejemplo trisomía 21. Los valores límites posibles que se establecieron para clasificar las aneuploidías de trisomía 13, trisomía 18, trisomía 21 y monosomía X como se describe en los Ejemplos, que describen el uso del método para determinar las aneuploidías cromosómicas por secuenciación de ADN libre de células extraído de una muestra materna comprenden una mezcla de ácidos nucleicos fetales y maternos.

En el paso **160**, la variación del número de copias de la secuencia de interés, por ejemplo aneuploidía cromosómica o parcial, se determina en la muestra de ensayo comparando al dosis de secuencia de ensayo para la secuencia de interés con al menos un valor límite establecido de las dosis de secuencia calificadas.

En el paso **160**, la dosis calculada para una secuencia de ensayo de interés se compara con ese conjunto como los valores límite que se eligen de acuerdo con un límite definido por el usuario de fiabilidad para clasificar al muestra como una "normal", una "afectada" o una "no designada" en el paso **165**. Las muestras "no designadas" son muestras para las que no se puede hacer un diagnóstico definitivo con fiabilidad.

Otra realización divulgada en la presente es un método para proporcionar diagnóstico prenatal de una aneuploidía cromosómica fetal en una muestra biológica que comprende moléculas de ácidos nucleicos fetales y

maternos. El diagnóstico se hace en base a la recepción de datos de la secuenciación en al menos una porción de la mezcla de moléculas de ácidos nucleicos fetales y maternos derivados de una muestra de ensayo biológica, por ejemplo una muestra de plasma materno, calculando de los datos de secuenciación una dosis de cromosoma normalizador para uno o más cromosomas de interés, determinando una diferencia estadísticamente significativa entre la dosis de cromosoma normalizador para el cromosoma de interés en la muestra de ensayo y un valor límite establecido en una pluralidad de muestras calificadas (normales), y proporcionando el diagnóstico prenatal en base a la diferencia estadística. Como se describe en el paso **165** del método, se hace un diagnóstico de normal o afectada. Se proporciona una "no designada" en el caso de que el diagnóstico para la normal o afectada no se pueda hacer con confianza.

Determinación del CNV para diagnósticos prenatales

El ARN y ADN fetal libre de células que circula en la sangre materna puede usarse para el diagnóstico prenatal no invasivo temprano (NIPD) de un número creciente de condiciones genéticas, tanto para la gestión del embarazo y para ayudar a la toma de decisiones reproductoras. La presencia de ADN libre de células que circulan en el torrente sanguíneo se ha conocido durante más de 50 años. Más recientemente, la presencia de pequeñas cantidades de ADN fetal circulante se descubrió en el torrente sanguíneo materno durante el embarazo (Lo et al., Lancet 350:485-487 [1997]). Creyendo que se origina de células de la placenta que se mueren, el ADN fetal libre de células (ADNcf) ha demostrado que consiste de fragmentos cortos típicamente de menos de 200 bp de longitud Chan et al., Clin Chem 50:88-92 [2004]), que se pueden distinguir tan pronto como desde las 4 semanas de gestación (Illanes et al, Early Human Dev 83:563-566 [2007]), y que se sabe desaparecen de la circulación materna pocas horas después del parto (Lo et al., Am J Hum Genet 64:218-224 [1999]). Además del ADN libre de células, también se pueden distinguir fragmentos de ARN fetal libre de células (ARNcf) en el torrente sanguíneo materno, originado de genes que es transcriben en el feto o la placenta. La extracción y el posterior análisis de estos elementos genéticos fetales de una muestra de sangre materna ofrece nuevas oportunidades para el NIPD.

El presente método es un método independiente de polimorfismos para el uso en NIPD y que no requiere que el ADN libre de células fetal se distinga del ADN libre de células materno para permitir la determinación de una aneuploidía fetal. En algunas realizaciones, la aneuploidía es una trisomía o monosomía cromosómica completa, o una trisomía o monosomía parcial. Las aneuploidías parciales son causadas por la pérdida o ganancia de parte de un cromosoma, y abarcan desequilibrios cromosómicos resultantes de translocaciones desequilibradas, inversiones desequilibradas, deleciones e inserciones. Con mucho, la aneuploidía más común conocida compatible con la vida es la trisomía 21, es decir Síndrome de Down (DS), que es causado por la presencia de parte o todo el cromosoma 21. Raramente, el DS puede ser causa de un defecto heredado o esporádico por el que una copia extra de todo o parte del cromosoma 21 se uno con otro cromosoma (habitualmente el cromosoma 14) para formar un único cromosoma anormal. El DS está asociado con deficiencia intelectual, dificultades de aprendizaje severas y exceso de mortalidad causado por problemas de salud a largo plazo como enfermedades cardiacas. Otras aneuploidías con significancia clínica conocida incluyen el síndrome de Edward (trisomía 18) y Síndrome de Patau (trisomía 13), que son frecuentemente fatales dentro de los primeros meses de vida. También se conocen anomalías asociadas con el número de cromosomas sexuales e incluyen monosomía X, por ejemplo síndrome de Turner (XO), y síndrome de triple X (XXX) en nacimientos de niñas y síndrome de Kleinefelter (XXY) y síndrome de XYY en nacimientos de niños, que están todos asociados con varios fenotipos incluyendo esterilidad y reducción en las habilidades intelectuales. El método de la invención puede ser usado para diagnosticar estas y otras anomalías cromosómicas prenatalmente.

De acuerdo con las realizaciones de la presente invención la trisomía determinada por la presente invención se selecciona de trisomía 21 (T21; Síndrome de Down), trisomía 18 (T18; Síndrome de Edward); trisomía 16 (T16), trisomía 22 (T22; Síndrome de Ojo de Gato); trisomía 15 (T15; Síndrome de Prader Willi), trisomía 13 (T13; Síndrome de Patau), trisomía 8 (T8; Síndrome de Warkany) y trisomías XXY (Síndrome de Kleinefelter), XYY o XXX. Se apreciará que varias otras trisomías y trisomías parciales se pueden determinar en ADN libre de células fetal de acuerdo con las enseñanzas de la presente invención. Estas incluyen, pero no están limitadas a, trisomía parcial 1q32-44, trisomía 9 p con trisomía, trisomía, mosaicismo 4, trisomía 17p, trisomía parcial 4q26-qter, trisomía 9, trisomía parcial 2p, trisomía parcial 1q, y/o trisomía parcial 6p/monosomía 6q.

El método de la presente invención también puede usarse para determinar la monosomía X cromosómica, y monosomías parciales como la monosomía 13, monosomía 15, monosomía 16, monosomía 21 y monosomía 22, que se sabe están implicadas en el aborto involuntario del embarazo. La monosomía parcial de cromosomas típicamente implicados en la aneuploidía completa también pude ser determinada por el método de la invención. La monosomía 18p es un trastorno cromosómico raro en el que todo o parte del brazo corto (p) del cromosoma 18 se elimina (monosómico). Este trastorno está caracterizado típicamente por estatura corta, grados variables de atraso mental, retrasos en el habla, malformaciones de la región craneal y facial (craneofacial), y/o anomalías físicas adicionales. Los defectos craneofaciales asociados pueden variar enormemente en rango y severidad de caso a caso. Las condiciones provocadas por cambios en la estructura o número de copias del cromosoma 15 incluyen el Síndrome de Angelman y el Síndrome de Prader-Willi, que implican una pérdida de actividad génica en la misma parte del cromosoma 15, la región 15q11-q13. Se apreciará que varias translocaciones y microdeleciones pueden

ser asintomáticas en el padre portador, sin embargo pueden provocar una enfermedad genética importante en la descendencia. Por ejemplo, una madre sana que porta la microdeleción 15q11-q13 puede dar a luz a un niño con el síndrome de Angelman, un trastorno neurodegenerativo severo. Por lo tanto, la presente invención puede usarse para identificar dicha deleción en el feto. La monosomía parcial 13q es un trastorno cromosómico raro que se produce cuando falta una pieza del brazo largo (q) del cromosoma 13 (monosómico). Los bebés nacidos con monosomía 13q parcial pueden mostrar bajo peso de nacimiento, malformaciones de la cabeza y la cara (región craneofacial), anomalías esqueléticas (especialmente de las manos y pies), y otras anomalías físicas. El retraso mental es característico de esta condición. La tasa de mortalidad durante la niñez es alta entre individuos nacidos con este trastorno. Casi todos los casos de monosomía 13q parcial tienen lugar aleatoriamente sin razón aparente (esporádica). El síndrome de deleción 22q11.2, también conocido como síndrome de DiGeorge, es un síndrome causado por la deleción de una pieza pequeña del cromosoma 22. La deleción (22q11.2) tiene lugar cerca del medio del cromosoma en el brazo largo de uno del par de cromosomas. Las características de este síndrome varían ampliamente, incluso entre miembros de la misma familia, y afectan a muchas partes del cuerpo. Los signos característicos y síntomas pueden incluir defectos de nacimiento como enfermedad cardiaca congénita, defectos en el paladar, más comúnmente relacionados con problemas neuromusculares con cleisis (insuficiencia velo-faríngea), problemas de aprendizaje, diferencias leves en los rasgos faciales e infecciones recurrentes. Las microdeleciones en la región cromosómica 22q11.2 están asociadas con un riesgo aumentado de 20 a 30 veces de esquizofrenia. En una realización, el método de la invención se usa para determinar monosomías parciales que incluyen per no están limitadas a monosomía 18p, monosomía parcial del cromosoma 15 (15q-11q13), monosomía parcial 13q, y la monosomía parcial del cromosoma 22 también pueden determinarse usando el método. El Ejemplo 6 y la Figura 16 ilustran el uso del método de la invención para determinar esa presencia de una deleción parcial del cromosoma 11.

El método de la invención también puede ser usado para determinar cualquier aneuploidía si uno de los padres es un portador conocido de dicha anomalía. Estas incluyen, pero no están limitadas a, cromosoma marcador supernumerario (SMC) pequeño; translocación t(11;14)(p15;p13); translocación desequilibrada t(8;11)(p23.2;p15.5); microdeleción 11q23; deleción 17p11.2 síndrome de Smith-Magenis; deleción 22q13.3; Microdeleción Xp22.3; 10p14 eliminación; Microdeleción 20p, el síndrome de Di-George [del (22) (q11.2q11.23)], el síndrome de Williams (deleciones 7q11.23 y 7q36); deleción 1p36; microdeleción 2p; neurofibromatosis tipo 1 (microdeleción 17q11.2), deleciónn Yq; Síndrome de Wolf-Hirschhorn (WHS, microdeleción 4p16.3); microdeleción 1p36.2; deleción 11q14; microdeleción 19q13.2; Rubinstein-Taybi (microdeleción 16 p13.3); microdeleción 7p21; síndrome de Miller-Dieker (17p13.3), deleción 17p11.2; y microdeleción 2q37.

Determinación del CNV de trastornos clínicos

Además de la determinación temprana de los defectos de nacimiento, los métodos descritos en la presente se pueden aplicar a la determinación de cualquier anomalía en la representación de las secuencias genéticas dentro del genoma. Se ha mostrado que el plasma sanguíneo y el ADN del suero de pacientes con cáncer contienen cantidades medibles de ADN tumoral, que se pueden recuperar y usadas como fuente sustituta de ADN tumoral. Los tumores están caracterizados por aneuploidía, o números inapropiados de secuencias de genes o incluso cromosomas completos. La determinación de una diferencia en la cantidad de una secuencia dad, es decir, una secuencia de interés, en una muestra de un individuo puede usarse por lo tanto en el diagnóstico de una condición médica, por ejemplo cáncer.

Las realizaciones divulgadas en la presente proporcionan un método para evaluar la variación del número de copias de una secuencia de interés, por ejemplo, una secuencia clínicamente relevante, en una muestra de ensayo que comprende una mezcla de ácidos nucleicos derivados de dos genomas diferentes, y que se sabe o se sospecha que difieren en la cantidad de una o más secuencias de interés. La mezcla de ácidos nucleicos se deriva de dos o más tipos de células. En una realización, la mezcla de ácidos nucleicos se deriva de células normales y cancerosas derivadas de un sujeto que sufre de una condición médica, por ejemplo cáncer.

Se cree que muchos tumores sólidos, como el cáncer de mama, progresan del inicio de la metástasis a través de la acumulación de varias aberraciones genéticas. [Sato et al., Cancer Res., 50: 7184-7189 [1990]; Jongsma et al., J Clin PAthol: Mol Path 55:305-309 [2002])]. Dichas aberraciones genéticas, a medida que se acumulan, pueden conferir ventajas proliferativas, inestabilidad genética y la capacidad auxiliar de desarrollar resistencia a los fármacos rápidamente, y angiogénesis, proteólisis y metástasis potenciadas. Las aberraciones genéticas pueden afectar o a los "genes supresores de tumores" recesivos o oncogenes que actúan dominantemente. Las deleciones y recombinación que llevan a la pérdida de heterocigosidad (LOH) se cree que juegan un papel principal en la progresión tumoral descubriendo alelos supresores de tumores mutados.

Se ha descubierto ADN libre de células en la circulación de pacientes diagnosticados con tumores malignos incluyendo pero no limitado a cáncer de pulmón (Pathak et al. Clin Chem 52:1833-1842 [2006]), cáncer de próstata (Schwartzenbach et al. Clin Cancer Res 15:1032-8 [2009]), y cáncer de mama (Schwartzenbach et al. disponible online en breast-cancer-research.com/content/11/5/R71 [2009]). La identificación de inestabilidades genómicas asociadas con cánceres que pueden determinarse en el ADN libre de células circulante en pacientes con cáncer es un diagnóstico potencial y herramienta de pronóstico. En una realización, el método divulgado en la presente evalúa

la CNV de una secuencia de interés en una muestra que comprende una mezcla de ácidos nucleicos derivados de un sujeto que se sospecha o se sabe tiene cáncer, por ejemplo carcinoma, sarcoma, linfoma, leucemia, tumores de células germinales y blastoma. En una realización, la muestra es una muestra de plasma derivada (procesos) de sangre periférica y que comprende una mezcla de ADN libre de células derivada de células normales y cancerosas. En otra realización, la muestra biológica que se necesita para determinar si hay una CNV presente se deriva de una mezcla de células cancerosas y no cancerosas de otros fluido biológicos que incluyen pero no están limitados a suero, sudor, lágrimas, esputo, orina, esputo, flujo del oído, linfa, saliva, líquido cefalorraquídeo, estragos, suspensión de médula ósea, flujo vaginal, lavado transcervical, fluido cerebral, ascitis, leche, secreciones de los tractos respiratorio, intestinal y genitourinario, y muestras de leucoforesis, o en biopsias de tejido, hisopos o frotis.

La secuencia de interés es una secuencia de ácidos nucleicos que se sabe o se sospecha que juega un papel en el desarrollo y/o progresión del cáncer. Los ejemplos de una secuencia de interés incluyen secuencias de ácidos nucleicos que se amplifican o eliminan en células cancerosas como se describe a continuación.

Los genes que actúan dominantemente asociados con tumores sólidos humanos ejercen típicamente su efecto por la sobre-expresión o la expresión alterada. La amplificación del gen es un mecanismo común que lleva a la regulación hacia arriba de la expresión del gen. La evidencia de estudios citogenéticos indica que tiene lugar amplificación significativa en por encima del 50% de los cánceres de mama humanos. Más notablemente, la amplificación del receptor del factor de crecimiento epidérmico humano de proto-oncogenes 2 (HER2) localizado en el cromosoma 17 (17(17q21-q22)), resulta en la sobreexpresión de los receptores HER2 en la superficie celular llevando a señalización excesiva y desregulada en el cáncer de mama y otros tumores malignos (Park et al., Clinical Breast Cancer 8:392-401 [2008]). Se ha descubierto una variedad de oncogenes que son amplificados en otros tumores malignos humanos. Ejemplos de la amplificación de oncogenes en tumores humanos incluyen amplificaciones de: c-myc en la línea celular de leucemia promielocítica HL60, y en líneas celulares de carcinoma de pulmón de células pequeñas. N-myc en neuroblastomsa primarios (etapas III y IV), líneas celulares de neuroblastoma, línea celular de retinoblastoma y tumores primarios, y líneas de carcinoma de pulmón de células pequeñas y tumores, L-myc en líneas celulares de carcinoma de pulmón de células pequeñas y tumores, c-myb en leucemia mieloide aguda y en líneas celulares de carcinoma de colon, c-erbb en célula de carcinoma epidermoide, y gliomas primarios, c-K-ras-2 en carcinomas primarios de pulmón, colon, vejiga y recto. N-ras en línea celular de carcinoma mamario Varmus H., Ann Rev Genetics 18: 553-612 (1984) [citado en Watson et al., Molecular Biology of the Gene (4ª ed.; Benjamin/Cummings Publishing Co. 1987)].

Las deleciones cromosómicas que implican genes supresores de tumores pueden jugar un papel importante en el desarrollo y progresión de tumores sólidos. El gen supresor de tumor retinoblastoma (Rb-1), localizado en el cromosoma 13q14, es el gen supresor de tumores más extensivamente caracterizado. El producto del gen Rb-1, una fosfoproteína nuclear de 105 kDa, juega aparentemente un papel importante en la regulación del ciclo celular (Howe et al., Proc Natl Acad Sci (USA) 87:5883-5887 [1990]). La expresión alterada o perdida de la proteína Rb es provocada por la inactivación de ambos alelos del gen o a través de una mutación puntual o una deleción cromosómica. Se ha descubierto que las alteraciones del gen Rb-i están presentes no solo en retinoblastomas sino también en otros tumores malignos como osteosarcomas, cáncer de tumor de células pequeñas (Rygaard et al., Cancer Res 50: 5312-5317 [1990])] y cáncer de mama. Los estudios de polimorfismos en la longitud del fragmento de restricción (RFLP) han indicado que dichos tipos de tumor han perdido frecuentemente heterocigosidad en 13q sugiriendo que uno de los alelos del gen Rb-1 ha sido perdido debido a una deleción cromosómica bruta (Bowcock et al., Am J Hum Genet, 46: 12 [1990]). Las anomalías del cromosoma 1 incluyendo duplicaciones, deleciones y translocaciones desequilibradas que implican el cromosoma 6 y otros cromosomas compañero indican que regiones del cromosoma 1, en particular 1q21-1q32 y 1p11-13, podrían albergar oncogenes o genes supresores de tumores que son patogénicamente relevantes para tanto la fase crónica como la avanzada de neoplasmas mieloproliferativos (Caramazza et al., Eur J Hematol84:191-200 [2010]). Los neoplasmas mieloproliferativos también están asociados con deleciones del cromosoma 5. La pérdida completa o deleciones intersticiales del cromosoma 5 son la anomalía cariotípica más común en síndromes mielodisplásicos (MDSs). Los pacientes de MDS del(5q)/5q aislados tienen un pronóstico más favorable que aquellos con defectos cariotípicos adicionales, que tienden a desarrollar neoplasmas mieloproliferativos (MPNs) y leucemia mieloide aguda. LA frecuencia de deleciones del cromosoma 5 desequilibradas ha llevado a la idea de que 5q alberga uno o más genes supresores de tumores que tienen papeles fundamentales en el control del crecimiento de células madre/progenitoras hematopoyéticas (HSCs/HPCs). El mapeo citogenético de regiones comúnmente eliminadas (CDRs) centradas en 5q31 y 5q32 identificó candidatos a genes supresores de tumores, incluyendo la subunidad ribosómica RPS14, el factor de transcripción Egr1/Krox20 y la proteína remodelación del citoesqueleto, alfa-catenina (Eisenmann et al., Oncogene 28:3429-3441 [2009]). Los estudios citogenéticos y de tipificación de alelos de tumores nuevos y líneas celulares de tumores han demostrado que la pérdida alélica de varias regiones distintas en el cromosoma 3p, incluyendo 3p25, 3p21-22, 3p21.3, 3p12-13 y 3p14, son las anomalías genómicas más tempranas y más frecuentes implicadas en un amplio espectro de cánceres epiteliales principales de pulmón, mama, riñón, cabeza y cuello, ovarios, cérvix, colon, páncreas, esófago, vejiga y otros órganos. Se han mapeado varios genes supresores de tumores para la región 3p del cromosoma, y se piensa que las deleciones intersticiales o hipermetilación de los promotores preceden la pérdida del 3p o del cromosoma 3 completo en el desarrollo de carcinomas (Angeloni D., Briefings Functional Genomics 6:19-39 [2007]).

Los recién nacidos y los niños con síndrome de Down (DS) presentan a menudo leucemia transitoria congénita y tienen un riesgo aumentado de leucemia mieloide aguda y leucemia linfoblástica aguda. El cromosoma 21, que alberga alrededor de 300 genes, puede estar implicado en numerosas aberraciones estructurales, por ejemplo, translocaciones, deleciones y amplificaciones, en leucemias, linfomas y tumores sólidos. La numérica somática así como las aberraciones del cromosoma 21 estructurales están sociedades con leucemias, y genes específicos incluyendo RUNX1, TMPRSS2, y TFF, que están localizados en 21q, juegan un papel en la tumorigénesis (Fonatsch C Gene Chromosomes Cancer 49:497-508 [2010]).

En una realización, el método proporciona un medio para evaluar la asociación entre la amplificación del gen y la extensión de la evolución del tumor. La correlación entra la amplificación y/o deleción y etapa o grado de un cáncer puede ser importante para la pronóstico ya que dicha información puede contribuir a la definición de un grado de tumor basado genéticamente que podría predecir mejor el curso futuro de la enfermedad con tumores más avanzados que tienen peor pronóstico. Además la información sobre la amplificación temprana y/o eventos de deleción puede ser útil al asociar esos eventos como indicadores de la progresión de la enfermedad posterior. La amplificación del gen y las deleciones como se identifican por el método se pueden asociar con otros parámetros conocidos como el grado del tumor, histología, índice de etiquetado Brd/Urd, estado hormonal, afectación ganglionar, tamaño del tumor, duración de la supervivencia y otras propiedades del tumor disponibles de estudios epidemiológicos y bioestadísticos. Por ejemplo, el ADN del tumor a ser probado por el método podría incluir hiperplasia atípica, carcinoma ductal in situ, cáncer de etapa I-III y ganglios linfáticos metastásicos para permitir la identificación de asociaciones entre amplificaciones y deleciones y etapa. Las asociaciones hechas pueden hacer posible la intervención terapéutica efectiva. Por ejemplo, las regiones amplificadas consistentemente pueden contener un gen sobre-expresado, el producto del cual puede ser capaz de ser atacado terapéuticamente (por ejemplo, la tirosina quinasa del receptor del factor de crecimiento, $p185^{HER2}$).

El método puede usarse para identificar eventos de amplificación y/o deleción que están asociados con resistencia a los fármacos determinando la variación del número de copias de ácidos nucleicos de cánceres primarios para aquellas células que se han hecho metástasis en otros sitios. Si la amplificación y/o deleción del gen es una manifestación de inestabilidad cariotípica que permite el desarrollo rápido de resistencia a fármacos, se esperará más amplificación y/o deleción en tumores primarios de pacientes quimioresistentes que en tumores en pacientes quimiosensibles. Por ejemplo, si la amplificación de genes específicos es responsable del desarrollo de resistencia a fármacos, las regiones que rodean estos genes se esperará que sean amplificadas consistentemente en las células tumorales de derrames pleurales de pacientes quimioresistentes pero no en los tumores primarios. El descubrimiento de asociaciones entre la amplificación del gen y/o deleción y el desarrollo de resistencia a fármacos puede permitir la identificación de pacientes que se beneficiarán o no de terapia adyuvante.

Determinación Simultánea de Aneuploidía y Fracción Fetal

En otra realización, el método permite la determinación simultánea de la fracción del componente de ácidos nucleicos fetal menor, es decir, la fracción fetal, en una muestra que comprende una mezcla de ácidos nucleicos fetales y maternos. En particular el método permite la determinación de la fracción de ADN libre de células aportada por un feto a la mezcla de ADN libre de células fetal y materno en una muestra materna, por ejemplo una muestra de plasma. La diferencia entre la fracción materna y fetal se determina por la aportación relativa de un alelo polimórfico derivado del genoma fetal con la aportación del alelo polimórfico correspondiente derivado del genoma materno. Las secuencias polimórficas se pueden usar en conjunción con pruebas de diagnóstico clínicamente relevantes como un control positivo para la presencia de ADN libre de células para destacar resultados de falsos negativos o falsos positivos derivados de niveles bajos de ADN libre de células por debajo del límite de identificación. El método descrito es útil en una variedad de edades gestacionales.

Las realizaciones ejemplares del método para determinar simultáneamente la fracción fetal y la presencia o ausencia de una aneuploidía se representan en las **Figuras 2-5** de la manera siguiente.

La **Figura 2** proporciona un diagrama de flujo de una realización del método de la invención **200** para determinar simultáneamente una aneuploidía fetal y la fracción fetal de ácidos nucleicos fetales en una muestra biológica materna. En el paso **210** se obtiene de un sujeto una muestra de ensayo que comprende una mezcla de ácidos nucleicos fetales y maternos. Las muestras de ensayo incluyen las muestras descritas en el paso **110** de la realización del método **100**. En algunas realizaciones, la muestra de ensayo es una muestra de sangre periférica obtenida de una hembra embarazada, por ejemplo mujer. En el paso **220** la mezcla de ácidos nucleicos presentes en la muestra está enriquecida para los ácidos nucleicos objetivo polimórficos que comprenden cada uno un sitio polimórfico. En algunas realizaciones, los ácidos nucleicos que están enriquecidos son ADN libre de células. Los ácidos nucleicos objetivo son segmentos de material genético que se sabe que comprenden al menos un sitio polimórfico. En algunas realizaciones, los ácidos nucleicos objetivo comprenden un SNP. En otras realizaciones, el ácido nucleico objetivo comprende una STR. En otras realizaciones, los ácidos nucleicos objetivo comprenden una STR en tándem. EL enriquecimiento de una mezcla de ácidos nucleicos fetales y maternos comprende amplificar secuencias objetivo de una porción de ácidos nucleicos contenidos en la muestra materna original, y combinar parte

de o el producto amplificado completo con el resto de la muestra materna original. En el paso **230**, se secuencia al menos una porción de la mezcla enriquecida, se identifican las diferencias de secuencia derivadas de la naturaleza polimórfica de las secuencias objetivo, y la contribución relativa de las de las secuencias polimórficas derivadas del genoma fetal, es decir la fracción fetal, se determina en el paso **240**. En algunas realizaciones, la muestra de ensayo materna original es una muestra de fluido biológico, por ejemplo plasma. En otras realizaciones, la muestra materna original es una fracción procesada de plasma que comprende ADN libre de células fetal y materno purificado.

Secuencias polimórficas

Los sitios polimórficos que están contenidos en los ácidos nucleicos objetivo incluyen sin limitación polimorfismos de nucleótido simples (SNPs) SNPs en tándem, deleciones o inserciones multi-base a pequeña escala, denominadas IN-DELS (también denominadas polimorfismos de deleción inserción o DIPs), Polimorfismos Multi-Nucleótido (MNPs) y Repeticiones Cortas en Tándem (STRs). Los sitios polimórficos que están abarcados por el método de la invención están localizados en cromosomas autosómicos, permitiendo de esta manera la determinación de la fracción fetal independientemente del sexo del feto. Cualquier sitio polimórfico que pueda ser abarcado por las lecturas generadas por los métodos de secuenciación descritos en la presente se pueden usar para determinar simultáneamente la fracción fetal y la presencia o ausencia de una aneuploidía en una muestra materna.

En una realización, la mezcla de ácidos nucleicos fetales y maternos en la muestra está enriquecida para los ácidos nucleicos que comprenden al menos un SNP. En algunas realizaciones, cada ácido nucleico objetivo comprende una única SNP, es decir una. Las secuencias de ácidos nucleicos objetivo que comprenden SNPs están disponibles en bases de datos públicamente accesibles incluyendo, pero no limitado a la Human SNP Database en la dirección web de la red mundial wi.mit.edu, la NCBI dbSNP Home Page en la dirección web de la red mundial ncbi.nlm.nih.gov, la dirección web de la red mundial lifesciences.perkinelmer.com, la base de datos Celera Human SNP en la dirección web de la red mundial celera.com, la base de datos de SNP del Genome Analysis Group (GAN) en la dirección web de la red mundial gan.iarc.fr. En una realización, los SNPs elegidos para enriquecer el ADN libre de células fetal y materno se seleccionan del grupo de 92SNPs de identificación de individuos (IISNPs) descrito por Pakstis et al. (Pakstis et al. Hum Genet 127:315-324 [2010]), que ha demostrado tener una variación muy pequeña en la frecuencia a través de las poblaciones (F_{st}z0,06), y ser altamente informativa en todo el mundo teniendo una heterocigosidad media de ≥0,4. Los SNPs que están abarcados por el método de la invención incluyen SNPs ligados y no ligados. Cada ácido nucleico objetivo comprende al menos un sitio polimórfico, por ejemplo un SNP individual, que difiere del presente en otro ácido nucleico objetivo para generar un panel de sitios polimórficos, por ejemplo SNPs, que contienen un número suficientes de sitios polimórficos de los cuales al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40 o más son informativos. Por ejemplo, un panel de SNPs puede ser configurado para comprender al menos un SNP informativo.

En una realización, los SNPs que son objetivos para la amplificación se seleccionan de rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022.

En otras realizaciones, cada ácido nucleico objetivo comprende dos o más SNPs, es decir cada ácido nucleico objetivo comprende SNPs en tándem. Preferiblemente, cada ácido nucleico objetivo comprende dos SNPs en tándem. Los SNPs en tándem son analizados como una única unidad como haplotipos cortos, y se proporcionan en la presente como conjuntos de dos SNPs. Para identificar secuencias de SNP en tándem adecuadas, se puede buscar la base de datos International HapMap Consortium (The International HapMap Project, Nature 426:789-796 [2003]). La base de datos está disponible en la red mundial en hapmap.org. En una realización, los SNPs en tándem que son objetivo para la amplificación se seleccionan de los siguientes conjuntos de parejas de tándem de SNPs rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980- rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986- rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672.

En otra realización, la mezcla de ácidos nucleicos fetales y maternos en la muestra está enriquecida para los ácidos nucleicos objetivo que comprenden al menos una STR. Los loci de STR se encuentran en casi caulquier

cromosoma en el genoma y pueden ser amplificados usando una variedad de cebadores de reacción en cadena de polimerasa (PCR). Se han preferido las repeticiones de tetranucleótidos entre los científicos forenses debido a su fidelidad en la amplificación por PCR aunque también se usan algunas repeticiones de tri- y pentanucleótidos. Un listado completo de referencias, hechos e información de secuencias sobre SRTs, cebadores de PCR publicados, sistemas múltiplex comunes, y datos de poblaciones relacionados están recopilados en la STRBase, a la que se puede acceder a través de la red mundial en ibm4.carb.nist.gov:8800/dna/home.htm. La información de secuencia del GenBank® (http://www2.ncbi.nlm.nih.gov/cgi-bin/genbank) para loci de STR usados comúnmente también es accesible a través de la STRBase. La naturaleza polimórfica de las secuencias de ADN repetidas en tándem que están ampliamente extendidas a través del genoma humano las han convertido en marcadores genéticos importantes para estudios de mapeado de genes, análisis de ligamiento, y pruebas de identidad humana. Debido al alto polimorfismo de las STRs, la mayoría de los individuos serán heterocigóticos, es decir la mayoría de la gente poseerá dos alelos (versiones) de cada uno heredado de cada padre con un número diferente de repeticiones. Por lo tanto, la secuencia de STR fetal heredad no maternalmente diferirá en el número de repeticiones de la secuencia materna. La amplificación de estas secuencias de STR resultará en dos productos de la amplificación principales correspondientes a los alelos maternos (y el alelo fetal heredado maternalmente) y un producto secundario correspondiente al alelo fetal no maternalmente heredado. Se informó primero de esta técnica en el 2000 (Pertl et al., Human Genetics 106:45-49 [2002]) y ha sido desarrollada posteriormente usando identificación simultanea de múltiples regiones diferentes de SRT usando PCR en tiempo real (Liu et al., Acta Obset Gyn Scand 86:535-541 [2007]). Por lo tanto, la fracción de ácido nucleico fetal en una muestra materna también puede determinarse secuenciando polimórficamente los ácidos nucleicos objetivo que comprenden STRs, que varían entre individuos en el número de unidades repetidas en tándem entre alelos. En una realización, la determinación simultánea de aneuploidía y fracción fetal comprende secuenciar al menos una porción de ácidos nucleicos fetales y maternos presentes en una muestra materna que se ha enriquecido para secuencias polimórficas que comprende STRs. Dado que el tamaño del ADN libre de células fetal es <300 bp, las secuencias polimórficas comprenden miniSTR, que pueden ser amplificadas para generar amplicones que son de longitudes de alrededor del tamaño de los fragmentos de ADN fetal circulante. El método puede usar uno o una combinación de cualquier número de miniSTRs informativos para determinar la fracción del ácido nucleico fetal. Por ejemplo, cualquiera o una combinación de cualquier número de miniSTRs, por ejemplo se pueden usar los miniSTRs divulgados en la Tabla 22. En una realización, la fracción de ácido nucleico fetal en una muestra materna se realiza usando un método que incluye determinar el número de copias del ácido nucleico materno y fetal presente en la muestra materna amplificando al menos un miniSTR autosómico elegido de CSF1PO, FGA, TH01, TPOX, vWA, D3S1358,D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, D2S1338, D1S1677, D2S441, D4S2364, D10S1248, D14S1434, D22S1045, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, y D1GATA113. En otra realización, el al menos un miniSTR autosómico es el grupo de mini STRs CSF1PO, FGA, D13S317, D16S539, D18S51, D2S1338, D21S11 y D7S820.

El enriquecimiento de la muestra para los ácidos nucleicos objetivo se consigue por métodos que comprenden amplificar específicamente secuencias de ácidos nucleicos objetivo que comprenden el sitio polimórfico. La amplificación de las secuencias objetivo puede realizarse por cualquier método que use PCR o variaciones del método incluyendo, pero no limitadas a, PCR asimétrico, amplificación dependiente de la helicasa, PCR hot-start, qPCR, PCR en fase sólida y PCR touchdown. Alternativamente, la replicación de las secuencias de ácidos nucleicos objetivo puede obtenerse por métodos independientes de enzimas, por ejemplo síntesis en fase sólida química usando las fosforamiditas. La amplificación de las secuencias objetivo se consigue usando pares de cebadores cada uno capaz de amplificar una secuencia de ácido nucleico objetico que comprende el sitio polimórfico, por ejemplo SNP, en una reacción de PCR multiplex. Las reacciones de PCR multiplex incluyen combinar al menos 2, al menos tres, al menos 3, al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30 al menos 30, al menos 35, al menos 40 o más conjuntos de cebadores en la misma reacción para cuantificar los ácidos nucleicos objetivo amplificados que comprenden al menos dos, al menos tres, al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 30, al menos 35, al menos 40 o más sitios polimórficos en la misma reacción de secuenciación. Cualquier panel de conjuntos de cebadores se puede configurar para amplificar al menos una secuencia polimórfica informativa.

Amplificación de secuencias polimórficas

Un número de cebadores de ácidos nucleicos están ya disponibles para amplificar fragmentos de ADN que contienen los polimorfismos de SNP y sus secuencias se pueden obtener, por ejemplo, de las bases de datos anteriormente identificadas. También se pueden diseñar cebadores adicionales, por ejemplo, usando un método similar al publicado por Vieux, E. F., Kwok, P-Y and Miller, R. D. en BioTechniques (Junio 2002) Vol. 32. Suplemento: "SNPs: Discovery of Marker Disease, pp. 28-32. En una realización, se elige al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40 o más conjuntos de cebadores para amplificar un ácido nucleico objetivo que comprende al menos un SNPs informativo en una porción de una mezcla de ADN libre de

células fetal y materno. En una realización, los conjuntos de cebadores comprenden cebadores directos e inversos que abarcan al menos un SNP informativo seleccionado de rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022. Conjuntos ejemplares de cebadores que se usan para amplificar los SNPs divulgados en la presente se proporcionan en el Ejemplo 7 y las Tablas 10 y 11, y se divulgan como SEQ ID NOs:57-112. En otra realización, el grupo de 13 conjuntos de cebadores SEQ ID NOs:57-82 se usa para amplificar un ácido nucleico objetivo cada uno comprendiendo al menos un SNP, por ejemplo un SNP individual, en una porción de una mezcla de ADN libre de células fetal o materno.

En otra realización, se usa al menos un conjunto de cebadores para amplificar un ácido nucleico objetivo comprendiendo cada uno al menos un SNP en tándem, por ejemplo un conjunto de dos SNPs en tándem, en una porción de una mezcla de ADN libre de células fetal y materno. En una realización, los conjuntos son de cebadores que comprenden cebadores directos e inversos que abarcan al menos un SNP en tándem informativo seleccionado de rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980- rs987981; rs4143392- rs4143391; rs1691324- rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102- rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 -rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672. Los cebadores usados para amplificar las secuencias objetivos que comprenden los SNPs en tándem están diseñados para abarcar ambos sitios de SNP. Conjuntos ejemplares de cebadores usados para amplificar los SNPs en tándem divulgados en la presente se proporcionan en el Ejemplo 12 y se divulgan como SEQ ID NOs:197-310.

La amplificación de ácidos nucleicos objetivo se realiza usando cebadores específicos de la secuencia que permiten la amplificación específica de la secuencia. Por ejemplo, los cebadores de PCR están diseñados para discriminar contra la amplificación de genes o parálogos similares que están en otros cromosomas tomando ventaja de las diferencias de secuencia entre el ácido nucleico objetivo y cualquier parálogo de otros cromosomas. Los cebadores de PCR directos o inversos están diseñados para hibridar cerca del sitio de SNP y para amplificar una secuencia de ácidos nucleicos de longitud suficiente para ser abarcada en las lecturas generadas por métodos de secuenciación masivamente paralelos. Algunos métodos de secuenciación masivamente paralelos requieren que la secuencia de ácidos nucleicos tenga una longitud mínima (bp) para permitir la amplificación por puente que puede usarse antes de la secuenciación. Por lo tanto, los cebadores de PCR usados para amplificar ácidos nucleicos objetivo están diseñados para amplificar secuencias que son de longitud suficiente para ser amplificadas por puente y para identificar SNPs que están abarcados por las lecturas de secuencia. En algunas realizaciones, el primero de dos cebadores en el conjunto de cebadores que comprende el cebador directo y el inverso para amplificar el ácido nucleico objetico está diseñado para identificar un SNP individual presente dentro de una lectura de secuencia de alrededor de 20bp, alrededor de 25bp, alrededor de 30bp, alrededor de 35bp, alrededor de 40bp, alrededor de 45bp, alrededor de 50bp, alrededor de 55bp, alrededor de 60bp, alrededor de 65bp, alrededor de 70bp, alrededor de 75bp, alrededor de 80bp, alrededor de 85bp, alrededor de90bp, alrededor de 95bp, alrededor de 100bp, alrededor de 110bp, alrededor de 120bp, alrededor de 130, alrededor de 140bp, alrededor de 150bp, alrededor de 200bp, alrededor de 250bp, alrededor de 300bp, alrededor de 350bp, alrededor de 400bp, alrededor de 450bp, o alrededor de 500bp. Se espera que los avances tecnológicos en tecnologías de secuenciación masivamente paralelas permitan lecturas de extremos individuales mayores de 500bp. En una realización, uno de los cebadores de PCR está diseñado para amplificar SNPs que están abarcados en lecturas de secuencia de 36 bp. El segundo cebador está diseñado para amplificar el ácido nucleico objetivo como un amplicón de longitud suficiente para permitir la amplificación por puente. en una realización, los cebadores de PCR ejemplares están diseñados para amplificar ácidos nucleicos objetivo que contienen un único SNP seleccionado de los SNPs rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022. En otras realizaciones, los cebadores directo e inverso están cada uno diseñados para amplificar ácidos nucleicos objetivo que comprenden cada uno un conjunto de dos SNPs en tándem, cada uno estando presente dentro de una lectura de secuencia de alrededor de 20bp, alrededor de 25bp, alrededor de 30bp, alrededor de 35bp, alrededor de 40bp, alrededor de 45bp, alrededor de 50bp, alrededor de 55bp, alrededor de 60bp, alrededor de 65bp, alrededor de 70bp, alrededor de 75bp, alrededor de 80bp, alrededor de 85bp, alrededor de90bp, alrededor de 95bp, alrededor de 100bp, alrededor de 110bp, alrededor de 120bp, alrededor de 130, alrededor de 140bp, alrededor de 150bp, alrededor de 200bp, alrededor de 250bp, alrededor de 300bp, alrededor de 350bp, alrededor de 400bp, alrededor de 450bp,o alrededor de 500bp. En una realización, al menos uno de los cebadores está diseñado para amplificar el ácido nucleico objetivo que comprende

un conjunto de dos SNPs en tándem como un amplicón de longitud suficiente para permitir la amplificación por puente.

Los SNPs, Los SNPs individuales o en tándem, están contenidos en amplicones de ácidos nucleicos objetivo amplificados de al menos alrededor de 100bp, al menos alrededor de 150bp, al menos alrededor de 200bp, al menos alrededor de250bp, al menos alrededor de 300bp, al menos alrededor de 350bp, o al menos alrededor de 400bp. En una realización, los ácidos nucleicos objetivo que comprenden un sitio polimórfico, por ejemplo un SNP, son amplificados como amplicones de al menos alrededor de 110 bp, y que comprenden un SNP dentro de 36 bp del extremo 3' o 5' del amplicón. En otra realización, los ácidos nucleicos objetivo que comprenden dos o más sitios polimórficos, por ejemplo dos SNPs en tándem, son amplificados como amplicones de al menos alrededor de 110 bp, y que comprenden el primer SNP dentro de 36 bp del extremo 3' del amplicón, y/o el segundo SNP dentro de 36 bp del extremo 5' del amplicón.

En una realización, se eligen al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40 o más conjuntos de cebadores para amplificar un ácido nucleico objetivo que comprende al menos un SNP en tándem informativo en una porción de una mezcla de ADN libre de células fetal o materno.

Amplificación de STRs

Hay disponible una variedad de cebadores de ácidos nucleicos para amplificar fragmentos de ADN que contienen las STRs y sus secuencias se pueden obtener, por ejemplo, de las bases de datos identificadas anteriormente. Se han usado varios amplicones de PCR dimensionados para distinguir las distribuciones de tamaño respectivas de las especies de ADN fetal y materno, y han mostrado que las moléculas de ADN fetal en el plasma de mujeres embarazadas son generalmente más cortas que las moléculas de ADN materno (Chan et al., Clin Chem 50:8892 [2004]). El fraccionamiento del tamaño del ADN fetal circulante ha confirmado que la longitud media de los fragmentos de ADN fetal circulante es <300 bp, mientras que se ha estimado que el del ADN materno es de entre alrededor de 0,5 y 1Kb (Li et al., Clin Chem, 50: 1002-1011 [2004]). Estos descubrimientos son consistentes con los de Fan et al., que determinó usando NGS que el ADN libre de células fetal raramente es >340bp (Fan et al., Clin Chem 56:1279-1286 [2010]). El método de la invención abarca determinar la fracción de ácido nucleico fetal en una muestra materna que ha sido enriquecido con ácidos nucleicos objetivo cada uno comprendiendo una miniSTR que comprende cuantificar al menos un alelo fetal y uno materno en una miniSTR polimórfica, que puede ser amplificado para generar amplicones que son de longitudes de alrededor del tamaño de los fragmentos de ADN fetal circulante.

En una realización, el método comprende determinar el número de copias de al menos un alelo fetal y al menos uno materno al menos en una miniSTR polimórfica que se amplifica para generar amplicones que son menores de alrededor 300 bp, menores de alrededor 250 bp, menores de alrededor 200 bp, menores de alrededor 150 bp, menores de alrededor 100 bp o menores de alrededor 50 bp. En otra realización, los amplicones que se generan amplificando las miniSTRs son menores de alrededor de 300 bp. En otra realización, los amplicones que se generan amplificando las miniSTRs son menores de alrededor de 250 bp. En otra realización, los amplicones que se generan amplificando las miniSTRs son menores de alrededor de 200 bp. La amplificación del alelo informativo incluye usar cebadores de miniSTR, que permiten la amplificación de amplicones de tamaño reducido para distinguir alelos de STR que son menores de alrededor de 500 bp, menores de alrededor de 450 bp, menores de alrededor de 400 bp, menores de alrededor de 350 bp, menores de alrededor de 300 pares de bases (bp), menores de alrededor de 250 bp, menores de alrededor de 200 bp, menores de alrededor de 150 bp, menores de alrededor de 100 bp o menores de alrededor de 50 bp. Los amplicones de tamaño reducido generados usando los cebadores de miniSTR son conocidos como miniSTRs que se identifican de acuerdo con el nombre del marcador correspondiente al locus para el que han sido mapeados. En una realización, los cebadores de miniSTR incluyen cebadores de miniSTR que han permitido la máxima reducción de tamaño en el tamaño del amplicón para todos los 13 loci CODIS STR además de D2S1338, Penta D, y pentaE encontrados en los kits de STR comercialmente disponibles (Butler et al., J Forensic Sci 48:1054-1064 [2003]), los loci de miniSTR que no están ligados a los marcadores CODIS se describen por Coble y Butler (Coble and Butler, J Forensic Sci 50:43-53 [2005]), y otros miniSTRs se han caracterizado en el NIST. La información referente a los miniSTRs caracterizados en el NIST es accesible a través de la red mundial en cstl.nist.gov/biotech/strbase/newSTRs.htm. Cualquier par o una combinación de dos o más pares de cebadores de miniSTR se pueden usar para amplificar al menos un miniSTR. Por ejemplo, se selecciona al menos un conjunto de cebadores de los conjuntos de cebadores proporcionados en la Tabla 22 (Ejemplo 11) y divulgados como SEQ ID NOs:113-196 puede usarse para amplificar secuencias objetivo polimórficas que comprenden un STR.

El enriquecimiento de la muestra se obtiene amplificando ácidos nucleicos objetivo contenidos en una porción de la mezcla de ácidos nucleicos fetales y maternos en la muestra original, y combinando al menos una porción o todo el producto amplificado con el resto de la muestra no amplificada original. El enriquecimiento comprende amplificar los ácidos nucleicos objetivo que están contenidos en una porción de la muestra de fluido biológico. En una realización, la muestra que es enriquecida es la fracción de plasma de una muestra de sangre (ver **Figura 3**). Por ejemplo, una porción de una muestra de plasma materna original se usa para amplificar secuencias de ácidos nucleicos objetivo.

Posteriormente, algo o todo el producto amplificado se combina con el resto de la muestra de plasma original no amplificada enriqueciéndola de esta manera (ver Ejemplo 10). En otra realización, la muestra que se enriquece en la muestra de ADN libre de células purificada que se extrae del plasma (ver **Figura 4**). Por ejemplo, el enriquecimiento comprende amplificar los ácidos nucleicos objetivo que están contenidos en una porción de una muestra original de mezcla purificada de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células que ha sido purificada de una muestra de plasma materno, y combinar posteriormente algo o todo el producto amplificado con el resto de la muestra purificada original no amplificada (ver Ejemplo 9). En otra realización, la muestra que es enriquecida es una muestra de biblioteca de secuenciación preparada de una mezcla purificada de ácidos nucleicos fetales y maternos (ver **Figura 5**). Por ejemplo, el enriquecimiento comprende amplificar los ácidos nucleicos objetivo que están contenidos en una porción de una muestra original de mezcla purificada de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células, que ha sido purificada de una muestra de plasma materno, preparar una primera biblioteca de secuenciación de secuencias de ácidos nucleicos no amplificados, preparar una segunda biblioteca de secuenciación de ácidos nucleicos polimórficos amplificados, y posteriormente combinar algo o toda la segunda biblioteca de secuenciación con algo o todo de la primera biblioteca de secuenciación (ver Ejemplo 8). La cantidad de producto amplificado que se usa para enriquecer la muestra original se selecciona para obtener suficiente información de secuenciación para determinar tanto la presencia como la ausencia de aneuploidía y la fracción fetal de la misma ejecución de secuenciación. Al menos alrededor del 3%, al menos alrededor del 5%, al menos alrededor del 7%, al menos alrededor del 10%, al menos alrededor del 15%, al menos alrededor del 20%, al menos alrededor del 25%, al menos alrededor del 30% o más del número total de etiquetas de secuencia obtenidas de la secuenciación se mapean para determinar la fracción fetal.

En una realización, el paso de enriquecer la mezcla de ácidos nucleicos fetales y maternos para ácidos nucleicos objetivo polimórficos comprende amplificar los ácidos nucleicos objetivo en una porción de una muestra de ensayo, por ejemplo una muestra de ensayo de plasma, y combinar todo o una porción del producto amplificado con el resto de la muestra de ensayo de plasma. La realización del método **300** se representa en el diagrama de flujo proporcionado en la **Figura 3**. En el paso **310**, se obtiene una muestra de ensayo, por ejemplo una muestra de fluido biológico como una muestra de sangre, de una mujer embarazada, y en el paso **320** se usa una porción del ADN libre de células contenido en la fracción de plasma de la muestra de sangre para amplificar ácidos nucleicos objetivo que comprenden sitios polimórficos, por ejemplo SNPs. En una realización, al menos alrededor del 1%, al menos alrededor del 1,5%, al menos alrededor del 2%, al menos alrededor del 10% del plasma materno se usó para amplificar los ácidos nucleicos objetivo. En el paso **330**, una porción o todos los ácidos nucleicos objetivo amplificados se combina con la mezcla de ADN libre de células fetal y materno presente en la muestra materna, y el ADN libre de células combinado y los ácidos nucleicos amplificados se purifican en el paso **340**, y se usan para preparar una biblioteca que se secuenció en el paso **350**. La biblioteca se purificó de ADN libre de células purificado y comprende al menos alrededor del 10%, al menos alrededor del 15%, al menos alrededor del 20%, al menos alrededor del 25% , al menos alrededor del 30% , al menos alrededor del 35% , al menos alrededor del 40% , al menos alrededor del 45%, o al menos alrededor del 50% del producto amplificado. En el paso **360**, los datos de las ejecuciones de secuenciación se analizan y se hace la determinación simultánea de la fracción fetal y la presencia o ausencia de aneuploidía.

En una realización, el paso de enriquecer al mezcla de ácidos nucleicos fetales y maternos para los ácidos nucleicos objetivo polimórficos comprende una pluralidad de ácidos nucleicos objetivo polimórficos en una porción de una mezcla de ácidos nucleicos fetales y maternos purificados de una muestra de ensayo materna. En una realización, una porción de una mezcla de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células, purificada de una muestra de plasma materno se usa para amplificar las secuencias de ácidos nucleicos polimórficos, y una porción del producto amplificado se combina con la mezcla no amplificada o ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células (ver **Figura 4**). La realización del método **400** se representa en el diagrama de flujo proporcionado en la **Figura 4**. En el paso **410**, se obtiene de una mujer embarazada una muestra de ensayo, por ejemplo una muestra de fluido biológico como una muestra de sangre, que comprende una mezcla de ácidos nucleicos fetales y maternos, y la mezcla de ácidos nucleicos fetales y maternos se purifica de la fracción de plasma en el paso **420**. Como se ha descrito anteriormente, los métodos para la separación de ADN libre de células del plasmo son bien conocidos. En el paso **430**, una porción del ADN libre de células contenido en la muestra purificada se usa para amplificar ácidos nucleicos objetivo que comprenden sitios polimórficos, por ejemplo SNPs. Al menos alrededor del 5%, al menos alrededor del 10%, al menos alrededor del 15%, al menos alrededor del 20% , al menos alrededor del 25% , al menos alrededor del 30% , al menos alrededor del 35% , al menos alrededor del 40% , al menos alrededor del 45%, o al menos alrededor del 50% del ADN libre de células purificado se usa para amplificar los ácidos nucleicos objetivo. Preferiblemente, la amplificación de las secuencias objetivo se puede realizar por cualquier método que use PCR o variaciones del método incluyendo pero no limitado a PCR asimétrico, amplificación dependiente de la helicasa, CR hot-start, qPCR, PCR en fase sólida y PCR touchdown. En el paso **440**, una porción, por ejemplo al menos alrededor del 0,01% del producto amplificado se combina con la muestra de ADN libre de células purificado no amplificado, y la mezcla de los ácidos nucleicos fetales y maternos amplificados y no amplificados se secuencia en el paso **450**. En una realización, la secuenciación se realiza usando cualquiera del as tecnologías NGS. En el paso **460**, los datos de las ejecuciones de secuenciación se analizan y se hace la determinación simultánea de la fracción fetal y la presencia o ausencia de aneuploidía como se describe en el paso **140** de la realización representada en la **Figura 1**.

En otra realización, el paso **220** de enriquecer la mezcla de ácidos nucleicos fetales y maternos para ácidos nucleicos objetivo polimórficos (**Figura 2**) comprende combinar al menos una porción de una primera biblioteca de secuenciación de moléculas de ácidos nucleicos fetales y maternos no amplificados con al menos una porción de una segunda biblioteca de secuenciación de ácidos nucleicos objetivo polimórficos. Por lo tanto, la muestra que es enriquecida es la muestra de la biblioteca que se prepara para la secuenciación (**Figura 5**). El enriquecimiento de la muestra de la biblioteca para los ácidos nucleicos objetivo se consigue por métodos que comprenden amplificar específicamente las secuencias de ácidos nucleicos que comprenden el sitio polimórfico. En el paso **510**, se obtiene de una mujer embarazada una muestra de ensayo, por ejemplo una muestra de fluido biológico como una muestra de sangre, que comprende una mezcla de ácidos nucleicos fetales y maternos, y la mezcla de ácidos nucleicos fetales y maternos se purifica de la fracción de plasma en el paso **520**. En el paso **530**, se usa una porción del ADN libre de células contenido en la muestra purificada para amplificar ácidos nucleicos objetivo que comprenden sitios polimórficos, por ejemplo SNPs. Al menos alrededor del 5%, al menos alrededor del 10%, al menos alrededor del 15%, al menos alrededor del 20%, al menos alrededor del 25%, o al menos alrededor del 30% del ADN libre de células se usa para amplificar secuencias de ácidos nucleicos objetivo. Preferiblemente, la amplificación de las secuencias objetivo puede realizarse por cualquier método que use PCR o variaciones del método incluyendo, pero no limitadas a, PCR asimétrico, amplificación dependiente de la helicasa, PCR hot-start, qPCR, PCR en fase sólida y PCR touchdown. En el paso **540**, los ácidos nucleicos objetivo amplificados que comprenden los sitios polimórficos, por ejemplo SNPs, se usan para preparar una biblioteca de secuenciación de ácidos nucleicos objetivo. De manera similar, la porción de ADN libre de células no amplificado purificado se usa para preparar una biblioteca de secuenciación primaria en el paso **550**. En el paso **560**, una porción de la biblioteca objetivo se combina con la biblioteca primaria generada de la mezcla no amplificada de ácidos nucleicos, y la mezcla de ácidos nucleicos fetales y maternos comprendida en las dos bibliotecas se secuencia en el paso **570**. La biblioteca enriquecida comprende al menos alrededor del 5%, al menos alrededor del 10%, al menos alrededor del 15%, al menos alrededor del 20%, o al menos alrededor del 25% de la biblioteca objetivo. En el paso **580**, los datos de las ejecuciones de secuenciación se analizan y se hace la determinación simultánea de la fracción fetal y la presencia o ausencia de aneuploidía como se describe en el paso **140** de la realización representada en la Figura 1.

Determinación de Aneuploidía a partir de Bibliotecas Enriquecidas de Secuenciación

La presencia o ausencia de aneuploidía se determina de la secuenciación de la biblioteca enriquecida para secuencias objetivo polimórficas como se describe para la biblioteca no enriquecida descrita en el método **100**.

Determinación de la Fracción Fetal de Bibliotecas Enriquecidas de Secuenciación

La determinación de la fracción fetal en los pasos **240 (Figura 2), 360 (Figura 3), 480 (Figura 4)** y **580 (Figura 5)** se basa en el número total de etiquetas que mapean para el primer alelo y el número total de etiquetas que mapean para el segundo alelo en un sitio polimórfico informativo, por ejemplo un SNP, contenido en un genoma de referencia. Por ejemplo, el genoma de referencia es la secuencia NCBI36/hg18 del genoma de referencia humano, o el genoma de referencia comprende la secuencia NCBI36/hg18 del genoma de referencia humano y un genoma de secuencias objetivo artificiales, lo que incluye las secuencias polimórficas objetivo. En una realización el genoma objetivo artificial abarca secuencias polimórficas que comprende los SNPs rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, y rs530022. En una realización, el genoma artificial incluye las secuencias objetivo polimórficas de las SEQ ID NOs: 1-56. En otra realización, el genoma artificial incluye las secuencias objetivo polimórficas de las SEQ ID NOs:1-26 (ver ejemplo 7). En otra realización, el genoma objetivo artificial abarca secuencias polimórficas que comprenden STRs seleccionados de CSF1PO, FGA, TH01, TPOX, vWA, D3S1358,D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, D2S1338, D1S1677, D2S441, D4S2364, D10S1248, D14S1434, D22S1045, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, y D1GATA113. En todavía otra realización, el genoma objetivo artificial abarca secuencias polimórficas que comprenden uno o más SNPs en tándem (SEQ ID NOs: 1-56). La composición del genoma de las secuencias objetivo artificiales variará dependiendo de las secuencias polimórficas que se usen para determinar la fracción fetal. Por consiguiente, un genoma de secuencias objetivo artificiales no está limitado a las secuencias de SNP o STR ejemplificadas en la presente.

El sitio polimórfico informativo, por ejemplo SNP, se identifica por la diferencia en las secuencias alélicas y la cantidad de cada uno de los alelos posibles. El ADN libre de células está presente a una concentración que es <10% del ADN libre de células materno. Por lo tanto, la presencia de una contribución menor de un alelo a la mezcla de ácidos nucleicos fetales y maternos en relación con la contribución principal del alelo materno puede ser asignada al feto. Los alelos que se derivan del genoma materno son referidos en la presente como alelos principales, y los alelos que se derivan del genoma fetal son referidos en la presente como alelos menores. Los alelos que se representan por niveles similares de etiquetas de secuencias mapeadas representan alelos maternos. Los

resultados de una amplificación multiplex ejemplar de ácidos nucleicos objetivo que comprenden SNPs y derivados de una muestra de plasma materno se muestran en la **Figura 18**. Los SNPs informativos son distinguidos del cambio del nucleótido individual en un sitio polimórfico predeterminado, y los alelos fetales se distinguen por su contribución menor relativa a la mezcla de ácidos nucleicos fetales y maternos en la muestra cuando se compara con la contribución mayor a la mezcla por los ácidos nucleicos maternos, es decir las secuencias de SNP son informativas cuando la madre es heterocigótica y hay presente un tercer alelo paternal, permitiendo una comparación cuantitativa entre el alelo heredado maternalmente y el alelo heredado paternalmente para calcular la fracción fetal. Por consiguiente, la abundancia relativa de ADN libre de células fetal en la muestra materna se determina como un parámetro del número total de etiquetas de secuencia únicas mapeadas para la secuencia de ácido nucleicos objetivo en un genoma de referencia para cada uno de los dos alelos del sitio polimórfico predeterminado. En una realización, la fracción de ácidos nucleicos fetales en la mezcla de ácidos nucleicos fetales y maternos se calcula para cada uno de los alelos informativos ($alelo_x$) de la manera siguiente:

$$\%\ \text{del alelo}_x\ \text{de la fracción fetal} = ((\Sigma\text{etiquetas de secuencia fetal para el alelo}_x)/(\Sigma\text{etiquetas de secuencia materna para el alelo}_x)) \times 100$$

y la fracción fetal para la muestra se calcula como la media de la fracción fetal de todos los alelos informativos.

Opcionalmente, la fracción de los ácidos nucleicos fetales en la mezcla de ácidos nucleicos fetales y maternos se calcula para cada uno de los alelos informativos ($alelo_x$) de la manera siguiente:

$$\%\ \text{del alelo}_x\ \text{de la fracción fetal} = ((2\text{X}\Sigma\text{etiquetas de secuencia fetal para el alelo}_x) / (\Sigma\text{etiquetas de secuencia materna para el alelo}_x)) \times 100$$

para compensar por la presencia de 2 alelos fetales, uno siendo enmascarado por el fondo materno.

El porcentaje de fracción fetal se calcula para al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, o más alelos informativos. En una realización, la fracción fetal es la fracción fetal media determinada para al menos 3 alelos informativos.

De manera similar, la fracción fetal se puede calcular del número de etiquetas mapeadas para alelos SNO en tándem como se hace para SNPs individuales, pero teniendo en cuenta las etiquetas mapeadas para los dos alelos de SNP en tándem x e y presentes en cada una de las secuencias de ácidos nucleicos objetivo polimórficas que se amplifican para enriquecer las muestras, es decir:

$$\%\ \text{del alelo}_{x+y}\ \text{de la fracción fetal} = ((\Sigma\text{etiquetas de secuencia fetal para el alelo}_{x+y}) / (\Sigma\text{etiquetas de secuencia materna para el alelo}_{x+y})) \times 100$$

Opcionalmente, la fracción de ácidos nucleicos fetales en la mezcla de ácidos nucleicos fetales y maternos se calcula para cada uno de los alelos informativos ($alelo_{x+y}$), de la manera siguiente:

$$\%\ \text{del alelo}_{x+y}\ \text{de la fracción fetal} = ((2\text{X}\Sigma\text{etiquetas de secuencia fetal para el alelo}_{x+y}) / (\Sigma\text{etiquetas de secuencia materna para el alelo}_{x+y})) \times 100$$

para compensar por la presencia de 2 conjuntos de alelos fetales en tándem, uno estando enmascarado por el fondo materno. Las secuencias de SNP en tándem son informativas cuando la madre es heterocigótica y hay presente un tercer haplotipo paterno, permitiendo una comparación cuantitativa entre el haplotipo heredado maternalmente y el haplotipo heredado paternalmente para calcular la fracción fetal.

La fracción fetal puede ser determinada de bibliotecas de secuenciación que comprenden secuencias objetivo polimórficas amplificadas que comprenden STRS contando el número de etiquetas mapeadas para un alelo principal (materno) y uno menor (fetal). Las etiquetas comprenden secuencias de longitud suficiente para abarcar los alelos de STR. Los alelos de STR informativos pueden resultar en una o dos secuencias de etiquetas principales correspondientes a los alelos maternos (y el alelo fetal heredado maternalmente) y una secuencia de etiqueta menor correspondiente al alelo fetal heredado no maternalmente. La fracción fetal se calcula como una proporción del número de etiquetas mapeadas para los alelos fetales y maternos.

Determinación de la fracción fetal por secuenciación masivamente paralela

Además de usar el presente método para determinar simultáneamente la fracción fetal y aneuploidía, la fracción fetal puede ser determinada independientemente de la determinación de la aneuploidía como se describe en la presente, pero puede determinarse independientemente y/o en conjunción con otros métodos usados para la determinación de aneuploidía como los métodos descritos en las Publicaciones de Solicitudes de Patente U.S: Nº US 2007/0202525A1; US2010/0112575A1, US 2009/0087847A1; US2009/0029377A1; US 2008/0220422A1;

US2008/0138809A1, US2008/0153090A1 y la Patente US 7.645.576. El método para determinar la fracción fetal también se puede combinar con ensayos para determinar otras condiciones prenatales asociadas con la madre y/o el feto. Por ejemplo, el método puede usarse en conjunción con análisis prenatales, por ejemplo, como se describe en las Publicaciones de Solicitudes de Patente U.S. Nº US2010/0112590A1, US2009/0162842A1, US2007/0207466A1, y US2001/0051341A1.

La **Figura 6** proporciona un diagrama de flujo de una realización del método de la invención para determinar la fracción de ácidos nucleicos fetales en una muestra biológica materna por secuenciación masivamente paralela de ácidos nucleicos objetivo polimórficos amplificados por PCR determinando independientemente o simultáneamente aneuploidía. El método comprende secuenciar una biblioteca de secuenciación de ácidos nucleicos maternos polimórficos de la manera siguiente. En el paso **610** se obtiene de un sujeto una muestra materna que comprende una mezcla de ácidos nucleicos fetales y maternos. La muestra es una muestra materna que se obtiene de una hembra embarazada, por ejemplo una mujer embarazada. Otras muestras maternas pueden ser de mamíferos, por ejemplo, vaca, caballo, perro o gato. Si el sujeto es un humano, la muestra se puede tomar en el primer o el segundo trimestre de embarazo. Los ejemplos de muestras biológicas maternas se describen anteriormente. En el paso **620**, se procesa adicionalmente la mezcla de ácidos nucleicos fetales y maternos de la fracción de la muestra, por ejemplo plasma, para obtener una muestra que comprende una mezcla purificada de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células, como se describe para la realización **100**. En el paso **630**, se usa una porción de la mezcla purificada de ADN libre de células fetal y materno para amplificar una pluralidad de ácidos nucleicos objetivo polimórficos cada uno comprendiendo un sitio polimórfico. Los sitios polimórficos que están contenidos en los ácidos nucleicos objetivo incluyen sin limitación polimorfismos de nucleótidos simples (SNPs), SNPs en tándem, deleciones o inserciones multi-base a pequeña escala, denominadas IN-DELS (también denominadas polimorfismos de deleción inserción o DIPs), Polimorfismos Multi-Nucleótido (MNPs), Repeticiones Cortas en Tándem (STRs), polimorfismo de longitud de fragmentos de restricción (RFLP), o un polimorfismo que comprenda cualquier otro cambio de secuencia en un cromosoma. Las secuencias polimórficas ejemplares y los métodos para amplificarlas se divulgan para las realizaciones mostradas en las Figuras **2-5**. En algunas realizaciones, los sitios polimórficos que están abarcados por el método de la invención están localizados en cromosomas autosómicos, permitiendo de esta manera la determinación de la fracción fetal independientemente del sexo del feto. Los polimorfismos asociados con los cromosomas distintos al cromosoma 13, 18, 21 e Y también se pueden usar en los métodos descritos en la presente.

En el paso 640, una porción o todas las secuencias polimórficas amplificadas se usan para preparar una biblioteca de secuenciación de una manera paralela como se describe. En una realización, la biblioteca se prepara para la secuenciación por síntesis usando química de secuenciación basada en terminadores reversible de Illumina, como se describe en el Ejemplo 13. En el paso 640, la información de secuencia que se necesita para determinar la fracción fetal se obtiene usando un método NGS. En el paso 650, la fracción fetal se determina en base al número total de etiquetas que mapean para el primer alelo y el número total de etiquetas que mapean para el segundo alelo en un sitio polimórfico informativo, por ejemplo un SNP, contenido en un genoma de referencia artificial, por ejemplo un genoma de referencia SNP. También se describen en la presente genomas objetivo artificiales. Se identifican los sitios polimórficos informativos y la fracción fetal se calcula como se ha descrito.

La determinación de la fracción fetal de acuerdo con la presente puede usarse en conjunción con pruebas de diagnóstico clínicamente relevantes como un control positivo para la presencia de ADN libre de células para resaltar resultados de falso negativo o falso positivo derivados de niveles bajos de ADN libre de células por debajo del límite de identificación. En una realización, la información de la fracción fetal puede usarse para establecer límites y estimar el tamaño mínimo de la muestra en detección de aneuploidías. Dicho uso se describe en el Ejemplo 16 siguiente. La información de la fracción fetal puede usarse en conjunción con información de secuenciación. Por ejemplo, los ácidos nucleicos de una muestra libre de células, por ejemplo una muestra de suero o plasma materno, se puede usar para enumerar secuencias en una muestra. Las secuencias pueden enumerarse usando cualquiera de las técnicas de secuenciación descritas anteriormente. El conocimiento de la fracción fetal se puede usar para establecer límites "de corte" para designar estados de "aneuploidía", "normal" o "marginal/no designado" (incierto). Después, se pueden realizar los cálculos para estimar el número mínimo de secuencias requeridas para alcanzar la sensibilidad adecuada (es decir probabilidad de identificar correctamente un estado de aneuploidía).

Los presentes métodos pueden aplicarse para determinar la fracción de cualquier población de ácidos nucleicos en una mezcla de ácidos nucleicos aportados por genomas diferentes. Además de determinar la fracción aportada a una muestra por dos individuos, por ejemplo los diferentes genomas son aportados por el feto y la madre que lleva el feto, los métodos pueden usarse para determinar la fracción de un genoma en una mezcla derivada de dos células diferentes de un individuo, por ejemplo los genomas son aportados a la muestra por células cancerosas aneuploides y células euploides normales del mismo sujeto.

Composiciones y kits

También se divulgan en la presente composiciones y kits de sistemas reactivos útiles para la práctica de los métodos descritos en la presente.

Las composiciones divulgadas en la presente se pueden incluir en kits para mezclas de secuenciación masivamente paralelas de moléculas de ácidos nucleicos fetales y maternos, por ejemplo ADN libre de células, presentes en una muestra materna, por ejemplo una muestra de plasma. Los kits comprenden una composición que comprende al menos un conjunto de cebadores para amplificar al menos un ácido nucleico objetivo polimórfico en dichas moléculas de ácidos nucleicos fetales y maternos. Los ácidos nucleicos polimórficos pueden comprender sin limitación polimorfismos de nucleótidos simples (SNPs), SNPs en tándem, deleciones o inserciones multi-base a pequeña escala, denominadas IN-DELS (también denominadas polimorfismos de deleción inserción o DIPs), Polimorfismos Multi-Nucleótido (MNPs), Repeticiones Cortas en Tándem (STRs), polimorfismo de longitud de fragmentos de restricción (RFLP), o un polimorfismo que comprenda cualquier otro cambio de secuencia en un cromosoma. Los métodos de secuenciación son métodos NGS de moléculas de ácidos nucleicos individuales o moléculas de ácidos nucleicos amplificadas como se describe en la presente. Los métodos NGS son métodos de secuenciación masivamente paralelos que incluyen pirosecuenciación, secuenciación por síntesis con terminadores de colorante reversibles, secuenciación en tiempo real, secuenciación por ligadura de sonda de oligonucleótidos o/y secuenciación de moléculas individuales.

En una realización, la composición incluye cebadores para amplificar ácidos nucleicos objetivo polimórficos que comprenden cada uno al menos un SNP. El al menos un SNP se selecciona de los SNPs rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005,y rs530022. Los conjuntos correspondientes de cebadores para amplificar los SNPs se proporcionan como las SEQ ID NOs:57-112.

En otra realización, la composición comprende cebadores para amplificar ácidos nucleicos objetivo polimórficos que comprenden cada una al menos un SNP en tándem. Los SNPs en tándem ejemplares incluyen rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980- rs987981; rs4143392- rs4143391; rs1691324- rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 -rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672. En una realización, la composición incluye cebadores para amplificar los SNPs en tándem ejemplares divulgados en la presente, y la composición comprende los cebadores ejemplares correspondientes de las SEQ ID NOS:197-310.

En otra realización, la composición comprende cebadores para amplificar ácidos nucleicos objetivo polimórficos que comprenden cada uno al menos una SRT. Las STRs ejemplares incluyen CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 y D1GATA113. En una realización, la composición incluye cebadores para amplificar las STRs en tándem ejemplares divulgadas en la presente, y la composición comprende los cebadores ejemplares correspondientes de las SEQ ID NOS:113-196.

Los kits pueden contener una combinación de reactivos incluyendo los elementos requeridos para realizar un ensayo de acuerdo con los métodos divulgados en la presente. El sistema de reactivos se presenta de una forma envasada comercialmente, como una composición o mezcla donde la compatibilidad de los reactivos permitirá, en una configuración del dispositivo de ensayo, o más típicamente como un kit de ensayo, es decir combinación envasada de uno o más contenedores, dispositivos, o similares mantener los reactivos necesarios, y preferiblemente incluyendo instrucciones escritas para la realización de los ensayos. El kit divulgado en la presente puede ser adaptado para cualquier configuración de ensayo y puede incluir composiciones para realizar cualquiera de los varios formatos de ensayo descritos en la presente. Los kits para determinar la fracción fetal comprenden composiciones que incluyen conjuntos de cebadores para amplificar ácidos nucleicos polimórficos presentes en una muestra materna como se describe y, donde sea aplicable, los reactivos para purificar ADN libre de células, están dentro del ámbito de la divulgación. En una realización, un kit diseñado para permitir la cuantificación de secuencias polimórficas fetales y materna, por ejemplo STRs y/o SNPs y/o SNPs en tándem en una muestra de plasma de ADN libre de células, incluye al menos un conjunto de oligonucleótidos específicos de alelos específicos para un SNP seleccionado y/o región de repeticiones en tándem. Preferiblemente, el kit incluye una pluralidad de conjuntos de cebadores para amplificar un panel de secuencias polimórficas. Un kit puede comprender otros reactivos y/o información para genotipificar o cuantificar alelos en una muestra (por ejemplo, tampones, nucleótidos,

instrucciones). Los kits también incluyen una pluralidad de contenedores de tampones y reactivos apropiados.

Productos informáticos

La determinación de aneuploidía y/o la determinación de la fracción fetal se deriva computacionalmente de la gran cantidad de información de secuenciación que se obtiene de acuerdo con los métodos descritos en la presente. En una realización, se divulga en la presente un medio legible por ordenador que tiene almacenado en el mismo instrucciones legibles por ordenador para determinar la presencia o ausencia de aneuploidía de información obtenida de la secuenciación dey ácidos nucleicos fetales y maternos en una muestra materna. En una realización, el medio legible por ordenador usa información de secuencia obtenida de una pluralidad de moléculas de ácidos nucleicos fetales y maternos en una muestra de plasma materno para identificar un número de etiquetas de secuencia mapeadas para un cromosoma de interés y para un cromosoma normalizador. Usando el número de etiquetas de secuencia mapeadas para un cromosoma de interés y el número de etiquetas de secuencia mapeadas identificadas para el al menos un cromosoma normalizador, el medio legible por ordenador calcula una dosis de cromosoma para un cromosoma de interés; y compara la dosis de cromosoma con al menos un valor límite, e identifica de esta manera la presencia o ausencia de aneuploidía fetal. Los ejemplos de cromosomas de interés incluyen sin limitación los cromosomas 21, 13, 18 y X.

En otra realización, se divulga en la presente un sistema de procesamiento por ordenador que está adaptado o configurado para determinar la presencia o ausencia de aneuploidía de la información obtenida de la secuenciación de ácidos nucleicos fetales y maternos en una muestra materna. El sistema de procesamiento por ordenador está adaptado o configurado para (a) usar información de secuencia obtenida de una pluralidad de moléculas de ácidos nucleicos fetales y maternos en una muestra de plasma materno para identificar un número de etiquetas de secuencia mapeadas para un cromosoma de interés; (b) usar la información de secuencia obtenida de una pluralidad de moléculas de ácidos nucleicos fetales y maternos en una muestra de plasma materno para identificar un número de etiquetas de secuencia mapeadas para al menos un cromosomas normalizador; (c) usar el número de etiquetas de secuencia mapeadas identificadas para un cromosoma de interés en el paso (a) y el número de etiquetas de secuencia mapeadas identificadas para al menos un cromosoma normalizador en el paso (b) para calcular una dosis de cromosoma para un cromosoma de interés; y (d) comparar la dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal. Los ejemplos de cromosomas de interés incluyen sin limitación los cromosomas 21, 13, 18 y X.

En otra realización, se divulga en la presente un aparato adaptado o configurado para determinar la presencia o ausencia de aneuploidía de la información obtenida de la secuenciación de ácidos nucleicos fetales y maternos en una muestra materna. El aparato está adaptado o configurado para comprender (a) un dispositivo de secuenciación adaptado o configurado para secuenciar al menos una porción de las moléculas de ácidos nucleicos en una muestra de plasma materno que comprende moléculas de ácidos nucleicos fetales y maternos, generando de esta manera información de secuencia; y (b) un sistema de procesamiento por ordenador configurado para realizar los pasos de: (i) usar información de secuencia generada por el dispositivo de secuenciación para identificar un número de etiquetas de secuencia mapeadas para un cromosoma de interés; (ii) usar la información de secuencia generada por el dispositivo de secuenciación para identificar un número de etiquetas de secuencia mapeadas para al menos un cromosoma normalizador; (iii) usar el número de etiquetas de secuencia mapeadas identificadas para un cromosoma de interés en el paso (i) y el número de etiquetas de secuencia mapeadas identificadas para el al menos un cromosoma normalizador en el paso (ii) para calcular una dosis de cromosoma para un cromosoma de interés; y (iv) comparar dicha dosis de cromosoma con al menos un valor límite, e identificar de esta manera la presencia o ausencia de aneuploidía fetal. Los ejemplos de cromosomas de interés incluyen sin limitación los cromosomas 21, 13, 18 y X.

La presente invención se describe con más detalle en los siguientes Ejemplos que no se pretende que limiten de ninguna manera el ámbito de la invención como se reivindica. Se pretende que las Figuras adjuntas se consideren como parte integral de la especificación y descripción de la invención. Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar la invención reivindicada.

Ejemplo 1

Procesamiento de Muestras y Extracción de ADN libre de células

Las muestras de sangre periférica se recogieron de mujeres embarazadas en su primer o segundo trimestre de embarazo y que se consideraron con riesgo de aneuploidía fetal. Se obtuvo el consentimiento informado de cada participante antes de la extracción de sangre. La sangre se recogió antes de la amniocentesis o la muestra de vellosidades coriónicas. El análisis de cariotipos se realizó usando las muestras de vellosidades coriónicas o amniocentesis para confirmar el cariotipo fetal.

La sangre periférica extraída de cada sujeto se recogió en tubos ACD. Se transmitió un tubo de muestra de sangre (aproximadamente 6-9 ml/tubo en un tubo de centrífuga de baja velocidad de 15 ml. La sangre se centrifugó

a 2640 rpm, 4ºC durante 10 minutos usando una centrífuga Beckman Allegra 6 R y un rotor modelo GA 3.8.

Para la extracción de plasma libre de células, la capa de plasma superior se transfirió a un tubo de centrífuga de alta velocidad de 15 ml y se centrifugó a 16000 x g, 4º C durante 10 min. usando una centrífuga Beckman Coulter Avanti J-E y un rotor JA-14. Los dos pasos de centrifugación se realizaron en el plazo de 72 horas después de la recogida de la sangre. El plasma libre de células que comprende ADN libre de células se almacenó a -80º C y se descongeló sólo una vez antes de la amplificación del ADN libre de células del plasma o para la purificación del ADN libre de células.

Se extrajo ADN libre de células (ADNcf) purificado de plasma libre de células usando el QIAamp Blood DNA Mini kit (Qiagen) de acuerdo esencialmente con las instrucciones del fabricante. Se añadieron un mililitro de tampón AL y 100 µl de solución de proteasa a 1ml de plasma. La mezcla se incubó durante 15 minutos a 56º C. Se añadió un mililitro de 100% de etanol a la digestión de plasma. La mezcla resultante se transfirió a minicolumnas QIAamp que se montaron con los VacValves y VacConnectors proporcionado en el montaje de columnas QIAvac 24 Plus (Qiagen). Se aplicó vacio a las muestras, y el ADN libre de células retenido en los filtros de la columna se lavó al vacio con 750 µl de tampón AW1, seguido por un segundo lavado con 750 µl de tampón AW24. La columna se centrifugó a 14.000 RPM durante 5 minutos para eliminar cualquier tampón residual del filtro. El ADN libre de células se eluyó con tampón AE por centrifugación a 14.000 RPM, y se determinó la concentración usando la Qubit™ Quantitation Platform (Invitrogen).

Ejemplo 2

Preparación y secuenciación de bibliotecas de secuenciación primarias y enriquecidas

a. Preparación de bibliotecas de secuenciación - protocolo abreviado

Todas las bibliotecas de secuenciación, es decir bibliotecas primarias y enriquecidas, se prepararon de aproximadamente 2 ng de ADN libre de células purificado que se extrajo de plasma materno. La preparación de la biblioteca se realizó usando reactivos del Conjunto 1 de Reactivos de ADN de Preparación de Muestras NEBNext™ (Parte No. E6000L; New England Biolabs, Ipswich, MA), para Illumina® como sigue. Debido a que el ADN de plasma libre de células está fragmentado en la naturaleza, no se hizo fragmentación adicional por nebulización o sonicación en las muestras de ADN de plasma. Los excesos de aproximadamente 2 ng de fragmentos de ADN libre de células purificados contenidos en 40 µl se convirtieron en extremos romos fosforilados de acuerdo con el NEBNext® End Repair Module incubando en un tubo de microcentrífuga de 1,5 ml el ADN libre de células con 5 µl de tampón de fosoforilación10X, 2 µl de mezcla de solución de desoxinucleótidos (10 mM cada dNTP), 1 µl de una dilución 1:5 de ADN polimerasa, 1 µl de ADN Polimerasa T4 y 1µl de Polinucleótido Quinasa T4 proporcionado en el Conjunto 1 de Reactivos de ADN de Preparación de Muestras de ADN NEBNext™ durante 15 minutos a 20º C. Los enzimas fueron entonces inactivados por calor incubando la mezcla de la reacción a 75º C durante 5 minutos. La mezcla se enfrió a 4º C, y se llevó a cabo la adición de colas de dA del ADN de extremo romo usando 10 µl de la mezcla maestra de adición de colas de dA que contenía el fragmento de Klenow (3' a 5' menos exo) (Conjunto 1 de Reactivos de ADN de Preparación de Muestras NEBNext™), e incubando durante 15 minutos a 37º C. Posteriormente, el fragmento de Klenow se inactivó por calor incubando la mezcla de la reacción a 75º C durante 5 minutos. Después de la inactivación del fragmento de Klenow, se usó 1 µl de una dilución 1:5 de Illumina Genomic Adaptor Oligo Mix (Parte No. 1000521; Illumina Inc., Hayward, CA) para ligar los adaptadores de Illumina (sin Índice Adaptadores Y) al ADN de la adición de colas de dA usando 4 µl de la ADN ligasa T4 proporcionada en el Conjunto 1 de Reactivos de ADN de Preparación de Muestras de ADN NEBNext™, incubando la mezcla de la reacción durante 15 minutos a 25º C. La mezcla se enfrió a 4º C, y el ADN libre de células ligado por adaptadores se purificó a partir de adaptadores no ligados, dímeros de adaptadores y otros reactivos usando microesferas magnéticas proporcionadas en el sistema de purificación Agencourt AMPure XP PCR (Parte No. A63881; Beckman Coulter Genomics, Danvers, MA). Se realizaron dieciocho ciclos de PCR para enriquecer selectivamente el ADN libre de células ligado por adaptadores (25 µl) usando Phusion ® High-Fidelity Master Mix (25ml; Finnzymes, Woburn, MA) y cebadores de PCR de Illumina 0,5 µM cada uno) complementarios a los adaptadores (Partes No. 1000537 y 1000537). El ADN ligado por adaptadores se sometió a PCR (98º C durante 30 segundos; 18 ciclos de 98º C durante 10 segundos, 65º C durante 30 segundos y 72º C durante 30; extensión final a 72º C durante 5 minutos, y mantener a 4º C) usando Cebadores de PCR Genómicos Illumina (Partes Nos. 100537 y 1000538) y la Mezcla Maestra de PCR Phusion HF proporcionado en el Conjunto 1 de Reactivos de ADN de Preparación de Muestras de ADN NEBNext™, de acuerdo con las instrucciones del fabricante. El producto amplificado se purificó usando el sistema de purificación Agencourt AMPure XP PCR (Agencourt Bioscience Corporation, Beverly, MA), de acuerdo con las instrucciones del fabricante disponibles en www.beckmangenomics.com/products/AMPmeXPProtocol_000387v001.pdf. El producto amplificado purificado se eluyó en 40 µl de tampón Qiagen EB, y la concentración y la distribución del tamaño de las bibliotecas amplificadas se analizó usando el Kit Agilent DNA 1000 para el 2100 Bioanalizador (Agilent technologies Inc., Santa Clara, CA).

b. Preparación de bibliotecas de secuenciación - protocolo de larga duración

El protocolo de larga duración descrito es esencialmente el protocolo estándar proporcionado por Illumina, y sólo difiere del protocolo de Illumina en la purificación de la biblioteca amplificada: El protocolo Illumina instruye que la biblioteca amplificada sea purificada usando electroforesis en gel, mientras que el protocolo descrito en la presente usa microesferas magnéticas para el mismo paso de purificación. Aproximadamente se usaron 2 ng de ADN libre de células purificado que se había extraído de plasma materno para preparar un biblioteca de secuenciación primaria usando el Conjunto 1 de Reactivos de ADN de Preparación de Muestras NEBNext™ (Parte No. E6000L; New England Biolabs, Ipswich, MA) para Illummina® de acuerdo esencialmente con las instrucciones del fabricante. Todos los pasos excepto para la purificación final de los productos ligados por adaptadores, que se realizó usando microesferas magnéticas de Agencourt y reactivos en lugar de columna de purificación, se realizaron de acuerdo con el protocolo que acompaña al NEBNext™ Reactivos para Preparación de Muestra para una biblioteca de ADN genómico que se secuencia usando Illumina® GAII. El protocolo NEBNext™ sigue esencialmente el proporcionado por Illumina, que está disponible en grcf.jhml.edu/hts/protocols/11257047_ChIP_Sample_Prep.pdf.

Los excesos de aproximadamente 2 ng de fragmentos de ADN libre de células purificados contenidos en 40µl se convirtieron en extremo romos fosforilados de acuerdo con el NEBNext® End Repair Module incubando los 40µl de ADN libre de células con 5 µl de tampón de fosforilación 10X, 2 µl de mezcla de solución de desoxinucleótidos (10 mM cada uno dNTP), 1 µl de una dilución 1:5 de ADN Polimerasa I, 1µl de ADN Polimerasa T4 y 1µl de Polinucleótido Quinasa proporcionada en el NEBNext™ DNA Sample Prep DNA Reagent Set 1 en un tubo de microcentrífuga de 200 µl en un ciclador térmico durante 30 minutos a 20º C. La muestra se enfrió a 4º C, y se purificó usando una columna QIAQuick proporcionada en el Kit de Purificación por PCR QIAQuick (QIAGEN Inc., Valencia, CA) de la manera siguiente. Los 50 µµl de reacción se transfirieron a un tubo de microcentrífuga, y se añadieron 250 µl de Tampón PB Qiagen. Los 300µl resultantes se transfirieron a una columna QIAquick, que se centrifugó a 13.000 RPM durante 1 minuto en una microcentrífuga. La columna se lavó con 750µl de Tampón PE Qiagen, y se volvió a centrifugar. El etanol residual se eliminó por una centrifugación adicional durante 5 minutos a 13.000 RPM. El ADN se eluyó en 39µl de Tampón EB Qiagen por centrifugación. La adición de colas de dA de 34µl del ADN de extremos romos se consiguió usando 16µl de la mezcla maestra de adición de colas de dA que contenía el fragmento de Klenow (3' a 5' menos exo) (NEBNext™ DNA Sample Prep DNA Reagent Set 1), e incubando durante 30 minutos a 37º C de acuerdo con el Módulo de Adición de colas de dA NEBNext® del fabricante. La muestra se enfrió a 4º C, y se purificó usando una columna proporcionada en el Kit de Purificación por PCR MinElute (QIAGEN Inc., Valencia, CA) como sigue. Los 50 µl de la reacción se transfirieron a un tubo de microcentrífuga de 1,5 ml, y se añadieron 250µl de tampón PB Qiagen. Los 300 µl se transfirieron a la columna MinElute, que se centrifugó a 13.000 RPM durante 1 minuto en una microcentrífuga. La columna se lavó con 750 µl de Tampón PE Qiagen, y se volvió a centrifugar. El etanol residual se eliminó por una centrifugación adicional durante 5 minutos a 13.000 RPM. El ADN se eluyó en 15 µl de Tampón EB Qiagen por centrifugación. Se incubaron diez microlitros del eluido de ADN con 1 µl de una dilución 1:5 del Illumina Genomic Adapter Oligo Mix (Parte No. 1000521), 15µl del Tampón de Reacción de Ligación Rápida 2x, y 4 µl de ADN Ligasa T4 Rápido, durante 15 minutos a 25º C de acuerdo con el Módulo de Ligación Rápida NEBNext®. La muestra se enfrió a 4º C, y se purificó usando una columna MinElute como sigue. Se añadieron ciento cincuenta microlitros de Tampón PE Qiagen a laos 30 µl de reacción, y el volumen completo se transfirió a una columna MinElute , que se centrifugó a 13.000 RPM durante un minuto en una microcentrífuga. La columna se lavó con 750 µl de Tampón PE Qiagen, y se volvió a centrifugar. El etanol residual se eliminó por una centrifugación adicional durante 5 minutos a 13.000 RPM. El ADN se eluyó en 28 µl de Tampón EB Qiagen por centrifugación. Se sometieron veintitrés microlitros del eluido de ADN ligado por adaptadores a 18 ciclos de PCR (98º C durante 30 segundos; 18 ciclos de 98º C durante 10 segundos, 65º C durante 30 segundos y 72º C durante 30; extensión final a 72º C durante 5 minutos, y mantener a 4º C) usando Cebadores de PCR Genómicos Illumina (Partes Nos. 100537 y 1000538) y la Mezcla Maestra de PCR Phusion HF proporcionada en el Conjunto 1 de Reactivos de ADN de Preparación de Muestras de ADN NEBNext™, de acuerdo con las instrucciones del fabricante. El producto amplificado se purificó usando el sistema de purificación por PCR Agencourt AMPure XP (Agencourt Bioscience Corporation, Beverly, MA) de acuerdo con las instrucciones del fabricante disponibles en www.beckmangenomics.com/products/AMPureXPProtocol_000387v001.pdf. El sistema de purificación por PCR Agencourt AMPure XP elimina dNTPs, cebadores, dímeros de cebadores sales y otros contaminantes no incorporados, y recupera amplicones mayores de 100bp. El producto amplificado purificado se eluyó de las microesferas Agencourt en 40 µl de Tampón EB Qiagen y la distribución del tamaño de las bibliotecas se analizó usando el Kit Agilent DNA 1000 para el 2100 Bioanalizador (Agilent technologies Inc., Santa Clara, CA).

c. Análisis de las bibliotecas de secuenciación preparadas de acuerdo con los protocolos abreviado (a) y de larga duración (b)

Los electroferogramas generados por el Bioanalizador se muestran en la **Figura 7**. La **Figura 7(A)** muestra el electroferograma del ADN de la biblioteca preparado de ADN libre de células purificado de la muestra de plasma M24228 usando el protocolo de larga duración descrito en (a), y la **Figura 7(B)** muestra el electroferograma del AND de la biblioteca preparado de ADN libre de células purificado de la muestra de plasma M24228 usando el protocolo de larga duración descrito en (b). En ambas figuras, los picos 1 y 4 representan el Marcador Inferior de 15 bp y el Marcador Superior de 1500, respectivamente; los números encima de los picos indican los tiempos de migración para los fragmentos de la biblioteca; y las líneas horizontales indican el límite establecido para la integración. El electroferograma en la **Figura 7(A)** muestra un pico menor de los fragmentos de 187bp y un pico principal de los

fragmentos de 263bo, mientras que el electroferograma en la **Figura 7(B)** muestra sólo un pico en 265 bp. La integración de las aéreas de los picos resultó en una concentración calculada de 0,40 ng/µl para el ADN del pico de 187bp en la **Figura 7(A)**, una concentración de 7,34 ng/µl para el ADN del pico de 263bp en la **Figura 7(A)**, y una concentración de 14,72 ng/µl para el ADN del pico de 265bp en la **Figura 7(B)**. Se sabe que los adaptadores Illumina que se ligaron al ADN libre de células eran de 92bp, que cuando se restan de 265bp, indican que el tamaño del pico para el ADN libre de células es de 173bp. Es posible que el pico menor en 187bp represente fragmentos de dos cebadores que se ligaron extremo a extremo. Los fragmentos de dos cebadores lineales se eliminan del producto de la biblioteca final cuando se usa el protocolo abreviado. El protocolo abreviado también elimina otros fragmentos más pequeños de menos de 187bp. En este ejemplo, la concentración de ADN libre de células ligado por adaptadores purificado es el doble que la del ADN libre de células ligado por adaptadores producido usando el protocolo de larga duración. Se ha observado que la concentración de los fragmentos de ADN libre de células ligado por adaptadores es siempre mayor que la obtenida usando el protocolo de larga duración (datos no mostrados).

Por lo tanto, una ventaja de preparar la biblioteca de secuenciación usando el protocolo abreviado es que la biblioteca obtenida comprende consistentemente sólo un pico principal en el intervalo 262-267bp mientras que la calidad de la biblioteca preparada usando el protocolo de larga duración varía como se refleja por el número y movilidad de los picos distintos que representan el ADN libre de células. Los productos no de ADN libre de células ocuparían espacio en la célula de flujo y disminuirían la calidad de la amplificación de clústeres y la posterior formación de imágenes de las reacciones de secuenciación, lo que es la base de la asignación general de estatus de aneuploidía. El protocolo abreviado mostró que no afecta a la secuenciación de la biblioteca (ver **Figura 8**).

Otra ventaja de preparar la biblioteca de secuenciación usando el protocolo abreviado es que los tres pasos enzimáticos de reparación de extremos, adición de colas de dA y ligado por adaptadores, tardan menos de una hora en completarse para apoyar la validación e implementación de un servicio de diagnóstico de aneuploidía rápido.

Otra ventaja es que los tres pasos enzimáticos de reparación de extremos, adición de colas de dA y ligado por adaptadores, se realizan en el mismo tubo de reacción, evitando así múltiples transferencias de muestras que podrían llevar potencialmente a pérdida de material y más importantemente a posibles confusiones de muestras y contaminación de la muestra.

Ejemplo 3

Secuenciación Masivamente Paralelo y Determinación de Aneuploidía

Se obtuvieron muestras de sangre periférica de sujetos embarazados y el ADN libre de células se purificó de la fracción de plasma como se describe en el ejemplo 1. Todas las bibliotecas de secuenciación se prepararon usando el protocolo de preparación de bibliotecas abreviado descrito en el Ejemplo 2. El ADN amplificado se secuenció usando El Analizador de Genomas II de Illumina para obtener lecturas de extremos individuales de 36 bp. Sólo se necesitaron alrededor de 30 bp de la información de la secuencia aleatoria para identificar una secuencia como perteneciente a un cromosoma humano específico. Las secuencias más largas pueden identificar únicamente objetivos más particulares. En el presente caso, se obtuvo un número mayor de lecturas de 36 bp, cubriendo aproximadamente un 10% del genoma. La secuenciación del ADN de la biblioteca se realizó usando el Analizador de Genomas II (Illumina Inc., San Diego, CA, USA) de acuerdo con los protocolos del fabricante. Se pueden encontrar copias del protocolo para la secuenciación del genoma completo usando tecnología Illumina/Solexa en BioTechniques.RTM. Protocol Guide 2007 Publicado en Diciembre del 2006: p 29, y en la red mundial en biotechniques.com/default.asp? page=protocol&subsection=article_display&id=112378. La biblioteca de ADN se diluyó a 1 nM y se desnaturalizó. La biblioteca de ADN (5pM) se sometió a amplificación de clústeres de acuerdo con el procedimiento descrito en la Cluster Station User Guide and Cluster Station Operations Guide de Illumina, disponible en la red mundial en illumina.com/systems/genome_analyzer/cluster_station.ilmn. Al completarse la secuenciación de la muestra, el "Software de Control del Secuenciador" transfirió los archivos de imagen y designación de bases a un servidor Unix que estaba ejecutando el software de Illumina "Genome Analyzer Pipeline" versión 1.51. Se ejecutó el programa "Gerald" de Illumina para alinear las secuencias con el genoma humano de referencia que se deriva del genoma hg18 proporcionado por el National Center for Biotechnology Information (NCBI36/hg18, disponible en la red mundial en http://genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgsid=166260105). Los datos de secuencia generados del procedimiento anterior que alinearon únicamente con el genoma se leyeron de las salidas de Gerald (archivos export.txt) por un programa (c2c.pl) ejecutado en un ordenador con el programa operativo Linux. Se permitieron los alineamientos de secuencia con discrepancias de bases y se incluyeron en los recuentos de alineamientos sólo si se alineaban únicamente con el genoma. Se excluyeron los alineamientos de secuencia con coordenadas de partida y finales idénticas (duplicados).

Se mapearon entre alrededor de 5 y 15 millones de etiquetas de 36 bp con 2 o menos discrepancias únicamente con el genoma humano. Todas las etiquetas mapeadas se contaron y se incluyeron en el cálculo de dosis de cromosomas en ambas muestras de ensayo y calificación. Las regiones que se extienden desde la base 0 a la base 2 x 106, base 10 x 106 a base 13 x 106 y base 23 x 106 al final del cromosoma Y, fueron excluidas

específicamente del análisis debido a que las etiquetas derivas de fetos masculinos o femeninos mapean para estas regiones del cromosoma Y.

Se observó alguna variación en el número total de etiquetas de secuencia mapeadas para cromosomas individuales a través de las muestras secuenciadas en la misma ejecución (variación inter-cromosómica), pero se observó que hubo sustancialmente mayor variación entre ejecuciones de secuenciación diferentes (variación de ejecución inter-secuenciación).

Ejemplo 4

Dosis y Varianza para los cromosomas 13, 18, 21, X e Y

Para examinar la extensión de la variación inter-cromosómica e inter-secuenciación en el número de etiquetas de secuencia mapeadas para todos los cromosomas, se extrajo y secuenció el ADN libre de células de plasma obtenido de sangre periférica de 48 sujetos embarazados voluntarios como se describe en el Ejemplo 1, y se analizó de la manera siguiente.

Se determinó el número total de etiquetas de secuencia que se mapearon para cada cromosoma (densidad de etiqueta de secuencia). Alternativamente, el número de etiquetas de secuencia mapeadas se puede normalizar a la longitud del cromosoma para generar una proporción de densidad de etiqueta de secuencia. La normalización de la longitud del cromosoma no es un paso requerido, y puede ser realizada únicamente para reducir el número de dígitos en e un número para simplificarlo para la interpretación humana. Las longitudes de cromosomas que se pueden usar para normalizar los recuentos de etiquetas de secuencia pueden ser las longitudes proporcionadas en la red mundial en genome.ucsc.edu/goldenPath/stats.html#hg18.

La densidad de etiqueta de secuencia resultante para cada cromosoma se relacionó con la densidad de etiqueta de secuencia de cada uno de los cromosomas restantes para derivar una dosis de cromosoma calificada, que se calculó como la proporción de la densidad de etiqueta de secuencia para el cromosoma de interés, por ejemplo cromosoma 21, y la densidad de etiqueta de secuencia para cada uno de los cromosomas restantes, es decir cromosomas 1-20, 22 y X. La Tabla 1 proporciona un ejemplo de la dosis de cromosoma calificada calculada para los cromosomas de interés 13, 18, 21, X e Y, determinada en una de las muestras calificadas. Las dosis de cromosomas se determinaron para todos los cromosomas en todas las muestras, y las dosis medias para los cromosomas de interés 13, 18, 21, X e Y en las muestras calificadas se proporcionan en las Tablas 2 y 3, y se representan en las Figuras 9-13. Las Figuras 9-13 también representan las dosis de cromosomas para las muestras de ensayo. Las dosis de cromosomas para cada uno de los cromosomas de interés en las muestras calificadas proporcionan una medición de la variación en el número total de etiquetas de secuencia mapeadas para cada cromosoma de interés en relación a la de cada uno de los cromosomas restantes. Por lo tanto, las dosis de cromosomas calificadas pueden identificar el cromosoma o un grupo de cromosomas, es decir cromosoma normalizador, que tiene una variación entre muestras que está más cercana a la variación del cromosoma de interés, y que serviría como secuencias ideales para valores normalizadores para evaluación estadística adicional. Las Figuras 14 y 15 representan las dosis de cromosomas medias calculadas determinadas en una población de muestras calificadas para los cromosomas 13, 18 y 21, y los cromosomas X e Y.

En algunas situaciones, el mejor cromosoma normalizador puede no tener la menor variación, pero puede tener una distribución de dosis calificadas que distingue mejor una muestra o muestras de ensayo de las muestras calificadas, es decir el mejor cromosoma normalizador puede no tener la variación más baja, pero puede tener la diferenciabilidad más alta. Por lo tanto, la diferenciabilidad cuenta para la variación en la dosis de cromosoma y la distribución de las dosis en las muestras calificadas.

Las Tablas 2 y 3 proporcionan el coeficiente de variación como la medición de la variabilidad, y los valores de la prueba t de student como una medición de diferenciabilidad para los cromosomas 18, 21, X e Y, en donde cuanto más pequeño el valor de la prueba T, mayor es la diferenciabilidad. La diferenciabilidad para el cromosoma 13 se determinó como la proporción de diferencia entre la dosis de cromosoma media en las muestras calificadas y la dosis para el cromosoma 13 en la única muestra de prueba de T13, y la desviación estándar de la media de la dosis calificada.

Las dosis de cromosomas calificadas también sirven como la base para determinar los valores límite cuando se identifican aneuploidías en muestras de ensayo como se describe a continuación.

TABLA 1

Dosis de Cromosoma Calificada para los Cromosomas 13, 18, 21, X e Y (n=1; muestra #11342, 46 XY)					
Cromosoma	**chr 21**	**chr 18**	**chr 13**	**chr X**	**chrY**
chr1	0.149901	0.306798	0.341832	0.490969	0.003958
chr2	0.15413	0.315452	0.351475	0.504819	0.004069
chr3	0.193331	0.395685	0.44087	0.633214	0.005104
chr4	0.233056	0.476988	0.531457	0.763324	0.006153
chr5	0.219209	0.448649	0.499882	0.717973	0.005787
chr6	0.228548	0.467763	0.521179	0.748561	0.006034
chr7	0.245124	0.501688	0.558978	0.802851	0.006472
chr8	0.256279	0.524519	0.584416	0.839388	0.006766
chr-9	0.309871	0.634203	0.706625	1.014915	0.008181
chr10	0.25122	0.514164	0.572879	0.822817	0.006633
chr11	0.257168	0.526338	0.586443	0.8423	0.00679
chr12	0.275192	0.563227	0.627544	0.901332	0.007265
chr13	0.438522	0.897509	1	1.436285	0.011578
chr14	0.405957	0.830858	0.925738	1.329624	0.010718
chr15	0.406855	0.832697	0.927786	1.332566	0.010742
chr16	0.376148	0.769849	0.857762	1.231991	0.009931
chr17	0.383027	0.783928	0.873448	1.254521	0.010112
chr18	0.488599	1	1.114194	1.600301	0.0129
chr19	0.535867	1.096742	1.221984	1.755118	0.014148
chr20	0.467308	0.956424	1.065642	1.530566	0.012338
chr21	1	2.046668	2.280386	3.275285	0.026401
chr22	0.756263	1.547819	1.724572	2.476977	0.019966
chrX	0.305317	0.624882	0.696241	1	0.008061
chrY	37.87675	77.52114	86.37362	124.0572	1

TABLA 2

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 21,18 y 13								
	21 (n=35)				**18 (n=40)**			
	Media	Desviación estándar	CV	Prueba T	Media	Desviación estándar	CV	Prueba T
chr1	0.15335	0.001997	1.30	3.18E-10	0.31941	0.008384	2.62	0.001675
chr2	0.15267	0.001966	1.29	9.87E-07	0.31807	0.001756	0.55	4.39E-05
chr3	0.18936	0.004233	2.24	1.04E-05	0.39475	0.002406	0.61	3.39E-05
chr4	0.21998	0.010668	4.85	0.000501	0.45873	0.014292	3.12	0.001349
chr5	0.21383	0.005058	2.37	1.43E-05	0.44582	0.003288	0.74	3.09E-05
chr6	0.22435	0.005258	2.34	1.48E-05	0.46761	0.003481	0.74	2.32E-05

(continuada)

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 21, 18 y 13								
	21 (n=35)				18 (n=40)			
	Media	Desviación estándar	CV	Prueba T	Media	Desviación estándar	CV	Prueba T
chr7	0.24348	0.002298	0.94	2.05E-07	0.50765	0.004669	0.92	9.07E-05
chr8	0.25269	0.003497	1.38	1.52E-06	0.52677	0.002046	0.39	4.89E-05
chr9	0.31276	0.003095	0.99	3.83E-09	0.65165	0.013851	2.13	0.000559
chr10	0.25618	0.003112	1.21	2.28E-10	0.53354	0.013431	2.52	0.002137
chr-11	0.26075	0.00247	0.95	1.08E-09	0.54324	0.012859	2.37	0.000998
chr12	0.27563	0.002316	0.84	2.04E-07	0.57445	0.006495	1.13	0.000125
chr13	0.41828	0.016782	4.01	0.000123	0.87245	0.020942	2.40	0.000164
chr14	0.40671	0.002994	0.74	7.33E-08	0.84731	0.010864	1.28	0.000149
chr15	0.41861	0.007686	1.84	1.85E-10	0.87164	0.027373	3.14	0.003862
chr16	0.39977	0.018882	4.72	7.33E-06	0.83313	0.050781	6.10	0.075458
chr17	0.41394	0.02313	5.59	0.000248	0.86165	0.060048	6.97	0.088579
chr18	0.47236	0.016627	3.52	1.3E-07				
chr19	0.59435	0.05064	8.52	0.01494	1.23932	0.12315	9.94	0.231139
chr20	0.49464	0.021839	4.42	2.16E-06	1.03023	0.058995	5.73	0.061101
chr21					2.03419	0.08841	4.35	2.81E-05
chr22	0.84824	0.070613	8.32	0.02209	1.76258	0.169864	9.64	0.181808
chrX	0.27846	0.015546	5.58	0.000213	0.58691	0.026637	4.54	0.064883

TABLA 3

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 13, X, e Y								
	13 (n=47)				X (n=19)			
	Media	**Desviación estándar**	**CV**	**Diff**	**Media**	**Desviación estándar**	**CV**	**Prueba T**
chr1	0.36536	0.01775	4.86	1.904	0.56717	0.025988	4.58	0.001013
chr2	0.36400	0.009817	2.70	2.704	0.56753	0.014871	2.62	9.6E-08
chr3	0.45168	0.007809	1.73	3.592	0.70524	0.011932	1.69	6.13E-11
chr4	0.52541	0.005264	1.00	3.083	0.82491	0.010537	1.28	1.75E-15
chr5	0.51010	0.007922	1.55	3.944	0.79690	0.012227	1.53	1.29E-11
chr6	0.53516	0.008575	1.60	3.758	0.83594	0.013719	1.64	2.79E-11
chr7	0.58081	0.017692	3.05	2.445	0.90507	0.026437	2.92	7.41E-07
chr8	0.60261	0.015434	2.56	2.917	0.93990	0.022506	2.39	2.11E-08
chr9	0.74559	0.032065	4.30	2.102	1.15822	0.047092	4.07	0.000228
chr10	0.61018	0.029139	4.78	2.060	0.94713	0.042866	4.53	0.000964
chr11	0.62133	0.028323	4.56	2.081	0.96544	0.041782	4.33	0.000419
chr12	0.65712	0.021853	3.33	2.380	1.02296	0.032276	3.16	3.95E-06
chr13					1.56771	0.014258	0.91	2.47E-15
chr14	0.96966	0.034017	3.51	2.233	1.50951	0.05009	3.32	8.24E-06
chr15	0.99673	0.053512	5.37	1.888	1.54618	0.077547	5.02	0.002925
chr16	0.95169	0.080007	8.41	1.613	1.46673	0.117073	7.98	0.114232
chr17	0.98547	0.091918	9.33	1.484	1.51571	0.132775	8.76	0.188271
chr18	1.13124	0.040032	3.54	2.312	1.74146	0.072447	4.16	0.001674
chr19	1.41624	0.174476	12.32	1.306	2.16586	0.252888	11.68	0.460752
chr20	1.17705	0.094807	8.05	1.695	1.81576	0.137494	7.57	0.08801
chr21	2.33660	0.131317	5.62	1.927	3.63243	0.235392	6.48	0.00675
chr22	2.01678	0.243883	12.09	1.364	3.08943	0.34981	11.32	0.409449
chrX	0.66679	0.028788	4.32	1.114				
chr2-6	0.46751	0.006762	1.45	4.066				
chr3-6	0.50332	0.005161	1.03	5.260				
chr_tot					1.13209	0.038485	3.40	2.7E-05
	Y (n=26)							
	Media	**Desviación estándar**	**CV**	**Prueba T**				
Chr 1-22, X	0.00734	0.002611	30.81	1.8E-12				

Los ejemplos de diagnósticos de T21, T13, T18 y un caso del síndrome de Turner obtenidos usando los cromosomas normalizadores, dosis de cromosomas y diferenciabilidad para cada uno de los cromosomas de interés se describen en el Ejemplo 3.

Ejemplo 5

Diagnóstico de Aneuploidía Fetal Usando Cromosomas Normalizadores

Para aplicar el uso de dosis de cromosomas para evaluar aneuploidía en una muestra de ensayo biológica, las muestras de ensayo de sangre materna se obtuvieron de voluntarios embarazados y se preparó ADN libre de células, y se secuenció y analizó una biblioteca de secuenciación de acuerdo con el protocolo abreviado descrito en el Ejemplo 2.

Trisomía 21

La Tabla 4 proporciona la dosis calculada para el cromosoma 21 en una muestra de ensayo ejemplar (#11403). El límite calculado para el diagnóstico positivo de la aneuploidía T21 se estableció en >2 desviaciones estándar de la medio de las muestras calificadas (normales). Un diagnóstico para T21 se dio en base a que la dosis de cromosoma en la muestra de ensayo es mayor que el límite establecido. Los cromosomas 14 y 15 se usaron como cromosomas normalizadores en cálculos separados para mostrar que o bien un cromosoma que tenga la variabilidad más baja, por ejemplo el cromosoma 14, o un cromosoma que tenga la diferenciabilidad más alta, por ejemplo el cromosoma 15, se puede usar para identificar la aneuploidía. Se identificaron trece muestras de T21 usando las dosis de cromosoma calculadas, y las muestras de aneuploidía se confirmaron que eran T21 por cariotipo.

TABLA 4

Dosis de Cromosoma para una Aneuploidía T21 (muestra #11403, 47 XY +21)			
Cromosoma	Densidad de Etiqueta de Secuencia	Dosis de Cromosoma para Chr 21	Límite
Chr21	333,660	0.419672	0.412696
Chr14	795,050		
Chr21	333,660	0.441038	0.433978
Chr15	756,533		

Trisomía 18

La Tabla 5 proporciona la dosis calculada para el cromosoma 18 en una muestra de ensayo (#11390). El límite calculado para el diagnóstico positivo de la aneuploidía T18 se estableció en 2 desviaciones estándar de la media de las muestras calificadas (normales). Se dio un diagnóstico para T18 en base a que la dosis de cromosoma en la muestra de ensayo era mayor que el límite establecido. El cromosoma 8 se usó como el cromosoma normalizador. En esta situación el cromosoma 8 tenía la variabilidad más baja y la diferenciabilidad más alta. Se identificaron ocho muestras de T18 usando dosis de cromosomas, y se confirmaron que eran T18 por cariotipo.

Estos datos muestran que un cromosoma normalizador puede tener tanto la variabilidad más baja como la diferenciabilidad más alta.

TABLA 5

Dosis de Cromosoma para una Aneuploidía T18 (muestra #11390, 47 XY +18)			
Cromosoma	Densidad de Etiqueta de Secuencia	Dosis de Cromosoma para Chr 18	Límite
Chr18	602,506	0.585069	0.530867
Chr8	1,029,803		

Trisomía 13

La Tabla 6 proporciona la dosis calculada para el cromosoma 13 en una muestra de ensayo (#51236). El límite calculado para el diagnóstico positivo de la aneuploidía T13 se estableció en 2 desviaciones estándar de la media de las muestras calificadas. Se dio un diagnóstico para T13 en base a que la dosis de cromosoma en la muestra de ensayo era mayor que el límite establecido. La dosis de cromosoma para el cromosoma 13 se calculó usando o el cromosoma 5 o el grupo de cromosomas 3, 4, 5 y 6 como el cromosoma normalizador. Se identificó una muestra de T13.

TABLA 6

Dosis de Cromosoma para una Aneuploidía T13 (muestra #51236, 47 XY +13)			
Cromosoma	Densidad de Etiqueta de Secuencia	Dosis de Cromosoma para Chr 13	Límite
Chr13	692,242	0.541343	0.52594
Chr5	1,278,749		
Chr13	692,242	0.530472	0.513647
Chr3-6 [media]	1,304,954		

La densidad de etiqueta de secuencia para los cromosomas 3-6 es la media de los recuentos de etiquetas para los cromosomas 3-6.

Los datos muestran que la combinación de cromosomas 3, 4, 5y 6 proporcionan una variabilidad que es más baja que la del cromosoma 5, y la diferenciabilidad más alta que cualquiera de la de los otros cromosomas.

Por lo tanto, se puede usar un grupo de cromosomas como el cromosoma normalizador para determinar las dosis de cromosomas e identificar aneuploidías.

Síndrome de Turner (monosomía X)

La Tabla 7 proporciona la dosis calculada para los cromosomas X e Y en una muestra de ensayo (#51238). El límite calculado para el diagnóstico positivo del Síndrome de Turner (monosomía X) se estableció para el cromosoma X en <-2 desviaciones estándar de la media, y para la ausencia del cromosoma Y en <-2 desviaciones estándar de la media para las muestras calificadas (normales).

TABLA 7

Dosis de Cromosoma para una aneuploidía de Turner (XO) (muestra #51238, 45 X)			
Cromosoma	Densidad de Etiqueta de Secuencia	Dosis de Cromosoma para Chr X y Chr Y	Límite
ChrX	873,631	0.786642	0.803832
Chr4	1,110,582		
ChrY	1,321	0.001542101	0.00211208
Chr_Total (1-22, X) (Media)	856,623.6		

Una muestra que tenía una dosis de cromosoma X menor que el límite establecido se identificó como la que tenía menos de un cromosoma X. Se determinó que la misma muestra tenía una dosis de cromosoma Y que era inferior al límite establecido, indicando que la muestra no tenía un cromosoma Y. Por lo tanto, la combinación de dosis de cromosomas para X e Y se usaron para identificar las muestras del Síndrome de Turner (monosomía X).

Por lo tanto, el método divulgado en la presente permite la determinación del CNV de cromosomas. En particular, el método permite la determinación de la sobre- o sub- representación de aneuploidías cromosómicas por secuenciación masivamente paralela de ADN libre de células de plasma materno y la identificación de cromosomas normalizadores para el análisis estadístico de los datos de secuenciación. La sensibilidad y fiabilidad del método permiten probar la aneuploidía en el primer y segundo trimestres con precisión.

Ejemplo 6

Determinación de aneuploidía parcial

El uso de las dosis de secuencia se aplicó para evaluar la aneuploidía parcial en una muestra de ensayo biológica de ADN libre de células que se preparó de plasma sanguíneo, y se secuenció como se describe en el Ejemplo 1. Se confirmó por cariotipado que la muestra se había derivado de un sujeto con una deleción parcial del cromosoma 11.

El análisis de los datos de secuenciación para la aneuploidía parcial (deleción parcial del cromosoma 11, es

decir q21-q23) se realizó como se describe para las aneuploidías cromosómicas en los ejemplos anteriores. El mapeado de las etiquetas de secuencia para el cromosoma 11 en una muestra de ensayo reveló una pérdida considerable de recuentos de etiquetas entre los pares de bases 81000082-103000103 en el brazo q del cromosoma en relación con los recuentos de etiquetas obtenidos para la secuencia correspondiente en el cromosoma 11 en las muestras calificadas (datos no mostrados). Las etiquetas de secuencia mapeadas para la secuencia de interés en el cromosoma 11 (810000082-103000103bp) en cada una de las muestras calificadas, y las etiquetas de secuencia mapeadas para los 20 segmentos de la megabase en el genoma completo en las muestras calificadas, es decir densidades de etiqueta de secuencia calificadas, se usaron para determinar las dosis de secuencia calificadas como proporciones de densidades de etiquetas en todas las muestras calificadas. La dosis de secuencia media, desviación estándar y coeficiente de variación se calcularon para los 20 segmentos de la megabase en el genoma completo y la secuencia de la megabase 20 que tenía menos variabilidad era la secuencia normalizadora identificada en el cromosoma 5 (13000014-33000033bp) (Ver Tabla 8), que se usó para calcular la dosis para la secuencia de interés en la muestra de ensayo (ver Tabla 9). La Tabla 8 proporciona la dosis de secuencia para la secuencia de interés en el cromosoma 11 (810000082-103000103bp) en la muestra de ensayo que se calculó como la proporción de etiquetas de secuencia mapeadas para la secuencia de interés y las etiquetas de secuencia mapeadas para la secuencia normalizadora identificada. La **Figura 16** muestra las dosis de secuencia para la secuencia de interés en las 7 muestras calificadas (O) y la dosis de secuencia para la secuencia correspondiente en la muestra de ensayo (◊). La media se muestra por la línea sólida, y el límite calculado para el diagnóstico positivo de la aneuploidía parcial que se estableció en 5 desviaciones estándar de la media se muestra por la línea discontinua. Un diagnóstico para la aneuploidía parcial se basó en que la dosis de secuencia en la muestra de ensayo era menor que el límite establecido. Se verificó por cariotipado que la muestra de ensayo tenía deleción q21-q23 en el cromosoma 11.

Por lo tanto, además de identificar aneuploidías cromosómicas, el método de la invención se puede usar para identificar aneuploidías parciales.

TABLA 8

Secuencia Normalizadora Cualificada, Dosis y Varianza para la Secuencia Chr11: 81000082-103000103 (muestras calificadas n=7)			
	Chr11:81000082-103000103		
	Media	**Desviación Estándar**	**CV**
Chr5: 13000014-33000033	1.164702	0.004914	0.42

TABLA 9

Dosis de Secuencia para la Secuencia de Interés (81000082-103000103) en el Cromosoma 11 (muestra de ensayo 11206)			
Segmento de Cromosoma	Densidad de Etiqueta de Secuencia	Dosis de Segmento de Cromosoma para Chr 11 (q21-q23)	Límite
Chr11: 81000082-103000103	27,052	1.0434313	1.1401347
Chr5: 13000014-33000033	25,926		

Ejemplo 7

Determinación Simultánea de Aneuploidía y Fracción Fetal por Secuenciación Masivamente Paralela: Selección de SNPs Autosómicos para la Determinación de la Fracción Fetal

Se seleccionó un conjunto de 28 SNPs autosómicos de una lista de 92 SNPs (Pakstis et al., Hum Genet 127:315-324 [2010]), y de las secuencias de SNP disponibles en Applied Biosystems en la dirección de la red mundial appliedbiosystems.com, y validada para su uso en amplificación por PCR multiplexado y para secuenciación masivamente paralela para determinar la fracción fetal determinando o no simultáneamente la presencia o ausencia de aneuploidía. Los cebadores se diseñaron para hibridar con una secuencia cercana con el sitio de SNPs en el ADN libre de células para asegurar que se incluya en la lectura de 36 bp generada de la secuenciación masivamente paralela en el Analizador GII Illumina, y para generar amplicones de longitud suficiente para someterse a amplificación por puente durante la formación de clústeres. Por lo tanto, los cebadores se diseñaron para generar amplicones que fueron de al menos 110 bp, que cuando se combinan con los adaptadores universales (Illumina Inc., San Diego, CA) usados para amplificación de clústeres, resultó en moléculas de ADN de al menos 200 bp. Se identificaron secuencias de cebadores, y los conjuntos de cebadores, es decir cebadores directos e inversos, se

sintetizaron por Integrated DNA Technologies (San Diego, CA), y se almacenaron como una solución de 1µM para ser usada para amplificar secuencias objetivo polimórficas como se describe en los Ejemplos 5-8. La Tabla 10 proporciona los números de ID de entrada de los RefSNP (rs), los cebadores usados para amplificar la secuencia de ADN libre de células objetivo, y las secuencias de los amplicones que comprenden los posibles alelos de SNP que serían generados usando los cebadores. Los SNPs dados en la Tabla 10 se usaron para la amplificación simultanea de 13 secuencias objetivo en un ensayo multiplexado para determinar simultáneamente la fracción fetal y la presencia o ausencia de una aneuploidía en muestras de ADN libre de células derivado de mujeres embarazadas. El panel proporcionado en la Tabla 10 es un panel de SNP ejemplar. Se pueden empelar menos o más SNPs para enriquecer el ADN fetal y materno para ácidos nucleicos objetivo polimórficos. SNPs adicionales que pueden usarse incluyen los SNPs dados en la Tabla 11. Los SNPs en la Tabla 11 han sido validados en amplificaciones por PCR multiplex, y secuenciados usando el analizador Genomell II como se ha descrito anteriormente. Los alelos de SNP en las Tablas 10 y 11 se muestran en negrita y están subrayados.

TABLA 10

Panel SNP para la Determinación de la Fracción Fetal				
SNP ID	**Chr**	**Amplicón: Alelo 1**	**Amplicón: Alelo 2**	**Secuencia del Cebador Directo, nombre y SEQ ID NO:**	**Secuencia del Cebador Inverso, nombre y SEQ ID NO:**
rs560681	1	CACATGCACAGCCA GCAACCCTGTCAGC AGGAGTTCCCACCA GTTTCTTTCTGAGAA CATCTGTTCAGGTTT CTCTCCATCTCTATT TACTCAGGTCACAG GACCTTGGGG (SEQ ID NO:1)	CACATGCACAGCCA GCAACCCTGTCAGC AGGAGTTCCCACCA GTTTCTTTCTGAGAA CATCTGTTCAGGTTT CTCTCCATCTCTGTT TACTCAGGTCACAG GACCTTGGGG (SEQ ID NO:2)	CACATGCA CAGCCAGC AACCC (rs560681_C 1_1_F; SEQ ID NO:57)	CCCCAAGGTC CTGTGACCTGA GT (rs560681_C1_1_ R; SEQ ID NO:58)
rs1109037	2	TGAGGAAGTGAGGC TCAGAGGGTAAGAA ACTTTGTCACAGAGC TGGTGGTGAGGGTG GAGATTTTACACTCC CTGCCTCCCACACCA GTTTCTCCAGAGTGG AAAGACTTTCATCTC GCACTGGCA (SEQ ID NO:3)	TGAGGAAGTGAGGC TCAGAGGGTAAGAA ACTTTGTCACAGAGC TGGTGGTGAGGGTG GAGATTTTACACTCC CTGCCTCCCACACCA GTTTCTCCGGAGTGG AAAGACTTTCATCTC GCACTGGCA (SEQ ID NO:4)	TGAGGAAG TGAGGCTC AGAGGGT (rs110937_C 2_1_F ; SEQ ID NO:59)	TGCCAGTGCG AGATGAAAGT CTTT (rs110937_C2_1_ R ; SEQ ID NO:60)
rs9866013	3	GTGCCTTCAGAACCT TTGAGATCTGATTCT ATTTTTAAAGCTTCT TAGAAGAGAGATTG CAAAGTGGGTTGTTT CTCTAGCCAGACAG GGCAGGCAAATAGG GGTGGCTGGTGGGA TGGGA (SEQ ID NO:5)	GTGCCTTCAGAACCT TTGAGATCTGATTCT ATTTTTAAAGCTTCT TAGAAGAGAGATTG CAAAGTGGGTTGTTT CTCTAGCCAGACAG GGCAGGTAAATAGG GGTGGCTGGTGGGA TGGGA (SEQ ID NO:6)	GTGCCTTC AGAACCTT TGAGATCT GAT (rs9866013_ C3_1_F; SEQ ID NO:61)	TCCCATCCCAC CAGCCACCC (rs9866013_C3_1 _R; SEQ ID NO:62)

Panel SNP para la Determinación de la Fracción Fetal				
SNP ID	**Chr**	**Amplicón: Alelo 1**	**Amplicón: Alelo 2**	**Secuencia del Cebador Directo, nombre y SEQ ID NO:**	**Secuencia del Cebador Inverso, nombre y SEQ ID NO:**
rs13182883	5	AGGTGTGTCTCTCTT TTGTGAGGGGAGGG GTCCCTTCTGGCCTA GTAGAGGGCCTGGC CTGCAGTGAGCATTC AAATCCTCAAGGAA CAGGGTGGGGAGGT GGGACAAAGG (SEQ ID NO:7)	AGGTGTGTCTCTCTT TTGTGAGGGGAGGG GTCCCTTCTGGCCTA GTAGAGGGCCTGGC CTGCAGTGAGCATTC AAATCCTCGAGGAA CAGGGTGGGGAGGT GGGACAAAGG (SEQ ID NO:8)	AGGTGTGT CTCTCTTT GTGAGGGG (rs13182883_ C5_1_F; SEQ ID NO:63)	CCTTTGTCCCA CCTCCCCACC (rs13182883_C5_ 1_R; SEQ ID NO:64)
rs13218440	6	CCTCGCCTACTGTGC TGTTTCTAACCATCA TGCTTTTCCCTGAAT CTCTTGAGTCTTTTT CTGCTGTGGACTGA AACTTGATCCTGAG ATTCACCTCTAGTCC CTCTGAGCAGCCTCC TGGAATACTCAGCT GGGATGG (SEQ ID NO:9)	CCTCGCCTACTGTGC TGTTTCTAACCATCA TGCTTTTCCCTGAAT CTCTTGAGTCTTTTT CTGCTGTGGACTGA AACTTGATCCTGAG ATTCACCTCTAGTCC CTCTGGGCAGCCTCC TGGAATACTCAGCT GGGATGG (SEQ ID NO:10)	CCTCGCCT ACTGTGCT GTTTCTAA CC (rs13218440_ C6_1_F; SEQ ID NO:65)	CCATCCCAGCT GAGTATTCCA GGAG (rs13218440_C6_ 1_R; SEQ ID NO:66)
rs7041158	9	AATTGCAATGGTGA GAGGTTGATGGTAA AATCAAACGGAACT TGTTATTTTGTCATT CTGATGGACTGGAA CTGAGGATTTTTCAAT TTCCTCTCCAACCCA AGACACTTCTCACTG G (SEQ ID NO:11)	AATTGCAATGGTGA GAGGTTGATGGTAA AATCAAACGGAACT TGTTATTTTGTCATT CTGATGGACTGGAA CTGAGGATTTTTCAAT TTCCTTTCCAACCCA AGACACTTCTCACTG G (SEQ ID NO:12)	AATTGCAA TGGTGAGA GGTTGATG GT (SEQ ID NO:67)	CCAGTGAGAA GTGTCTTGGGT TGG (SEQ ID NO:68)

Panel SNP para la Determinación de la Fracción Fetal				
SNP ID	**Chr**	**Amplicón: Alelo 1**	**Amplicón: Alelo 2**	**Secuencia del Cebador Directo, nombre y SEQ ID NO:**	**Secuencia del Cebador Inverso, nombre y SEQ ID NO:**
rs740598	10	GAAATGCCTTCTCAG GTAATGGAAGGTTA TCCAAATATTTTTCG TAAGTATTTCAAATA GCAATGGCTCGTCTA TGGTTAGTCTCACAG CCACATTCTCAGAAC TGCTCAAACC (SEQ ID NO:13)	GAAATGCC TTCTCAGG TAATGGAA GGT (SEQ ID NO:69)	GGTTTGAGCA GTTCTGAGAAT GTGGCT (SEQ ID NO:70)
rs10773760	12	ACCCAAAACACTGG AGGGGCCTCTTCTCA TTTTCGGTAGACTGC AAGTGTTAGCCGTC GGGACCAGCTTCTGT CTGGAAGTTCGTCA AATTGCAGTTAAGTC CAAGTATGCCACAT AGCAGATAAGGG (SEQ ID NO:15)	ACCCAAAA CACTGGAG GGGCCT (SEQ ID NO:71)	CCCTTATCTGC TATGTGGCATA CTTGG (SEQ ID NO:72)
rs4530059	14	GCACCAGAATTTAA ACAACGCTGACAAT AAATATGCAGTCGA TGATGACTTCCCAGA GCTCCAGAAGCAAC TCCAGCACACAGAG AGGCGCTGATGTGC CTGTCAGGTGC (SEQ ID NO:17)	GCACCAGA ATTTAAAC AACGCTGA CAA (SEQ ID NO:73)	GCACCTGACA GGCACATCAG CG (SEQ ID NO:74)

Panel SNP para la Determinación de la Fracción Fetal				
SNP ID	**Chr**	**Amplicón: Alelo 1**	**Amplicón: Alelo 2**	**Secuencia del Cebador Directo, nombre y SEQ ID NO:**	**Secuencia del Cebador Inverso, nombre y SEQ ID NO:**
rs7205345	16	TGACTGTATACCCCA GGTGCACCCTTGGGT CATCTCTATCATAGA ACTTATCTCACAGAG TATAAGAGCTGATTT CTGTGTCTGCCTC_TC ACACTAGACTTCCAC ATCCTTAGTGC (SEQ ID NO:19)	TGACTGTATACCCCA GGTGCACCCTTGGGT CATCTCTATCATAGA ACTTATCTCACAGAG TATAAGAGCTGATTT CTGTGTCTGCCTG_TC ACACTAGACTTCCAC ATCCTTAGTGC (SEQ ID NO:20)	TGACTGTA TACCCCAG GTGCACCC (SEQ ID NO:75)	GCACTAAGGA TGTGGAAGTCT AGTGTG (SEQ ID NO:76)
rs8078417	17	TGTACGTGGTCACCA GGGGACGCCTGGCG CTGCGAGGGAGGCC CCGAGCCTCGTGCCC CCGTGAAGCTTCAG CTCCCCTCCCC_GGCT GTCCTTGAGGCTCTT CTCACACT (SEQ ID NO:21)	TGTACGTGGTCACCA GGGGACGCCTGGCG CTGCGAGGGAGGCC CCGAGCCTCGTGCCC CCGTGAAGCTTCAG CTCCCCTCCCT_GGCT GTCCTTGAGGCTCTT CTCACACT (SEQ ID NO:22)	TGTACGTG GTCACCAG GGGACG (SEQ ID NO:77)	AGTGTGAGAA GAGCCTCAAG GACAGC (SEQ ID NO:78)
rs576261	19	CAGTGGACCCTGCT GCACCTTTCCTCCCC TCCCATCAACCTCTT TTGTGCCTCCCCCTC CGTGTACCACCTTCT CTGTCACCAA_CCCTG GCCTCACAACTCTCT CCTTTGCCAC (SEQ ID NO:23)	CAGTGGACCCTGCT GCACCTTTCCTCCCC TCCCATCAACCTCTT TTGTGCCTCCCCCTC CGTGTACCACCTTCT CTGTCACCAC_CCCTG GCCTCACAACTCTCT CCTTTGCCAC (SEQ ID NO:24)	CAGTGGAC CCTGCTGC ACCTT (SEQ ID NO:79)	GTGGCAAAGG AGAGAGTTGT GAGG (SEQ ID NO:80)

Panel SNP para la Determinación de la Fracción Fetal				
SNP ID	**Chr**	**Amplicón: Alelo 1**	**Amplicón: Alelo 2**	**Secuencia del Cebador Directo, nombre y SEQ ID NO:**	**Secuencia del Cebador Inverso, nombre y SEQ ID NO:**
rs2567608	20	CAGTGGCATAGTAG TCCAGGGGCTCCTCC TCAGCACCTCCAGC ACCTTCCAGGAGGC AGCAGCGCAGGCAG AGAACCCGCTGGAA GAATCGGCGGAAGT TGTCGGAGAGG (SEQ ID NO:25)	CAGTGGCATAGTAG TCCAGGGGCTCCTCC TCAGCACCTCCAGC ACCTTCCAGGAGGC AGCAGCGCAGGCAG AGAACCCGCTGGAA GGATCGGCGGAAGT TGTCGGAGAGG (SEQ ID NO:26)	CAGTGGCA TAGTAGTC CAGGGGCT (SEQ ID NO:81)	CCTCTCCGACA ACTTCCGCCG (SEQ ID NO:82)

TABLA 11

SNPs Adicionales para la Determinación de la Fracción Fetal

SNP ID	Chr	Amplicón: Alelo 1	Amplicón: Alelo 2	Secuencia del Cebador Directo, nombre y SEQ ID NO:	Secuencia del Cebador Inverso, nombre y SEQ ID NO:
rs430046	16	AGGTCTGGGGGCC GCTGAATGCCAAGC TGGGAATCTTAAAT GTTAAGGAACAAG GTCATACAATGAAT GGTGTGATGTAAAA GCTTGGGAGGTGAT TTCTGAGGGTAGGT GCTGGGTTTAATGG GAGGA (SEQ ID NO:27)	AGGTCTGGGGGCCGC TGAATGCCAAGCTGG GAATCTTAAATGTTA AGGAACAAGGTCATA CAATGAATGGTGTGA TGTAAAAGCTTGGGA GGTGATTTTGAGGG TAGGTGCTGGGTTTA ATGGGAGGA (SEQ ID NO:28)	AGGTCTGG GGGCCGCT GAAT (rs430046_C 1_1_F; SEQ ID NO:83)	TCCTCCCATTA AACCCAGCAC CT (rs430046_C1_1_ R; SEQ ID NO:84)
rs9951171	18	ACGGTTCTGTCCTG TAGGGGAGAAAAG TCCTCGTTGTTCCT CTGGGATGCAACAT GAGAGAGCAGCAC ACTGAGGCTTTATG GATTGCCCTGCCAC AAGTGAACAGG (SEQ ID NO:29)	ACGGTTCTGTCCTGT AGGGGAGAAAAGTCC TCGTTGTTCCTCTGGG ATGCAACATGAGAGA GCAGCACACTGAGGC TTTATGGGTTGCCCTG CCACAAGTGAACAGG (SEQ ID NO:30)	ACGGTTCT GTCCTGTA GGGGAGA (rs9951171_ C1_1_F; SEQ ID NO:85)	CCTGTTCACTT GTGGCAGGGC A (rs9951171_C1_1 _R; SEQ ID NO:86)
rs338882	5	GCGCAGTCAGATG GGCGTGCTGGCGTC TGTCTTCTCTCTCTC CTGCTCTCTGGCTT CATTTTTCTCTCCTT CTGTCTCACCTTCT TTCGTGTGCCTGTG CACACACACGTTTG GGACAAGGG CTGGA (SEQ ID NO:31)	GCGCAGTCAGATGGG CGTGCTGGCGTCTGT CTTCTCTCTCTCCTGC TCTCTGGCTTCATTTT TCTCTCCTTCTGTCTC ACCTTCTTTCGTGTGC CTGTGCATACACACG TTTGGGACAAGGG CTGGA (SEQ ID NO:32)	GCGCAGTC AGATGGGC GTGC (rs338882_C 1_1_F; SEQ ID NO:87)	TCCAGCCCTTG TCCCAAACGT GT (rs338882_C1_1_ R; SEQ ID NO:88)

SNPs Adicionales para la Determinación de la Fracción Fetal

SNP ID	Chr	Amplicón: Alelo 1	Amplicón: Alelo 2	Secuencia del Cebador Directo, nombre y SEQ ID NO:	Secuencia del Cebador Inverso, nombre y SEQ ID NO:
rs10776839	9	GCCGGACCTGCGA AATCCCAAAATGCC AAACATTCCCGCCT CACATGATCCCAGA GAGAGGGGACCCA GTGTTCCCAGCTTG CAGCTGAGGAGCC CGAGGTTGCCGTCA GATCAGAGCCCCA GTTGCCCG (SEQ ID NO:33)	GCCGGACCTGCGAAA TCCCAAAATGCCAAA CATTCCCGCCTCACA TGATCCCAGAGAGAG GGGACCCAGTGTTCC CAGCTTGCAGCTGAG GAGCCCGAGTTTGCC GTCAGATCAGAGCCC CAGTTGCCCG (SEQ ID NO:34)	GCCGGACC TGCGAAAT CCCAA (rs10776839 C1_1_F; SEQ ID NO:89)	CGGGCAACTG GGGCTCTGATC (rs10776839_C1_ 1_R; SEQ ID NO:90)
rs9905977	17	AGCAGCCTCCCTCG ACTAGCTCACACTA CGATAAGGAAAAT TCATGAGCTGGTGT CCAAGGAGGGCTG GGTGACTCGTGGCT CAGTCAGCATCAAG ATTCCTTTCGTCTTT CCCCTCTGCC (SEQ ID NO:35)	AGCAGCCTCCCTCGA CTAGCTCACACTACG ATAAGGAAAATTCAT GAGCTGGTGTCCAAG GAGGGCTGGGTGACT CGTGGCTCAGTCAGC GTCAAGATTCCTTTC GTCTTTCCCCTCTGCC (SEQ ID NO:36)	AGCAGCCT CCCTCGAC TAGCT (rs9905977_ C1_1_F; SEQ ID NO:91)	GGCAGAGGGG AAAGACGAAA GGA (rs9905977_C1_1 _R; SEQ ID NO:92)
rs1277284	4	TGGCATTGCCTGTA ATATACATAGCCAT GGTTTTTTATAGGC AATTTAAGATGAAT AGCTTCTAAACTAT AGATAAGTTTCATT ACCCCAGGAAGCT GAACTATAGCTACT TTACCCAAAATCAT TAGAATGGTGCTT (SEQ ID NO:37)	TGGCATTGCCTGTAA TATACATAGCCATGG TTTTTTATAGGCAATT TAAGATGAATAGCTT CTAAACTATAGATAA GTTTCATTACCCCAG GAAGCTGAACTATAG CTACTTTCCCCAAAA TCATTAGAATGGTGC TT (SEQ ID NO:38)	TGGCATTG CCTGTAAT ATACATAG (rs1277284_ C4_1_F; SEQ ID NO:93)	AAGCACCATT CTAATGATTTT GG (rs1277284_C4_1 _R; SEQ ID NO:94)

SNPs Adicionales para la Determinación de la Fracción Fetal

SNP ID	Chr	Amplicón: Alelo 1	Amplicón: Alelo 2	Secuencia del Cebador Directo, nombre y SEQ ID NO:	Secuencia del Cebador Inverso, nombrey SEQ ID NO:
rs258684	7	ATGAAGCCTTCCAC CAACTGCCTGTATG ACTCATCTGGGGAC TTCTGCTCTATACT CAAAGTGGCTTAGT CACTGCCAATGTAT TTCCATATGAGGGA CGATGATTACTAAG GAAATATAGAAAC AACAACTGATC (SEQ ID NO:39)	ATGAAGCCTTCCACC AACTGCCTGTATGAC TCATCTGGGGACTTC TGCTCTATACTCAAA GTGGCTTAGTCACTG CCAATGTATTTCCAT ATGAGGGACGGTGAT TACTAAGGAAATATA GAAACAACAACTGAT C (SEQ ID NO:40)	ATGAAGCC TTCCACCA ACTG (rs258684_C 7_1_F; SEQ ID NO:95)	GATCAGTTGTT GTTTCTATATT TCCTT (rs258684_C7_1_ R; SEQ ID NO:96)
rs1347696	8	ACAACAGAATCAG GTGATTGGAGAAA AGATCACAGGCCTA GGCACCCAAGGCTT GAAGGATGAAAGA ATGAAAGATGGAC GGAACAAAATTAG GACCTTAATTCTTT GTTCAGTTCAG (SEQ ID NO:41)	ACAACAGAATCAGGT GATTGGAGAAAAGAT CACAGGCCTAGGCAC CCAAGGCTTGAAGGA TGAAAGAATGAAAGA TGGACGGAAGAAAAT TAGGACCTTAATTCTT TGTTCAGTTCAG (SEQ ID NO:42)	ACAACAGA ATCAGGTG ATTGGA (rs1347696_ C8_4_F; SEQ ID NO:97)	CTGAACTGAA CAAAGAATTA AGGTC (rs1347696_C8_4 _F; SEQ ID NO:98)
rs508485	11	TTGGGGTAAATTTT CATTGTCATATGTG GAATTTAAATATAC CATCATCTACAAAG AATTCCACAGAGTT AAATATCTTAAGTT AAACACTTAAAATA AGTGTTTGCGTGAT ATTTTGATGACAGA TAAACAGAGTCTAA TTCCCACCCC (SEQ ID NO:43)	TTGGGGTAAATTTTC ATTGTCATATGTGGA ATTTAAATATACCAT CATCTACAAAGAATT CCACAGAGTTAAATA TCTTAAGTTAAACAC TTAAAATAAGTGTTT GCGTGATATTTTGAT GATAGATAAACAGAG TCTAATTCCCACCCC (SEQ ID NO:44)	TTGGGGTA AATTTTCA TTGTCA (rs508485_C 11_1_F; SEQ ID NO:99)	GGGGTGGGAA TTAGACTCTG (rs508485_C11_1 _R; SEQ ID NO100)

SNPs Adicionales para la Determinación de la Fracción Fetal

SNP ID	Chr	Amplicón: Alelo 1	Amplicón: Alelo 2	Secuencia del Cebador Directo, nombre y SEQ ID NO:	Secuencia del Cebador Inverso, nombre y SEQ ID NO:
rs9788670	15	TGCAATTCAAATCA GGAAGTATGACCA AAAGACAGAGATC TTTTTTGGATGATC CCTAGCCTAGCAAT GCCTGGCAGCCATG CAGGTGCAATGTCA ACCTTAAATAATGT ATTGCAAACTCAGA GCTGACAAACCTCG ATGTTGC (SEQ ID NO:45)	TGCAATTCAAATCAG GAAGTATGACCAAAA GACAGAGATCTTTTT TGGATGATCCCTAGC CTAGCAATGCCTGGC AGCCATGCAGGTGCA ATGTCAACCTTAAAT AATGTATTGCAAATT CAGAGCTGACAAACC TCGATGTTTGC (SEQ ID NO:46)	TGCAATTC AAATCAGG AAGTATG (rs9788670_c 15_2_F; SEQ ID NO:101)	GCAACATCGA GGTTTGTCAG (rs9788670_c15_ 2_R; SEQ ID NO:102)
rs8137254	22	CTGTGCTCTGCGAA TAGCTGCAGAAGTA ACTTGGGGACCCAA AATAAAGCAGAAT GCTAATGTCAAGTC CTGAGAACCAAGC CCTGGGACTCTGGT GCCATTTCGGATTC TCCATGAGCATGGT (SEQ ID NO:47)	CTGTGCTCTGCGAAT AGCTGCAGAAGTAAC TTGGGGACCCAAAAT AAAGCAGAATGCTAA TGTCAAGTCCTGAGA ACCAAGCCCTGGGAC TCTGGTGCCATTTTG ATTCTCCATGAGCAT GGT (SEQ ID NO:48)	CTGTGCTC TGCGAATA GCTG (rs8137254_c 22_2_F: SEQ ID NO:103)	ACCATGCTCAT GGAGAATCC (rs8137254_c22_ 2_R; SEQ ID NO:104)
rs3143	19	TTTTTCCAGCCAAC TCAAGGCCAAAAAA AAATTTCTTAATAT AGTTATTATGCGAG GGGAGGGGAAGCA AAGGAGCACAGGT AGTCCACAGAATA AGACACAAGAAAC CTCAAGCTGTG (SEQ ID NO:49)	TTTTTCCAGCCAACTC AAGGCCAAAAAAAAT TTCTTAATATAGTTAT TATGCGAGGGGAGGG GAAGCAAAGGAGCA CAGGTAGTCCACAGA ATAGGACACAAGAAA CCTCAAGCTGTG (SEQ ID NO:50)	TTTTTCCA GCCAACTC AAGG (rs3143_c19_ 2_F: SEQ ID NO:105)	CACAGCTTGA GGTTTCTTGTG (rs3143_c19_2_R; SEQ ID NO:106)

SNPs Adicionales para la Determinación de la Fracción Fetal					
SNP ID	Chr	Amplicón: Alelo 1	Amplicón: Alelo 2	Secuencia del Cebador Directo, nombre y SEQ ID NO:	Secuencia del Cebador Inverso, nombre y SEQ ID NO:
rs2182957	13	TCTTCTCGTCCCCTAAAGCAAACAACATCCGCTTGCTTCTGTCTGTGTAACCACACAGTGAATGGGTGTGCACGCTTGATGGGCCTCTGAGCCCCTGTTGCACAAACCAGAAA (SEQ ID NO:51)	TCTTCTCGTCCCCTAAAGCAAACAACATCCGCTTGCTTCTGTCTGTGTAACCACAGTGAATGGGTGTGCACGCTTGGTGGGCCTCTGAGCCCCTGTTGCACAAACCAGAAA (SEQ ID NO:52)	TCTTCTCG TCCCCTAA GCAA (rs2182957_c 13_1_F; SEQ ID NO:107)	TTTCTGGTTTG TGCAACAGG (rs2182957_c13_ 1_R; SEQ ID NO:108)
rs3739005	2	CACATGGGGGCATTAAGAATCGCCCAGGGAGGAGGAGGGAGAACGCGTGCTTTTCACATTTGCATTTGAATTTTTGAGTTCCCAGGATGTGTTTTTTGTGCTCATCGATGT (SEQ ID NO:53)	CACATGGGGGCATTAAGAATCGCCCAGGGAGGAGGAGGGAGAACGCGTGCTTTTCACATTTGAATTTTTGAGTTCCCAGGATGTGTTTTTTGTGCTCATCGA TGT (SEQ ID NO:54)	CACATGGG GGCATTAA GAAT (rs3739005_c 2_2_F; SEQ ID NO:109)	ACATCGATGA GCACAAAAAC AC (rs3739005_c2_2 _R; SEQ ID NO:110)
rs530022	1	GGGCTCTGAGGTGTGGTGAAATAAAAACAAAAATGTCCATGTCTGTCCTTTTATGGCATTTTTGGGACTTTACATTTCAAACATTTCAGACATGTATCACACACGAAGGAATAACAGTTCCAGGGATATCT (SEQ ID NO:55)	GGGCTCTGAGGTGTGTGAAATAAAAACAAATGTCCATGTCTGTCCTTTATGGCATTTTTGGGACTTTACATTTCAAACATTTCAGACATGTATCACACACGAGGGAATAACAGTTCCAGGG ATATCT (SEQ ID NO:56)	GGGCTCTG AGGTGTGT GAAA (rs530022_c1 _2_F; SEQ ID NO:111)	AGATATCCCTG GAACTGTTATT CC (rs530022_c1_2_ R; SEQ ID NO:112)

Ejemplo 8

Determinación Simultánea de Aneuploidía y fracción Fetal: Enriquecimiento de ácidos nucleicos fetales y maternos en una muestra de biblioteca de secuenciación de ADN libre de células.

Para determinar simultáneamente la fracción fetal y la presencia o ausencia de aneuploidía en una muestra materna se enriqueció una biblioteca de secuenciación primaria de ácidos nucleicos fetales y maternos para las secuencias de ácidos nucleicos objetivo polimórficas, y se secuenciaron de la manera siguiente:

Se preparó ADN libre de células purificado de una muestra de plasma materno como se describe en el Ejemplo 1. Una primera porción de la muestra de ADN libre de células purificado se usó para preparar una biblioteca de secuenciación primaria usando el protocolo abreviado descrito en el Ejemplo 2. Una segunda porción de la muestra de ADN libre de células purificado se usó para amplificar secuencias de ácidos nucleicos objetivo polimórficos, es decir SNPs, y preparar una biblioteca de secuenciación objetivo de la manera siguiente. El ADN libre de células contenido en 5µl de ADN libre de células purificado se amplificó en un volumen de reacción que contenía 7,5µl de una mezcla de cebadores de 1 µM (Tabla 5), 10 µl de Mezcla maestra NEXB 5X y 27 µl de agua. Se realizó ciclado térmico con el Gene Amp9700 (Applied Biosystems). Usando las siguientes condiciones de ciclado: incubación a 95º C durante 1 minuto, seguido por 30 ciclos a 95º C durante 20 segundos, 68ºº C durante 1 minuto, y 68º C durante 30 segundos, a lo que siguió una incubación final a 68º C durante 5 minutos. Se añadió un mantenimiento final a 4º C hasta que se retiraron las muestras para combinar con la porción no amplificada de la muestra de ADN libre de células purificado. El producto amplificado se purificó usando el sistema de purificación por PCR Agencourt AMPure XP (Parte No. A63881; Beckman Coulter Genomics, Danvers, MA). Se añadió un mantenimiento final a 4º C hasta que se retiraron las muestras para preparar la biblioteca objetivo. El producto amplificado se analizó con un Bioanalizador 2100 (Agilent Technologies, Sunnyvale, CA) y se determinó la concentración del producto amplificado. Un quinto del producto amplificado purificado se usó para preparar una biblioteca de secuenciación objetivo de ácidos nucleicos polimórficos amplificados como se describe en el Ejemplo 2. Las bibliotecas de secuenciación primaria y objetivo fueron cada una diluidas a 10 nM, y la biblioteca objetivo se combinó a una proporción de 1:9 con la biblioteca de secuenciación para proporcionar una biblioteca de secuenciación enriquecida. La secuenciación de la biblioteca enriquecida se realizó como se describe en el Ejemplo. El análisis de los datos de secuenciación para determinar la aneuploidía se realizó como se describe en el Ejemplo 3 usando el genoma humano hg18 como genoma de referencia. El análisis de los datos de secuenciación para determinar la fracción fetal se realizó de la manera siguiente. Concomitante al análisis para determinar aneuploidía, los datos de secuenciación se analizaron para determinar la fracción fetal. Después de la transferencia de los archivos de imagen y designación de bases al servidor Unix ejecutando el software Illumina "Genome Analyzer Pipeline" versión 1.51 como se describe en el Ejemplo 2c, las lecturas de 36bp se alinearon con un 'genoma de SNP' usando el programa BOWTIE. El genoma de SNP se identificó como la agrupación de las secuencias de ADN polimórficas, es decir las SEQ ID NOS:1-56, que abarcan los alelos de las 13 SNP divulgadas en la Tabla 10 en el Ejemplo 7. Sólo se usaron lecturas que mapeaban únicamente para el genoma de SNP para el análisis de la fracción fetal. Las lecturas que equivalían perfectamente con el genoma de SNP se contaron como etiquetas y se filtraron. De las lecturas restantes, sólo se contaron como etiquetas y se incluyeron en el análisis lecturas que tenían una o dos discrepancias. Se contaron las etiquetas mapeadas para cada uno de los alelos de SNP, y se determinó la fracción fetal. Alrededor de un millón del número total de etiquetas de secuencia obtenidas de secuenciar la biblioteca enriquecida correspondían a etiquetas que mapeaban para el genoma de referencia de SNP. La **Figura 17** muestra un gráfico de la proporción del número de etiquetas de secuencia mapeadas para cada cromosoma y el número total de etiquetas mapeadas para todos los cromosomas (1-22, X e Y) obtenidas de la secuenciar una biblioteca de ADN libre de células no enriquecida (●), y biblioteca de ADN libre de células enriquecida con un 5% (■) o 10% (♦) de biblioteca de SNP multiplex amplificada. El gráfico indica que combinar una biblioteca de secuencias polimórficas amplificadas con una biblioteca de secuencias no amplificadas de la muestra materna no afecta a la información de secuenciación usada para determinar aneuploidía. Los ejemplos de determinación de la fracción fetal para muestras obtenidas de sujetos que llevan un feto con una aneuploidía cromosómica se dan en las Tablas 12, 13 y 14 siguientes.

a. Determinación de la fracción fetal

La fracción fetal se calculó como:

$$\%\ \text{de alelo}_x\ \text{de la fracción fetal} = ((\Sigma\text{Etiquetas de secuencia fetal para el alelo}_x) / (\Sigma\text{Etiquetas de secuencia materna para el alelo}_x)) \times 100$$

donde el alelo_x es un alelo informativo.

TABLA 12

Determinación Simultánea de Aneuploidía y Fracción Fetal: Determinación de la Fracción Fetal			
ID de la Muestra (cariotipo)	**SNP**	**RECUENTOS DE ETIQUETAS DE SNP**	**FRACCION FETAL (%)**
11409 (47, XY+21)	rs13182883.1|Chr.5|longitud=111|alelo=A	261	4.4 1
	rs13182883.2|Chr.5|longitud=111|alelo=G	5918	
	rs740598.1|Chr.10|longitud=114|alelo=A	5545	7.30
	rs740598.2|Chr.10|longitud=114|alelo=G	405	
	rs8078417.1|Chr.17|longitud=110|alelo=C	8189	6.74
	rs8078417.2|Chr.17|longitud=110|alelo=T	121470	
	rs576261.1|Chr.19|longitud=114|alelo=A	58342	7.62
	rs576261.2|Chr.19|longitud=114|alelo=C	4443	
Fracción Fetal (Media±D.S.) = 6.53±1.45			
ID de la muestra			
95133 (47, XX+18)	rs1109037.1|Chr.2|longitud=126|alelo=A	12229	2.15
	rs1109037.2|Chr.2|longitud=126|alelo=G	263	
	rs13218440.1|Chr.6|longitud=139|alelo=A	55949	3.09
	rs13218440.2|Chr.6|longitud=139|alelo=G	1729	
	rs7041158.1|Chr.9|longitud=117|alelo=C	7281	4.1 2
	rs7041158.2|Chr.9|longitud=117|alelo=T	300	
	rs7205345.1|Chr.16|longitud=116|alelo=C	53999	2.14
	rs7205345.2|Chr.16|longitud=116|alelo=G	1154	
Fracción Fetal (Media±D.S.) = 2.9±0.9			
ID de la muestra			
51236 (46,XY+13)	rs13218440.1|Chr.6|longitud=139|alelo=A	1119	1.65
	rs13218440.2|Chr.6|longitud=139|alelo=G	67756	
	rs560681.1|Chr.1|longitud=111|alelo=A	14123	5.18
	rs560681.2|Chr.1|longitud=111|alelo=G	732	
	rs7205345.1|Chr.16|longitud=116|alelo=C	18176	1.63
	rs7205345.2|Chr.16|longitud=116|alelo=G	296	
	rs9866013.1|Chr.3|longitud=121|alelo=C	117	2.33
	rs9866013.2|Chr.3|longitud=121|alelo=T	5024	
Fracción Fetal (Media±D.S.) = 2.7±1.7			

(continuada)

ID de la muestra			
54430 (45,XO)	rs1109037.1|Chr.2|longitud=126|alelo=A	19841	1.80
	rs1109037.2|Chr.2|longitud=126|alelo=G	357	
	rs9866013.1|Chr.3|longitud=121|alelo=C	12931	3.81
	rs9866013.2|Chr.3|longitud=121|alelo=T	493	
	rs7041158.1|Chr.9|longitud=117|alelo=C	2800	4.25
	rs7041158.2|Chr.9|longitud=117|alelo=T	119	
	rs740598.1|Chr.10|longitud=114|alelo=A	12903	4.87
	rs740598.2|Chr.10|longitud=114|alelo=G	628	
	rs10773760.1|Chr.12|longitud=128|alelo=A	46324	4.65
	rs10773760.2|Chr.12|longitud=128|alelo=G	2154	
Fracción Fetal (Media±D.S.) = 3.961.2			

b. Determinación de aneuploidía

La determinación de aneuploidía de los cromosomas 21, 13, 18 y X se realizó usando dosis de cromosomas como se describe en el Ejemplo 4. La dosis de cromosomas calificada, la varianza y la diferenciabilidad para los cromosomas 21, 18, 13, X e Y se dan en las Tablas X e Y. La clasificación de los cromosomas normalizadores identificados por dosis de cromosomas determinadas a partir de la secuenciación de la biblioteca enriquecida fue la misma que la determinada a partir de secuenciar una biblioteca primaria (no enriquecida) del Ejemplo 4. La **Figura 17** muestra que la secuenciación de una biblioteca que ha sido enriquecida para secuencias objetivo polimórficas, por ejemplo SNPs, no se ve afectada por la inclusión de productos de SNP amplificados.

TABLA 13

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 21 y 18								
	21 (n=35)				**18 (n=40)**			
	Media	Desviación Estándar	CV	Prueba T	Media	Desviación Estándar	CV	Prueba T
chr1	0.15332	0.002129	1.39	1.06E-10	0.32451	0.008954	2.76	2.74E-03
chr2	0.15106	0.002053	1.36	8.52E-08	0.31984	0.001783	0.56	5.32E-05
chr3	0.18654	0.004402	2.36	8.07E-07	0.39511	0.002364	0.60	1.93E-05
chr4	0.21578	0.011174	5.18	1.47E-04	0.45714	0.014794	3.24	1.37E-03
chr5	0.21068	0.005332	2.53	1.08E-06	0.44626	0.003250	0.73	3.18E-05
chr6	0.22112	0.005453	2.47	1.74E-06	0.46818	0.003434	0.73	2.24E-05
chr7	0.24233	0.002314	0.96	2.39E-08	0.51341	0.005289	1.03	1.24E-04
chr8	0.24975	0.003772	1.51	1.06E-07	0.52898	0.002161	0.41	6.32E-05
chr9	0.31217	0.003050	0.98	1.60E-09	0.66100	0.014413	2.18	8.17E-04
chr10	0.25550	0.003164	1.24	2.42E-11	0.54091	0.013953	2.58	2.26E-03
chr11	0.26053	0.002596	1.00	1.32E-10	0.55158	0.013283	2.41	1.29E-03
chr12	0.27401	0.002061	0.75	1.40E-08	0.58032	0.007198	1.24	1.57E-04
chr13	0.41039	0.017637	4.30	3.09E-05	0.86961	0.021614	2.49	2.36E-04
chr14	0.40482	0.002908	0.72	1.10E-08	0.85732	0.011748	1.37	2.16E-04
chr15	0.41821	0.008238	1.97	1.24E-10	0.88503	0.029199	3.30	5.72E-03

(continuada)

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 21 y 18								
	21 (n=35)				18 (n=40)			
	Media	Desviación Estándar	CV	Prueba T	Media	Desviación Estándar	CV	Prueba T
chr16	0.40668	0.021232	5.22	2.91E-05	0.86145	0.056245	6.53	1.04E-01
chr17	0.42591	0.027001	6.34	5.85E-04	0.90135	0.068151	7.56	1.24E-01
chr18	0.46529	0.016239	3.49	8.02E-09				
chr19	0.63003	0.063272	10.04	3.30E-02	1.33522	0.150794	11.29	3.04E-01
chr20	0.49925	0.023907	4.79	1.65E-05	1.05648	0.064440	6.10	7.98E-02
chr21					2.06768	0.087175	4.22	5.10E-05
chr22	0.88726	0.083330	9.39	3.43E-02	1.87509	0.198316	10.58	2.43E-01
chrX	0.27398	0.016109	5.88	1.16E-04	0.58665	0.027280	4.65	7.50E-02

TABLA 14

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 13, X e Y								
	13 (n=47)				X (n=20)			
	Media	Desviación Estándar	CV	Dif.	Media	Desviación Estándar	CV	Prueba T
chr1	0.37213	0.018589	5.00	2.41	0.58035	0.02706	4.66	5.68E-05
chr2	0.36707	0.010067	2.74	3.03	0.57260	0.01432	2.50	1.53E-09
chr3	0.45354	0.008121	1.79	3.67	0.70741	0.01126	1.59	9.04E-13
chr4	0.52543	0.005306	1.01	2.39	0.82144	0.01192	1.45	5.86E-16
chr5	0.51228	0.008273	1.61	3.95	0.79921	0.01100	1.38	2.32E-13
chr6	0.53756	0.008901	1.66	3.91	0.83880	0.01261	1.50	3.64E-13
chr7	0.58908	0.018508	3.14	2.83	0.91927	0.02700	2.94	1.86E-08
chr8	0.60695	0.015797	2.60	3.05	0.94675	0.02173	2.30	3.40E-10
chr9	0.75816	0.033107	4.37	2.59	1.18180	0.04827	4.08	9.63E-06
chr10	0.62018	0.029891	4.82	2.56	0.96642	0.04257	4.40	4.55E-05
chr11	0.63248	0.029204	4.62	2.55	0.98643	0.04222	4.28	1.82E-05
chr12	0.66574	0.023047	3.46	2.76	1.03840	0.03301	3.18	1.26E-07
chr13					1.56355	0.01370	0.88	6.33E-17
chr14	0.98358	0.035331	3.59	2.67	1.58114	0.08076	5.11	2.29E-04
chr15	1.01432	0.055806	5.50	2.39	1.53464	0.12719	8.29	2.01E-02
chr16	0.98577	0.085933	8.72	2.17	1.61094	0.14829	9.21	2.68E-02
chr17	1.03217	0.100389	9.73	2.13	1.74904	0.07290	4.17	1.62E-04
chr18	1.13489	0.040058	3.53	2.62	2.38397	0.30515	12.80	1.07E-01
chr19	1.52678	0.203732	13.34	1.98	1.88186	0.14674	7.80	1.56E-02
chr20	1.20919	0.100371	8.30	2.27	3.71853	0.22406	6.03	4.21E-04
chr21	2.38087	0.132418	5.56	2.29	3.35158	0.40246	12.01	8.66E-02
chr22	2.14557	0.271281	12.64	2.13	0.58035	0.02706	4.66	5.68E-05

(continuada)

Dosis de Cromosoma Calificada, Varianza y Diferenciabilidad para los cromosomas 13, X e Y								
	13 (n=47)				X (n=20)			
	Media	Desviación Estándar	CV	Dif.	Media	Desviación Estándar	CV	Prueba T
chrX	0.66883	0.029157	4.36	1.04				
chr2-6	0.46965	0.006987	1.49	4.17				
chr3-6	0.50496	0.005373	1.06	5.16				
	Y (n=25)							
	Media	Desviación Estándar	CV	Prueba T				
Chr 1-22, X	0.00728	0.00227	31.19	1.30E-13				

La dosis de cromosoma 21 se determinó usando el cromosoma 14 como el cromosoma normalizador; la dosis de cromosoma 13 se determinó usando el grupo de cromosomas 3, 4, 5 y 6como el cromosoma normalizador; la dosis de cromosoma 18 se determinó usando el cromosoma 8 como el cromosoma normalizador; y la dosis de cromosoma X se determinó usando el cromosoma 4 como el cromosoma normalizador. Los límites se calcularon para ser 2 desviaciones estándar por encima y por debajo de la media determinada en las muestras calificadas.

La Tabla 12 muestra los datos para la determinación de la fracción fetal en muestras ejemplares. Los valores de dosis de cromosoma calculados para los cromosomas 21, 18, 13, X e Y en las muestras de ensayo ejemplares correspondientes se dan en las Tablas 15, 16, 17 y 18, respectivamente.

Trisomía 21

La Tabla 8 proporciona la dosis calculada para el cromosoma 21 en la muestra de ensayo (11409). El cromosoma 14 se usó como el cromosoma normalizador. El límite calculado para el diagnóstico positivo de la aneuploidía T21 es estableció en 2 desviaciones estándar de la media de las muestras calificadas (normales). Se dio un diagnóstico para la T21 en base a que la dosis de cromosoma en la muestra de ensayo era mayor que la del límite establecido. Las doce muestras de T21 que se conformaron para ser T21 por el cariotipo se identificaron en una población de 48 muestras de sangre.

TABLA 15

Dosis de Cromosoma para una aneuploidía T21			
Cromosoma	**Densidad de Etiqueta de Secuencia**	**Dosis de Cromosoma para Chr 21**	**Límite**
Chr21	264,404	0.439498	0.410634
Chr14	601,605		

Trisomía 18

La Tabla 9 proporciona la dosis calculada para el cromosoma 18 en una muestra de ensayo (95133). El cromosoma 8 se usó como el cromosoma normalizador. En esta situación, el cromosoma 8 tenía la variabilidad más baja y la diferenciabilidad más alta. El límite calculado para el diagnóstico positivo de la aneuploidía T18 se estableció en >2 desviaciones estándar de la media de las muestras calificadas (no T18). Se dio un diagnóstico para la T18 en base a que la dosis de cromosoma en la muestra de ensayo era mayor que el límite establecido. Se identificaron ocho muestras de T18 usando dosis de cromosomas, y se confirmaron que eran T18 por cariotipado

TABLA 16

Dosis de Cromosoma para una aneuploidía T18			
Cromosoma	**Densidad de Etiqueta de Secuencia**	**Dosis de Cromosoma para Chr 18**	**Límite**
Chr18	604,291	0.550731	0.533297
Chr8	1,097,253		

Trisomía 13

Las Tablas 10 y 11 proporcionan la dosis calculada para el cromosoma 13 en una muestra de ensayo (51236). El límite calculado para el diagnóstico positivo de la aneuploidía T13 se estableció en 2 desviaciones estándar de la media de las muestras calificadas (no T13). La dosis de cromosoma para el cromosoma 13 proporcionada en la Tabla 10 se calculó usando la densidad de etiqueta de secuencia para el cromosoma 4 como el cromosoma normalizador, mientras que la dosis dada en la Tabla 11 se determinó usando la media de las proporciones de densidades de etiquetas de secuencia para el grupo de cromosomas 3, 4, 5 y 6 como el cromosoma normalizador. Se dio un diagnóstico para T13 en base a que la dosis de cromosoma en la muestra de ensayo era mayor que el límite establecido. Una muestra de T13 se identificó usando dosis de cromosomas, y se confirmaron que eran T13 por cariotipado.

Los datos muestran que la combinación de cromosomas 3, 4, 5 y 6 proporcionan una variabilidad (1,06) que es similar a la del cromosoma 4 (1,01), demostrando que un grupo de cromosomas se puede usar como el cromosoma normalizador para determinar las dosis de cromosomas e identificar aneuploidías.

TABLA 17

Dosis de Cromosoma para una Aneuploidía T13			
Cromosoma	**Densidad de Etiqueta de Secuencia**	**Dosis de Cromosoma para Chr 13**	**Límite**
Chr13	669,872	0.538140	0.536044
Chr4	1,244,791		

TABLA 18

Dosis de Cromosoma para una Aneuploidía T13			
Cromosoma	**Densidad de Etiqueta de Secuencia**	**Dosis de Cromosoma para Chr 13**	**Límite**
Chr13	669,872	0.532674	0.515706
Chr3	1,385,881		
Chr4	1,244,791		
Chr5	1,229,257		
Chr6	1,170,331		

Síndrome de Turner (monosomía X)

Se identificaron tres muestras que tenían una dosis de cromosoma menor que la del límite establecido que tenían menos de un cromosoma X. Se determinó que las mismas muestras tenían una dosis de cromosoma Y que era menor que el límite establecido, indicando que las muestras no tenían cromosoma Y.

Las dosis calculadas para los cromosomas X e Y en la muestra de ensayo de monosomía Y ejemplar (54430) se dan en la Tabla 12. El Cromosoma 4 se seleccionó como el cromosoma normalizador para calcular la dosis para el cromosoma X; y todos los cromosomas, es decir 1-22 e Y, se usaron como los cromosomas normalizadores. El límite calculado para el diagnóstico positivo del Síndrome de Turner (monosomía X) se estableció para el cromosoma X en <-2 desviaciones estándar de la medio y para la ausencia del cromosoma Y en <-2 desviaciones estándar de la media para las muestras calificadas (no monosomía Y)

TABLA 19

Dosis de Cromosoma para un Síndrome de Turner (monosomía X)			
Cromosoma	**Densidad de Etiqueta de Secuencia**	**Dosis de Cromosoma para Chr X**	**Límite**
ChrX	904,049	0.777990	0.797603
Chr4	1,162,031		
ChrY	390	0.0004462	0.002737754
Chr(1-22, X) (Average)	874,108.1		

Por lo tanto, el método permite la determinación simultánea de aneuploidías cromosómicas y la fracción fetal por secuenciación masivamente paralela de una muestra materna que comprende una mezcla de ADN libre de células fetal y materno que ha sido enriquecido para una pluralidad de secuencias polimórficas comprendiendo cada una un SNP. En este ejemplo, la mezcla de ácidos nucleicos fetales y maternos se enriqueció combinando una porción de una biblioteca de secuenciación que se construyó de secuencias polimórficas fetales y maternas con una biblioteca de secuenciación que se construyó de la mezcla de ADN libre de células feral y materno original no amplificado.

Ejemplo 9

Determinación Simultánea de Aneuploidía y Fracción fetal:

Enriquecimiento de ácidos nucleicos fetales y maternos en una muestra de ADN libre de células purificada

Para enriquecer el ADN libre de células fetal y materno contenido en una muestra purificada de ADN libre de células extraído de una muestra de plasma materno, se usó una porción del ADN libre de células purificado para amplificar secuencias de ácidos nucleicos objetivo polimórficos que comprenden cada una un SNP elegido del panel de SNPs dado en la tabla 6.

El plasma libre de células se obtuvo de una muestra de sangre materna, y el ADN libre de células se purificó de la muestra de plasma como se describe en el Ejemplo 1. La concentración final se determinó que era de 92,8 pg/µl.

El ADN libre de células contenido en 5 µl de ADN libre de células purificado se amplificó en un volumen de reacción de 50 µl que contenía 7,5µl de una mezcla de cebadores de 1 uM (Tabla 5), 10µl de Mezcla maestra NEB 5X y 27 µl de agua. El ciclado térmico se realizó con el Gene Amp9700 (Applied Biosystems). Usando las siguientes condiciones de ciclado: incubar a 95º C durante 1 minuto, seguido por 30 ciclos a 95º C durante 20 segundos, 68º C durante 1 minuto y 68º C durante 30 segundos, que fue seguido por una incubación final a 68º C durante 5 minutos. Se añadió un mantenimiento final a 4º C hasta que las muestras fueron retiradas para combinarlas con la porción no amplificada de la muestra de ADN libre de células purificada. El producto amplificado se purificó usando el sistema de purificación por PCR Agencourt AMPure XP (Parte No. A63881; Beckman Coulter Genomics, Danvers, MA), y se cuantificó la concentración usando el Nanodrop 2000 (Thermo Scientific, Wilmington, DE). El producto de la amplificación purificado se diluyó al 1:10 en agua y se añadieron 0,9µl (371 pg) a 40 µl de muestra de ADN libre de células purificado para obtener una espiga del 10%. El ADN libre de células fetal y materno enriquecido presente en la muestra de ADN libre de células purificado se usó para preparar una biblioteca de secuenciación, y se secuenció como se describe en el Ejemplo 2.

La Tabla 13 proporciona los recuentos de etiquetas obtenidos para cada uno de los cromosomas 21, 18, 13, X e Y, es decir la densidad de etiqueta de secuencia, y los recuentos de etiquetas obtenidos para las secuencias polimórficas informativas contenidas en el genoma de referencia de SNP, es decir densidad de etiqueta de SNP. Los datos muestran que la información de la secuenciación se puede obtener secuenciando una única biblioteca construida de una muestra de ADN libre de células materna purificada que ha sido enriquecida para secuencias que comprenden SNPs para determinar simultáneamente la presencia o ausencia de aneuploidía y la fracción fetal. En el ejemplo dado los datos muestran que la fracción fetal de ADN en la muestra de plasma AFR105 era cuantificable que los resultados de secuenciación de cinco SNPs informativos y se determinó que era del 3,84%. Las densidades de etiquetas de secuencia se proporcionan para los cromosomas 21, 13, 18, X e Y. La muestra AFR105 fue la única muestra que se sometió al protocolo de enriquecer ADN libre de células purificado para secuencias polimórficas amplificadas. Por lo tanto, no se proporcionaron los coeficientes de variación y ensayos para la diferenciabilidad. Sin embargo, el ejemplo muestra que el protocolo de enriquecimiento proporciona los recuentos de etiquetas requeridos para determinar la aneuploidía y la fracción fetal a partir de un único proceso de secuenciación.

TABLA 20

Determinación Simultánea de Aneuploidía y Fracción Fetal: Enriquecimiento de ácidos nucleicos fetales y maternos en una muestra de ADN libre de células purificada					
Aneuploidía					
	Cromosoma 21	Cromosoma 18	Cromosoma 13	Cromosoma X	Cromosoma Y
Densidad de Etiqueta de Secuencia	178763	359529	388204	572330	2219
Cariotipo	No afectada	No afectada	No afectada	No afectada	No afectada
Fracción Fetal					
SNP		**DENSIDAD DE ETIQUETA DE SNP**		**FRACCION FETAL (%)**	
rs10773760.1|Chr.12|longitud=128|alelo=A		18903		2.81	
rs10773760.2|Chr.12|longitud=128|alelo=G		532			
rs1109037. 1|Chr.2|longitud=126|alelo=A		347		5.43	
rs1109037.2|Chr.2|longitud=126|alelo=G		6394			
rs2567608.1|Chr.20|longitud=10|alelo=A		94503		1.74	
rs2567608.2|Chr.20|longitud=110|alelo=G		1649			
rs7041158.1|Chr.9|longitud= 17|alelo=C		107		5.61	
rs7041158.2|Chr.9|longitud=117|alelo=T		6			
rs8078417.1|Chr.17|longitud=110|alelo=C		162668		3.61	
rs8078417.2|Chr.17|longitud=110|alelo=T		5877			
Fracción Fetal (Media±D.S.) = 3.8±1.6					

Ejemplo 10

Determinación Simultánea de Aneuploidía y Fracción Fetal: Enriquecimiento de ácidos nucleicos fetales y maternos en una muestra de sangre

Para enriquecer el ADN libre de células feral y materno contenido en una muestra de plasma original derivada de una mujer embarazada, se usó una porción de la muestra de plasma original para amplificar secuencias de ácidos nucleicos objetivo polimórficas que comprendía cada una un SNP del panel de SNPs proporcionado en la Tabla 14, y una porción del producto amplificado se combinó con el resto de la muestra de plasma original.

El ADN libre de células contenido en 15µl de plasma libre de células se amplificó en un volumen de reacción de 50µl que contenía 9µl de una mezcla de cebadores de 1µm (15 plex Tabla 5), 1µl de ADN polimerasa de sangre Phusion, 25µl del tampón de PCR de sangre Phusion que contenía desoxinucleótidos trifosfatos (dNTPs: dATP, dCTP, dGTP y dTTP). El ciclado térmico se realizó con el Gene Amp9700 (Applied Biosystems) usando las siguientes condiciones de ciclado: incubar a 95º C durante 3 minutos, seguido por 35 ciclos a 95º C durante 20 segundos, 55º C durante 30 segundos, y 70º C durante 1 minuto, a lo que siguió una incubación final a 68º C durante 5 minutos. Se añadió un mantenimiento final a 4º C hasta que las muestras se retiraron para la combinación con la porción no amplificada del plasma libre de células. El producto amplificado se diluyó a 1:2 con agua y se analizó usando el Bioanalizador. Se diluyeron 3 µl adicionales del producto amplificado con 11,85µl de agua para obtener una concentración final de 2 ng/µl. Se combinaron 2,2µl del producto amplificado diluido con el resto de la muestra de plasma. El ADN libre de células fetal y materno enriquecido presente en la muestra de plasma se purificó como se describe en el Ejemplo 1, y se usó para preparar una biblioteca de secuenciación. La secuenciación y el análisis de los datos de secuenciación se realizaron como se describe en los Ejemplos 2 y 3.

Los resultados se dan en la Tabla 21. En el ejemplo dado, los datos muestran que la fracción de ADN fetal en la muestra de plasma SAC2517 era cuantificable de los resultados de secuenciación de un SNP informativo y se determinó que era del 9,5%. En el ejemplo dado, se mostró por cariotipado que la muestra SAC2517 no se vio afectada por aneuploidías de los cromosomas 21, 13, 18, X e Y. Las densidades de etiquetas de secuencia se proporcionan para los cromosomas 21, 13, 18, X e Y. La muestra SAC2517 fue la única muestra que se sometió al protocolo de enriquecimiento de ADN libre de células de plasma para secuencias polimórficas amplificadas. Por lo tanto, los coeficientes de variación y ensayos de diferenciabilidad no se pudieron determinar. El ejemplo demuestra que enriquecer la muestra de ADN libre de células fetal y materno presente en una muestra de plasma para secuencias de ácidos nucleicos que comprenden al menos un SNP informativo se puede usar para proporcionar la secuencia requisito y los recuentos de etiquetas de SNP para determinar la aneuploidía y la fracción fetal a partir de un único proceso de secuenciación

TABLA 21

Determinación Simultánea de Aneuploidía y Fracción Fetal: Enriquecimiento de ácidos nucleicos fetales y maternos en una muestra de plasma					
Aneuploidía					
	Cromosoma 21	Cromosoma 18	Cromosoma 13	Cromosoma X	Cromosoma Y
Densidad de Etiqueta de Secuencia	183851	400582	470526	714055	2449
Cariotipo	No afectada	No afectada	No afectada	No afectada	No afectada
Fracción Fetal					
SNP		**RECUENTOS DE ETIQUETAS**		**FRACCION FETAL (%)**	
rs10773760.1|Chr.12|longitud=128|alelo=A		8536		9.49	
rs10773760.2|Chr.128|longitud=128|alelo=G		89924			

Ejemplo 11

Determinación Simultánea de Aneuploidía y Fracción Fetal en muestras maternas enriquecidas para secuencias polimórficas que comprenden STRs

Para determinar simultáneamente la presencia o ausencia de una aneuploidía y la fracción fetal en una mezcla de ADN libre de células fetal y materno obtenido de una muestra materna, la mezcla se enriquece para secuencias polimórficas que comprenden STRs, se secuencia y se analizan los datos. El enriquecimiento puede ser de una biblioteca de secuenciación como se describe en el Ejemplo 8, de una muestra de ADN libre de células purificado como se describe en el ejemplo 9 o de una muestra de plasma como se describe en el ejemplo10. En cada caso, la información de secuenciación se obtiene de secuenciar una biblioteca individual, lo que permite determinar simultáneamente la presencia o ausencia de una aneuploidía y la fracción fetal. Preferiblemente, la biblioteca de secuenciación se prepara usando el protocolo abreviado proporcionado en el Ejemplo 2.

Las STRs que se amplifican se eligen de las STRs codis y no codis divulgadas en la Tabla 22, y la amplificación de las secuencias de STRs polimórficas se obtiene usando los conjuntos correspondientes de cebadores proporcionados. Algunas de las STRs que se han divulgado y/o analizado anteriormente para determinar la fracción fetal se enumeran en la Tabla 22, y se divulgan en las Solicitudes provisionales US 61/296,358 y 61/360,837.

TABLA 22

miniSTRs CODIS y NO-CODIS				
Locus de STR Locus (Nombre del Marcador)	**Localización del Cromosoma**	**Intervalo de Tamaño**	**Acceso del GenBank**	**Secuencias de Cebadores (Directo/Inversoe)**
loci de miniSTR Codis*				
CSF1PO	5q33.1	89-129	X14720	ACAGTAACTGCCTTCATAGATAG (CSF1PO_F; SEQ ID NO:113) GTGTCAGACCCTGTTCTAAGTA (CSF1PO_R; SEQ ID NO:114)
FGA	4q31.3	125-281	M64982	AAATAAAATTAGGCATATTTACAAGC (FGA_F; SEQ ID NO: 115) GCTGAGTGATTTGTCTGTAATTG(FGA_R; SEQ ID NO:116)
TH01	11p15.5	51-98	D00269	CCTGTTCCTCCCTTATTTCCC(TH01_F; SEQ ID NO:117) GGGAACACAGACTCCATGGTG(THO 1_R; SEQ ID NO:118)
TPOX	2p25.3	65-101	M68651	CTTAGGGAACCCTCACTGAATG(TPOX_F ; SEQ ID NO:119) GTCCTTGTCAGCGTTTATTTGC(TPOX_R; SEQ ID NO:120)
vWA	12p13.31	88-148	M25858	AATAATCAGTATGTGACTTGGATTGA(v WA_F; SEQ ID NO:121) ATAGGATGGATGGATAGATGGA(vWA_R ; SEQ ID NO:122)
D3S1358	3p21.31	72-120	NT_005997	CAGAGCAAGACCCTGTCTCAT(D3S1358 _F; SEQ ID NO:123) TCAACAGAGGCTTGCATGTAT(D3 S1358_ R; SEQ ID NO:124)
D5S818	5q23.2	81-117	AC008512	GGGTGATTTTCCTCTTTGGT(D5S818_F; SEQ ID NO:125) AACATTTGTATCTTTATCTGTATCCTTAT TTAT(DSS818_R; SEQ ID NO:126)
D7S820	7q21.11	136-176	AC004848	GAACACTTGTCATAGTTTAGAACGAAC(D7S820_F; SEQ ID NO:127) TCATTGACAGAATTGCACCA(D7S820_R; SEQ ID NO:128)
D8S1179	8q24.13	86-134	AF216671	TTTGTATTTCATGTGTACATTCGTATC(D 7S820_F; SEQ ID NO:129) ACCTATCCTGTAGATTATTTTCACTGTG(D7S820_R; SEQ ID NO:130)
D13S317	13q31.1	88-132	AL353628	TCTGACCCATCTAACGCCTA(D13S317_F; SEQ ID NO:131) CAGACAGAAAGATAGATAGATGATTGA(D13S317_R; SEQ ID NO:132)
D16S539	16q24.1	81-121	AC024591	ATACAGACAGACAGACAGGTG(D165539 _F; SEQ ID NO:133) GCATGTATCTATCATCCATCTCT(D16553 9_R; SEQ ID NO: 134)
D18S51	18q21.33	113-193	AP001534	TGAGTGACAAATTGAGACCTT(D18S51_F ; SEQ ID NO:135) GTCTTACAATAACAGTTGCTACTATT(D1 8S51_R; SEQ ID NO:136)
D21S11	21q21.1	153-221	AP000433	ATTCCCCAAGTGAATTGC(D21S11_F; SEQ ID NO:137)

miniSTRs CODIS y NO-CODIS				
Locus de STR Locus (Nombre del Marcador)	**Localización del Cromosoma**	**Intervalo de Tamaño**	**Acceso del GenBank**	**Secuencias de Cebadores (Directo/Inversoe)**
loci de miniSTR Codis*				
				GGTAGATAGACTGGATAGATAGACGA(D 21511_R; SEQ ID NO:138)
D2S1338	2q35	90-142	AC01036	TGGAAACAGAAATGGCTTGG(D2S1338_F ; SEQ ID NO:139) GATTGCAGGAGGGAAGGAAG(D2 S13 38_ R; SEQ ID NO:140)
Penta D	21q22.3	94-167	AP001752	GAGCAAGACACCATCTCAAGAA(Penta D_F; SEQ ID NO:141) GAAATTTTACATTTATGTTTATGATTCTC T(Penta D_R; SEQ ID NO:142)
Penta E	15q26.2	80-175	AC027004	GGCGACTGAGCAAGACTC(Penta E _F; SEQ ID NO:143) GGTTATTAATTGAGAAAACTCCTTACA(P enta E_R; SEQ ID NO:144)
loci de miniSTR No Codis*				
D22S1045	22q12.3	82 - 115	AL022314 (17)	ATTTTCCCCGATGATAGTAGTCT(D22S1045_ F; SEQ ID NO:145) GCGAATGTATGATTGGCAATATTTTTT(D22S 1045_R; SEQ ID NO:146)
D20S1082	20q13.2	73 - 101	AL158015	ACATGTATCCCAGAACTTAAAGTAAAC(D2 0S1082_F;SEQIDNO:147) GCAGAAGGGAAAATTGAAGCTG(D20S 1082 _R; SEQ ID NO:148)
D20S482	20p13	85 - 126	AL121781 (14)	CAGAGACACCGAACCAATAAGA(D20S482_ F; SEQ ID NO:149) GCCACATGAATCAATTCCTATAATAAA(D20S482_R; SEQ ID NO:150)
D18S853	18p11.31	82 - 104	AP005130 (11)	GCACATGTACCCTAAAACTTAAAAT(D18S8 53_F;SEQIDNO:151) GTCAACCAAAACTCAACAAGTAGTAA(D18 S853_R; SEQ ID NO:152)
D17S1301	17q25.1	114 - 139	AC016888 (12)	AAGATGAAATTGCCATGTAAAAATA(D17S1 301_F; SEQ ID NO:153) GTGTGTATAACAAAAATTCCTATGATGG(D17 S1301_R;SEQIDNO:154)
D17S974	17p13.1	114 - 139	AC034303 (10)	GCACCCAAAACTGAATGTCATA(D175974_F ; SEQ ID NO:155) GGTGAGAGTGAGACCCTGTC(D175974_R; SEQ ID NO:156)
D14S1434	14q32.13	70-98	AL121612 (13)	TGTAATAACTCTACGACTGTCTGTCTG(D14 S1434_F; SEQ ID NO:157) GAATAGGAGGTGGATGGATGG(D 14S1434_R ; SEQ ID NO:158)
D12ATA63	12q23.3	76-106	AC009771 (13)	GAGCGAGACCCTGTCTCAAG(D12ATA63_F; SEQ ID NO:159) GGAAAAGACATAGGATAGCAATTT(D 12AT A63_R; SEQ ID NO: 160)
D11S4463	11q25	88 - 116	AP002806 (14)	TCTGGATTGATCTGTCTGTCC(D1154463_F; SEQ ID NO:161)

loci de miniSTR No Codis*				GAATTAAATACCATCTGAGCACTGAA(D 11S 4463_R; SEQ ID NO:162)
D10S1435	10p15.3	82 - 139	AL354747 (11)	TGTTATAATGCATTGAGTTTTATTCTG(D10S 1435_F; SEQ ID NO: 163) GCCTGTCTCAAAAATAAAGAGATAGACA(D 10S1435_R; SEQ ID NO:164)
D10S1248	10q26.3	79 - 123	AL391869 (13)	TTAATGAATTGAACAAATGAGTGAG(D10S1 248 _F; SEQ ID NO:165) GCAACTCTGGTTGTATTGTCTTCAT(D 1 0S12 48_R; SEQ ID NO:166)
D9S2157	9q34.2	71-107	AL162417 (10)	CAAAGCGAGACTCTGTCTCAA(D9S2157_F; SEQ ID NO:167) GAAAATGCTATCCTCTTTGGTATAAAT(D9S 2157_R; SEQ ID NO:168)
D9S1122	9q21.2	93 - 125	AL161789 (12)	GGGTATTTCAAGATAACTGTAGATAGG(D9S1122_F; SEQIDNO:168) GCTTCTGAAAGCTTCTAGTTTACC(D9S 1122 _R;SEQ ID NO:170)
D8S1115	8p11.21	63 - 96	AC090739 (9)	TCCACATCCTCACCAACAC(D8S1115_F; SEQ ID NO:171) GCCTAGGAAGGCTACTGTCAA(D8S 1115_R; SEQ ID NO:172)
D6S1017	6p21.1	81 - 110	AL035588 (10)	CCACCCGTCCATTTAGGC(D6S1017_F; SEQ ID NO:173) GTGAAAAAGTAGATATAATGGTTGGTG(D6 S1017_R; SEQ ID NO:174)
D6S474	6q21	107 - 136	AL357514 (17)	GGTTTTCCAAGAGATAGACCAATTA(D6S47 4_F; SEQ ID NO: 175) GTCCTCTCATAAATCCCTACTCATATC(D6S 474_R; SEQ ID NO: 176)
D5S2500	5q11.2	85 - 126	AC008791 (17)	CTGTTGGTACATAATAGGTAGGTAGGT(D5S 2500_F; SEQ ID NO:177) GTCGTGGGCCCCATAAATC(D5S2500_R; SEQ ID NO:178)
D4S2408	4p15.1	85 - 109	AC110763 (9)	AAGGTACATAACAGTTCAATAGAAAGC(D4 S2408_F; SEQ ID NO:179) GTGAAATGACTGAAAAAATAGTAACCA(D4S 2408_R; SEQ ID NO:180)
D4S2364	4q22.3	67-83	AC022317 (9)	CTAGGAGATCATGTGGGTATGATT(D4S2364 U_F; SEQ ID NO:181) GCAGTGAATAAATGAACGAATGGA(D4S236 4_R; SEQ ID NO: 182)
D3S4529	3p12.1	111 - 139	AC117452 (13)	CCCAAAATTACTTGAGCCAAT(D3S452_F; SEQ ID NO:183) GAGACAAAATGAAGAAACAGACAG(D3S45 2_R; SEQ ID NO:184)
D3S3053	3q26.31	84 - 108	AC069259 (9)	TCTTTGCTCTCATGAATAGATCAGT(D3S305 3_F; SEQ ID NO:185) GTTTGTGATAATGAACCCACTCAG(D3S3053 _R; SEQ ID NO:186)
D2S1776	2q24.3	127 - 161	AC009475 (11)	TGAACACAGATGTTAAGTGTGTATATG(D2S 1776_F; SEQ ID NO: 187) GTCTGAGGTGGACAGTTATGAAA(D2S 1776_ R; SEQ ID NO:188)
D2S441	2p14	78-110	AC079112 (12)	CTGTGGCTCATCTATGAAAACTT(D2S441_F; SEQ ID NO:189)

loci de miniSTR No Codis*					
					GAAGTGGCTGTGGTGTTATGAT(D2S441_R; SEQ ID NO:190)
D1S1677	1q23.3	81 - 117	AL513307 (15)		TTCTGTTGGTATAGAGCAGTGTTT(D 1 S 1677 _F; SEQ ID NO:191) GTGACAGGAAGGACGGAATG(D 1S1677_R; SEQ ID NO:192)
D1S1627	1p21.1	81 - 100	AC093119 (13)		CATGAGGTTTGCAAATACTATCTTAAC(D1S 1627_F; SEQ ID NO:193) GTTTTAATTTTCTCCAAATCTCCA(D1S1627_ R; SEQ ID NO: 194)
D1GATA113	1p36.23	81 - 105	Z97987(11)		TCTTAGCCTAGATAGATACTTGCTTCC(D1G ATA113_F; SEQ ID NO:195) GTCAACCTTTGAGGCTATAGGAA(D1GATA 113_R; SEQ ID NO:196)

*(Butler *et coil,* J Forensic Sci 5:1054-1064; Hill *et al.,* Poster #44- 17th International Symposium on Human Identification - 2006)

Las miniSTRs proporcionadas en la Tabla 22 se han usado con éxito para determinar la fracción fetal en las muestras de ADN libre de células de plasma obtenidas de mujeres embarazadas con fetos o masculinos o femeninos, usando electroforesis capilar (ver Tabla 24 en el Ejemplo 15) para identificar y cuantificar los alelos fetales y maternos. Por lo tanto, se espera que las secuencias polimórficas que comprenden otras STRs, por ejemplo las STRs restantes de la Tabla 22, se puedan usar para determinar la fracción fetal por métodos de secuenciación masivamente paralelos.

La secuenciación de la biblioteca enriquecida para secuencias de STR polimórficas se realiza usando una tecnología NGS, por ejemplo secuenciación masivamente paralela por síntesis. Las lecturas de secuencia de longitudes de al menos 100 bp se alinean con un genoma de referencia, por ejemplo la secuencia del genoma de referencia humano NCBI36/hg18, y con un genoma STR, y el número de etiquetas de secuencia mapeadas para el genoma humano de referencia y el genoma de referencia STR obtenido para los alelos informativos se usa para determinar la presencia o ausencia de aneuploidía y la fracción fetal, respectivamente. El genoma de referencia STR incluye las secuencias de amplicones amplificados de los cebadores dados.

Ejemplo 12

Determinación Simultánea de Aneuploidía y Fracción Fetal por Secuenciación Masivamente Paralela de Muestras Maternas Enriquecidas para Secuencias Polimórficas que Comprenden SNPs en tándem

Para determinar simultáneamente aneuploidía y fracción fetal en muestras maternas que comprenden ácidos nucleicos fetales y maternos, se enriquecen muestras de ADN libre de células purificado y muestras de bibliotecas de secuenciación para secuencias de ácidos nucleicos objetivo polimórficos que comprenden cada una un par de SNPs en tándem seleccionados de rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980- rs987981; rs4143392- rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs1 1088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 -rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; y rs2837381-rs4816672. Los cebadores usados para amplificar las secuencias objetivo que comprenden los SNPs en tándem están diseñados para abarcar ambos sitios de SNP. Por ejemplo, el cebador directo está diseñado para abarcar el primer SNP, y el cebador inverso está diseñado para abarcar el segundo del par de SNP en tándem, es decir cada uno de los sitios SNP en el par en tándem está abarcado dentro de los 36 bp generados por el método de secuenciación. La secuenciación de extremos emparejados se puede usar para identificar todas las secuencias que abarcan los sitios de SNP en tándem. Los conjuntos ejemplares de cebadores que se usan para amplificar los SNPs en tándem divulgados en la presente son rs7277033-rs2110153_F: TCCTGGAAACAAAAGTATT (SEQ ID NO:197) y rs7277033-rs2110153_R: AACCTTACAACAAAGCTAGAA (SEQ ID NO:198), set rs2822654-rs1882882_F: ACTAAGCCTTGGGGATCCAG (SEQ ID NO:199) y rs2822654-rs1882882_R: TGCTGTGGAAATACTAAAAGG (SEQ ID NO:200), set rs368657-rs376635 F:CTCCAGAGGTAATCCTGTGA (SEQ ID NO:201) y rs368657-rs376635_R:TGGTGTGAGATGGTATCTAGG (SEQ ID NO:202), rs2822731-rs2822732_F:GTATAATCCATGAATCTTGTTT (SEQ ID NO:203) y rs2822731-rs2822732_R:TTCAAATTGTATATAAGAGAGT (SEQ ID NO:204), rs1475881-rs7275487_F:GCAGGAAAGTTATTTTTAAT (SEQ ID NO:205) y rs1475881-rs7275487_R:TGCTTGAGAAAGCTAACACTT (SEQ ID NO:206), rs1735976-rs2827016F:CAGTGTTTGGAAATTGTCTG (SEQ ID NO:207) y rs1735976-rs2827016_R:GGCACTGGGAGATTATTGTA (SEQ ID NO:208), rs447349-rs2824097_F:TCCTGTTGTTAAGTACACAT (SEQ ID NO:209) y rs447349-rs2824097_R:GGGCCGTAATTACTTTTG (SEQ ID NO:210), rs418989-rs13047336_F:ACTCAGTAGGCACTTTGTGTC (SEQ ID NO:211) y rs418989-rs13047336_R:TCTTCCACCACACCAATC (SEQ ID NO:212), rs987980-rs987981_F:TGGCTTTTCAAAGGTAAAA (SEQ ID NO:213) y rs987980- rs987981_R: GCAACGTTAACATCTGAATTT (SEQ ID NO:214), rs4143392-rs4143391_F: rs4143392- rs4143391 (SEQ ID NO:215) y rs4143392- rs4143391_R:ATTTTATATGTCATGATCTAAG (SEQ ID NO:216), rs1691324-rs13050434_F: AGAGATTACAGGTGTGAGC (SEQ ID NO:217) y rs1691324-rs13050434_R: ATGATCCTCAACTGCCTCT (SEQ ID NO:218), rs11909758-rs9980111_F: TGAAACTCAAAAGAGAAAAG (SEQ ID NO:219) y rs11909758-rs9980111_R: ACAGATTTCTACTTAAAATT (SEQ ID NO:220), rs2826842-rs232414_F: TGAAACTCAAAAGAGAAAAG (SEQ ID NO:221) y rs2826842-rs232414_R: ACAGATTTCTACTTAAAATT (SEQ ID NO:22), rs2826842-rs232414_F: GCAAAGGGGTACTCTATGTA (SEQ ID NO:223) y rs2826842-rs232414_R: TATCGGGTCATCTTGTTAAA (SEQ ID NO:224), rs1980969-rs1980970_F: TCTAACAAAGCTCTGTCCAAAA (SEQ ID NO:225) y rs1980969-rs1980970-R: CCACACTGAATAACTGGAACA (SEQ ID NO:226), rs9978999-rs9979175_F: GCAAGCAAGCTCTCTACCTTC (SEQ ID NO:227) y rs9978999-

rs9979175_R: TGTTCTTCCAAAATTCACATGC (SEQ ID NO:228), rs1034346-rs12481852_F: ATTTCACTATTCCTTCATTTT (SEQ ID NO:229) y rs1034346-rs12481852_R: TAATTGTTGCACACTAAATTAC (SEQ ID NO:230), rs4817013-rs7277036_F: AAAAAGCCACAGAAATCAGTC (SEQ ID NO:231) y rs4817013-rs7277036_R: TTCTTATATCTCACTGGGCATT (SEQ ID NO:232), rs9981121-rs2829696_F: GGATGGTAGAAGAGAAGAAAGG (SEQ ID NO:233) y rs9981121-rs2829696_R: GGATGGTAGAAGAGAAGAAAGG (SEQ ID NO:234), rs455921-rs2898102_F: TGCAAAGATGCAGAACCAAC (SEQ ID NO:235) y rs455921-rs2898102_R: TTTTGTTCCTTGTCCTGGCTGA (SEQ ID NO:236), rs2898102- rs458848_F: TGCAAAGATGCAGAACCAAC (SEQ ID NO:237) y rs2898102- rs458848_R: GCCTCCAGCTCTATCCAAGTT (SEQ ID NO:238), rs961301-rs2830208_F: CCTTAATATCTTCCCATGTCCA (SEQ ID NO:239) y rs961301-rs2830208_R: ATTGTTAGTGCCTCTTCTGCTT (SEQ ID NO:240), rs2174536-rs458076_F: GAGAAGTGAGGTCAGCAGCT (SEQ ID NO:241) y rs2174536-rs458076_R: TTTCTAAATTTCCATTGAACAG (SEQ ID NO:242), rs11088023-rs11088024_F: GAAATTGGCAATCTGATTCT (SEQ ID NO:243) y rs11088023-rs11088024_R: CAACTTGTCCTTTATTGATGT (SEQ ID NO:244), rs1011734-rs1011733_F: CTATGTTGATAAAACATTGAAA (SEQ ID NO:245) y rs1011734-rs1011733_R: GCCTGTCTGGAATATAGTTT (SEQ ID NO:246), rs2831244-rs9789838_F: CAGGGCATATAATCTAAGCTGT (SEQ ID NO:247) y rs2831244-rs9789838_R: CAATGACTCTGAGTTGAGCAC (SEQ ID NO:248), rs8132769-rs2831440_F: ACTCTCTCCCTCCCCTCT (SEQ ID NO:249) y rs8132769-rs2831440_R: TATGGCCCCAAAACTATTCT (SEQ ID NO:250), rs8134080-rs2831524_F: ACAAGTACTGGGCAGATTGA (SEQ ID NO:251) y rs8134080-rs2831524_R: GCCAGGTTTAGCTTTCAAGT (SEQ ID NO:252), rs4817219-rs4817220_F: TTTTATATCAGGAGAAACACTG (SEQ ID NO:253) y rs4817219-rs4817220_R: CCAGAATTTTGGAGGTTTAAT (SEQ ID NO:254), rs2250911-rs2250997_F: TGTCATTCCTCCTTTATCTCCA (SEQ ID NO:255) y rs2250911-rs2250997_R: TTCTTTTGCCTCTCCCAAAG (SEQ ID NO:256), rs2831899-rs2831900_F: ACCCTGGCACAGTGTTGACT (SEQ ID NO:257) y rs2831899-rs2831900_R: TGGGCCTGAGTTGAGAAGAT (SEQ ID NO:258), rs2831902-rs2831903_F: AATTTGTAAGTATGTGCAACG (SEQ ID NO:259) y rs2831902-rs2831903_R: TTTTTCCCATTTCCAACTCT (SEQ ID NO:260), rs11088086-rs2251447_F: AAAAGATGAGACAGGCAGGT (SEQ ID NO:261) y rs11088086-rs2251447 _R: ACCCCTGTGAATCTCAAAAT (SEQ ID NO:262), rs2832040-rs11088088_F: GCACTTGCTTCTATTGTTTGT (SEQ ID NO:263) y rs2832040-rs11088088_R: CCCTTCCTCTCTTCCATTCT (SEQ ID NO:264), rs2832141-rs2246777_F: AGCACTGCAGGTA (SEQ ID NO:265) y rs2832141-rs2246777_R: ACAGATACCAAAGAACTGCAA (SEQ ID NO:266), rs2832959 _ rs9980934_F: TGGACACCTTTCAACTTAGA (SEQ ID NO:267) y rs2832959 -rs9980934_R: GAACAGTAATGTTGAACTTTTT (SEQ ID NO:268), rs2833734-rs2833735_F: TCTTGCAAAAAGCTTAGCACA (SEQ ID NO:269) y rs2833734-rs2833735_R: AAAAAGATCTCAAAGGGTCCA (SEQ ID NO:270), rs933121-rs933122_F: GCTTTTGCTGAACATCAAGT (SEQ ID NO:271) y rs933121-rs933122_R: CCTTCCAGCAGCATAGTCT (SEQ ID NO:272), rs2834140-rs12626953_F: AAATCCAGGATGTGCAGT (SEQ ID NO:273) y rs2834140-rs12626953_R: ATGATGAGGTCAGTGGTGT (SEQ ID NO:274), rs2834485-rs3453_F: CATCACAGATCATAGTAAATGG (SEQ ID NO:275) y rs2834485-rs3453_R: AATTATTATTTTGCAGGCAAT (SEQ ID NO:276), rs9974986-rs2834703_F: CATGAGGCAAACACCTTTCC (SEQ ID NO:277) y rs9974986-rs2834703_R: GCTGGACTCAGGATAAAGAACA (SEQ ID NO:278), rs2776266-rs2835001_F: TGGAAGCCTGAGCTGACTAA (SEQ ID NO:279) y rs2776266-rs2835001_R:CCTTCTTTTCCCCCAGAATC (SEQ ID NO:280), rs1984014-rs1984015_F:TAGGAGAACAGAAGATCAGAG (SEQ ID NO:281) y rs1984014-rs1984015_R:AAAGACTATTGCTAAATGCTTG (SEQ ID NO:282), rs7281674-rs2835316_F: TAAGCGTAGGGCTGTGTGTG (SEQ ID NO:283) y rs7281674-rs2835316_R: GGACGGATAGACTCCAGAAGG (SEQ ID NO:284), rs13047304-rs13047322_F: GAATGACCTTGGCACTTTTATCA (SEQ ID NO:285) y rs13047304-rs13047322_R: AAGGATAGAGATATACAGATGAATGGA (SEQ ID NO:286), rs2835735-rs2835736_F: CATGCACCGCGCAAATAC (SEQ ID NO:287) y rs2835735-rs2835736_R: ATGCCTCACCCACAAACAC (SEQ ID NO:288), rs13047608-rs2835826_F: TCCAAGCCCTTCTCACTCAC (SEQ ID NO:289) y rs13047608-rs2835826_R: CTGGGACGGTGACATTTTCT (SEQ ID NO:290), rs2836550-rs2212596_F: CCCAGGAAGAGTGGAAAGATT (SEQ ID NO:291) y rs2836550-rs2212596_R: TTAGCTTGCATGTACCTGTGT (SEQ ID NO:292), rs2836660-rs2836661_F: AGCTAGATGGGGTGAATTTT (SEQ ID NO:293) y _R: TGGGCTGAGGGGAGATTC (SEQ ID NO:294), rs465612-rs8131220_F: ATCAAGCTAATTAATGTTATCT (SEQ ID NO:295) y rs465612-rs8131220_R: AATGAATAAGGTCCTCAGAG (SEQ ID NO:296), rs9980072-rs8130031_F:TTTAATCTGATCATTGCCCTA (SEQ ID NO:297) y rs9980072-rs8130031_R: AGCTGTGGGTGACCTTGA (SEQ ID NO:298), rs418359-rs2836926_F: TGTCCCACCATTGTGTATTA (SEQ ID NO:299) y rs418359-rs2836926_R: TCAGACTTGAAGTCCAGGAT (SEQ ID NO:300), rs7278447-rs7278858_F: GCTTCAGGGGTGTTAGTTTT (SEQ ID NO:301) y rs7278447-rs7278858_R: CTTTGTGAAAAGTCGTCCAG (SEQ ID NO:302), rs385787-rs367001_F:CCATCATGGAAAGCATGG (SEQ ID NO:303) y rs385787-rs367001_R: TCATCTCCATGACTGCACTA (SEQ ID NO:304), rs367001-rs386095_F: GAGATGACGGAGTAGCTCAT (SEQ ID NO:305) y rs367001-rs386095_R: CCCAGCTGCACTGTCTAC (SEQ ID NO:306), rs2837296-rs2837297_F: TCTTGTTCCAATCACAGGAC (SEQ ID NO:307) y rs2837296-rs2837297_R: ATGCTGTTAGCTGAAGCTCT (SEQ ID NO:308), y rs2837381-rs4816672_F: TGAAAGCTCCTAAAGCAGAG (SEQ ID NO:309) y rs2837381-rs4816672_R:TTGAAGAGATGTGCTATCAT (SEQ ID NO:310). Se pueden incluir secuencias de polinucleótidos, por ejemplo secuencias de abrazadera GC, para asegurar la hibridación específica de los cebadores ricos en AT (Ghanta et al., PLOS ONE 5(10): doi10.1371/journal.pone.0013184 [2010], disponible en la red mundial en plosone.org). Un ejemplo de una secuencia de abrazadera GC que se puede incluir o en el 5' del cebador directo o en el 3' del cebador inverso es la GCCGCCTGCAGCCCGCGCCCCCCGTGCCCCCGCCCCGCCGCCGGCCCGGGCGCC (SEQ ID NO:311).

Se realiza la preparación de la muestra y el enriquecimiento de la biblioteca de secuenciación de ADN libre de células, una muestra de ADN libre de células purificado y una muestra de plasma de acuerdo con el método descrito en los Ejemplos 8, 9 y 10, respectivamente. Todas las bibliotecas de secuenciación se preparan como se describe en el Ejemplo 2a, y la secuenciación se realiza como se describe en el Ejemplo 2b incluyendo secuenciación de extremos emparejados. El análisis de los datos de secuenciación para la determinación de aneuploidía fetal se realiza como se describe en los Ejemplos 4 y 5. Concomitante al análisis para determinar la aneuploidía, los datos de secuenciación se analizan para determinar la fracción fetal de la manera siguiente. Después de la transferencia de los archivos de imagen y designación de bases al servidor Unix que ejecuta el software "Genome Analyzer Pipeline" versión 1.51 de Illumina como se ha descrito, las lecturas de 36bp se alinean con un 'genoma de SNP en tándem' usando el programa BOWTIE. El genoma de SNP en tándem se identifica como el agrupamiento de secuencias de ADN que abarcan los alelos de los 58 pares de SNP en tándem divulgados anteriormente. Sólo se usan lecturas que mapean únicamente para el genoma de SNP en tándem para el análisis de la fracción fetal. Las lecturas que equivalen perfectamente con el genoma de SNP en tándem se cuentan como etiquetas y se filtran. De las lecturas restantes, sólo se cuentan como etiquetas y se incluyen en el análisis las lecturas que tienen una o dos discrepancias. Se cuentan las etiquetas mapeadas para cada uno de los alelos de SNP en tándem, y se determina la fracción fetal esencialmente como se ha descrito en el Ejemplo 6 anterior pero contando para las etiquetas mapeadas para los dos alelos x e y de ENP en tándem presentes en cada una de las secuencias de ácidos nucleicos objetivo polimórficos amplificadas que se amplifican para enriquecer las muestras, ese decir % de $alelo_{x+y}$ de la fracción fetal = ((ΣEtiquetas de secuencia fetal para el $alelo_{x+y}$) / (ΣEtiquetas de secuencia materna para el $alelo_{x+y}$)) x 100. Opcionalmente, la fracción de los ácidos nucleicos fetales en la mezcla de los ácidos nucleicos fetales y maternos se calcula para cada uno de los alelos informativos ($alelo_{x+y}$) de la manera siguiente.

$$\%\ \text{de alelo}_{x+y}\ \text{de la fracción fetal} = ((2\ X\ \Sigma\text{Etiquetas de secuencia fetal para el alelo}_{x+y}) / (\Sigma\text{Etiquetas de secuencia materna para el alelo}_{x+y})) \times 100$$

para compensar por la presencia de 2 conjuntos de alelos fetales en tándem, uno estando enmascarado por el fondo materno. Las secuencias de SNP en tándem son informativas cuando la madre es heterocigótica y hay presente un tercer haplotipo paterno, permitiendo una comparación cuantitativa entre el haplotipo heredado maternalmente y el haplotipo heredado paternalmente para calcular la fracción fetal calculando la Proporción de Haplotipos (HR). El porcentaje de fracción fetal se calcula para al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 25, al menos 30, al menos 40 o más conjuntos informativos de alelos en tándem. En una realización, la fracción fetal es la fracción fetal media determinada para al menos 3 conjuntos informativos de alelos en tándem.

Ejemplo 13

Determinación de Fracción Fetal por Secuenciación Masivamente Paralela de una Biblioteca Objetivo que Comprende Ácidos Nucleicos Polimórficos que Comprenden SNPs

Para determinar la fracción de ADN libre de células fetal en una muestra materna, se amplificaron secuencias de ácidos nucleicos polimórficos objetivo que compendia cada una un SNP y se usaron para preparar una biblioteca objetivo para secuenciar de una manera masivamente paralela.

El ADN libre de células se extrajo como se ha descrito en el Ejemplo 1. Se preparó una biblioteca de secuenciación de la manera siguiente. Se amplificó ADN libre de células contenido en 5µl de ADN libre de células purificado en un volumen de reacción de 50µl que contenía 7,5µl de una mezcla de cebadores de 1µM (Tabla 10), 10 µl de Mezcla maestra NEB 5X y 27 µl de agua. El ciclado térmico se realizó con el Gene Amp9700 (Applied Biosystems) usando las siguientes condiciones de ciclado: incubación a 95º C durante 1 minuto, seguido por 20-30 ciclos a 95º C durante 20 segundos, 68º C durante 1 minuto y 68º C durante 30 segundos, a lo que siguió una incubación final a 68º C durante 5 minutos. Se añadió un mantenimiento final a 4º C hasta que las muestras se retiraron para la combinación con la porción no amplificada de la muestra de ADN libre de células purificado. El producto amplificado se purificó usando el sistema de purificación por PCR Agencourt AMPure XP (Parte No. A63881; Beckman Coulter Genomics, Danvers, MA), y la concentración se cuantificó usando el Nanodrop 2000 (Thermo Scientific, Wilmington, DE). Se añadió un mantenimiento final a 4º C hasta que se retiraron las muestras para preparar la biblioteca objetivo. El producto amplificado se analizó con un Bioanalizador 2100 (Agilent Technologies, Sunnyvale, CA), y se determinó la concentración del producto amplificado. Se preparó una biblioteca de secuenciación de los ácidos nucleicos objetivo amplificados usando el protocolo abreviado descrito en el Ejemplo 2, y se secuenció de una manera masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles y de acuerdo con el protocolo de Illumina. Se realizó el análisis y recuento de etiquetas mapeadas para un genoma de referencia que consistía de 26 secuencias (13 paras cada representando dos alelos) que comprendían un SNP, es decir SEQ ID NO: 1-56 como se ha descrito.

La Tabla 23 proporciona los recuentos de etiquetas obtenidos de la secuenciación de la biblioteca objetivo,

y la fracción fetal calculada derivada de los datos de secuenciación.

TABLA 23

Determinación de la Fracción Fetal por Secuenciación Masivamente Paralela de una Biblioteca de Ácidos nucleicos Polimórficos que Comprenden SNPs		
SNP	**RECUENTOS DE ETIQUETAS DE SNP**	**Fracción Fetal (%)**
rs10773760.1|Chr.12|longitud=128|alelo=A	236590	1.98
rs10773760.2|Chr.12|longitud=128|alelo=G	4680	
rs13182883.1|Chr.5|longitud=111|alelo=A	3607	4.99
rs13182883.2|Chr.5|longitud=111|alelo=G	72347	
rs4530059.1|Chr.14|longitud=110|alelo=A	3698	1.54
rs4530059.1|Chr.14|longitud=110|alelo=G	239801	
rs8078417.1|Chr.17|longitud=110|alelo=C	1E+06	3.66
rs8078417.2|Chr.17|longitud=110|alelo=T	50565	
Fracción Fetal (**Media±D**.S.) = **12**.**4±6**.**6**		

Los resultados muestran que las secuencias de ácidos nucleicos polimórficos que comprenden cada una al menos un SNP pueden amplificarse de ADN libre de células derivado de una muestra de plasma materno para construir una biblioteca que puede ser secuenciada de una manera masivamente paralela para determinar la fracción de ácidos nucleicos fetales en la muestra materna. Los métodos de secuenciación masivamente paralelos para determinar la fracción fetal se pueden usar en combinación con otros métodos para proporcionar diagnóstico de aneuploidía fetal y otras pruebas prenatales.

Ejemplo 14

Determinación de la Fracción Fetal por Secuenciación Masivamente Paralela de una Biblioteca Objetivo que Comprende Ácidos Nucleicos Polimórficos que comprenden STRs o SNPs en tándem

La fracción fetal se puede determinar independientemente de la determinación de la aneuploidía usando una biblioteca objetivo que comprende SNPs o STRs como se describe para la biblioteca objetivo de SNP del Ejemplo 13. Para preparar una biblioteca objetivo de SNP en tándem, se usa una porción de una biblioteca de ADN libre de células purificada que comprende ácidos nucleicos fetales y maternos para amplificar secuencias objetivo usando una mezcla de cebadores, por ejemplo Tablas 10 y 11. Para preparar una biblioteca de STR, se usa una porción de una biblioteca de ADN libre de células purificada que comprende ácidos nucleicos fetales y maternos para amplificar secuencias objetivo usando una mezcla de cebadores, por ejemplo Tabla 22. La biblioteca objetivo de SNP en tándem se secuencia como se describe en el Ejemplo 12.

Las bibliotecas objetivo se secuencian como se ha descrito, y la fracción fetal se determina a partir del número de etiquetas de secuencia mapeadas para el genoma de referencia de SRT o SNP en tándem respectivamente comprendiendo todos los alelos de STR o SNP en tándem abarcados por los cebadores. Se identifican los alelos informativos, y la fracción fetal se determina usando el número de etiquetas mapeadas para los alelos de las secuencias polimórficas.

Ejemplo 15

Determinación de la Fracción fetal por Electroforesis Capilar de Secuencias Polimórficas que Comprenden STRs

Para determinar la fracción fetal en muestras maternas que comprenden ADN libre de células fetal y materno, se recogieron muestras de sangre periférica de mujeres embarazadas voluntarias que llevaban fetos masculinos o femeninos. Las muestras de sangre periférica se obtuvieron y procesaron para proporcionar ADN libre de células purificado como se describe en el Ejemplo 1.

Se analizaron diez microlitros de muestras de ADN libre de células usando el kit de amplificación por PCR AmpF1STR® MiniFiler™ (Applied Biosystems, Foster City, CA) de acuerdo con las instrucciones del fabricante. Brevemente, se amplificó el ADN libre de células contenido en 10 µl en un volumen de reacción de 25 µl que

contenía 5µl de cebadores etiquetados por fluorescencia (Conjunto de cebadores AmpFlSTR® MiniFiler™) y la Mezcla Maestra AmpFlSTR® MiniFiler™, que incluye ADN polimerasa AmpliTaq Gold® y el tampón asociado, sal (1,5 mM MgCl2) y 200 µM de desoxinucleótido trifosfatos (dNTPs: dATP, dCTP, dGTP y dTTP). Los cebadores etiquetados por fluorescencia son cebadores directos que están etiquetados con colorantes 6FAM™, VIC™, NED™, y PET™. Se realizó el ciclado térmico con el Gene Amp9700 (Applied Biosystems) usando las siguientes condiciones de ciclado: incubación a 95º C durante 10 minutos, seguido por 30 ciclos a 94º C durante 20 segundos, 59º C durante 2 minutos y 72º C durante 1 minuto, a lo que siguió una incubación final a 60º C durante 45 minutos. Se añadió un mantenimiento final a 4º C hasta que las muestras se retiraron para análisis. El producto amplificado se preparó diluyendo 1ul de producto amplificado en 8,7ul de Hi-DiTM formamida (Applied Biosystems) y 0,3 µl de estándar de tamaño interno GeneScanTM-500 LIZ_ (Applied Biosystems), y se analizó con un Analizador Genético ABI PRISM3130x1 (Applied Biosystems) usando el Data Collection HID_G5_POP4 (Applied Biosystems), y una matriz capilar de 36-cm. Todo el genotipado se realizó con el software GeneMapper_ID v3.2 (Applied Biosystems) usando las escaleras alélicas e intervalos y paneles proporcionados por el fabricante.

Toda medición del genotipado se realizó en el Analizador Genético Applied Biosystems 3130xl, usando una "ventana" de ±0,5-nt de alrededor del tamaño obtenido para cada alelo para permitir la detección y asignación correcta de los alelos. Cualquier alelo de la muestra cuyo tamaño estaba fuera de la ventana de ±0,5-nt se determinó que era OL, es decir "Fuera de la escalera". Los alelos OL son alelos de un tamaño que no está representado en la Escalera Alélica AmpFlSTR® MiniFiler™ e un alelo que no corresponde con una escalera alélica, pero cuyo tamaño está justo fuera de una ventana debido a un error de medición. Se estableció el límite de altura del pico mínimo de >50 RFU en base a los experimentos de validación realizados para evitar el tipificado cuando es probable que los efectos estocásticos interfieran con la interpretación precisa de las mezclas. El cálculo de la fracción fetal se basa en hacer la media de todos los marcadores informativos. Los marcadores informativos se identifican por la presencia de picos en el electroferograma que cae dentro de los parámetros de los intervalos preestablecidos para las STRs que se analizan.

Los cálculos de la fracción fetal se realizaron usando la altura de pico media para los alelos principales y menores en cada locus de STR determinado por inyecciones por triplicado. Las reglas aplicadas al cálculo son:

1. Los datos de alelos fuera de escalera (OL) para alelos no incluidos en el cálculo; y
2. Sólo se incluyen en el cálculo alturas de pico derivadas de > 50 RFU (unidades de fluorescencia relativas)
3. Si uno de los intervalos está presente el marcador se considera no informativo; y
4. Si se designa un segundo intervalo pero los picos del primer y segundo intervalos están dentro del 50-70% de sus unidades de fluorescencia relativa (RFU) en la altura del pico, la fracción minoritaria no se mide y el marcador se considera no informativo.

La fracción del alelo menor para cualquier marcador informativo dado se calcula dividiendo la altura del pico del componente menor por la suma de la altura del pico para el componente principal, y se expresa como un porcentaje que fue primero calculado para cada locus informativo como

fracción fetal = (Σaltura del pico del alelo menor / Σaltura del pico del alelo(s) principal)) x 100

La fracción fetal para una muestra que comprende dos o más STRs informativas, se calcularía como la media de las fracciones fetales calculadas para dos o más marcadores informativos.

La Tabla 8 proporciona los datos obtenido de analizar ADN libre de células de un sujeto embarazado con un feto masculino.

Fracción Fetal Determinada en ADN libre de células de un Sujeto Embarazado por análisis de STRs								
STR	Alelo 1	Aleleo 2	Aleleo 3	Altura del Alelo 1	Altura del Alelo 2	Altura del Alelo 3	Fracción Fetal	Fracción Fetal (Media/STR)
AMEL	X	Y		3599	106		2.9	
AMEL	X	Y		3602	110		3.1	
AMEL	X	Y		3652	109		3.0	3.0
CSF1PO	11	12		2870	2730			
CSF1PO	11	12		2924	2762			
CSF1PO	11	12		2953	2786			
D13S317	11	12		2621	2588			
D13S317	11	12		2680	2619			
D13S317	11	12		2717	2659			
D16S539	9	11		1056	1416			
D16S539	9	11		1038	1394			
D16S539	9	11		1072	1437			
D18S51	13	15		2026	1555			
D18S51	13	15		2006	1557			
D18S51	13	15		2050	1578			
D21S11	28	31.2		2450	61		2.5	
D21S11	28	31.2		2472	62		2.5	
D21S11	28	31.2		2508	67		2.7	2.6
D2S1338	20	23		3417	3017			
D2S1338	20	23		3407	3020			
D2S1338	20	23		3493	3055			
D7S820	9	12	13	2373	178	1123	5.1	
D7S820	9	12	13	2411	181	1140	5.1	
D7S820	9	12	13	2441	182	1156	5.1	5.1
FGA	17.2	22	25	68	1140	896	3.3	
FGA	17.2	22	25	68	1144	909	3.1	
FGA	17.2	22	25	68	1151	925	3.3	3.2
Fracción Fetal = 3.5								

Los resultados muestran que el ADN libre de células puede usarse para determinar la presencia o ausencia de ADN fetal como se indica por la detección de un componente menor en uno o más alelos de STR, para determinar el porcentaje de fracción fetal y para determinar el género fetal como se indica por la presencia o ausencia del alelo de Amelogenina.

Ejemplo 16

Uso de la Fracción Fetal para Establecer Límites y Estimar el Tamaño de Muestra Mínimo en Detección de Aneuploidías

Los recuentos de equivalencias de secuencias para cromosomas diferentes se manipulan para generar una

puntuación que variará con el número de copias de cromosomas que pueden ser interpretadas para identificar la amplificación o deleción cromosómica. Por ejemplo, dicha puntuación podría ser generada comparando la cantidad relativa de un etiqueta de secuencia en un cromosoma sometido a cambios de número copias para un cromosoma que se sabe es una euploidia. Ejemplos de puntuaciones que se pueden usar para identificar la amplificación o deleción incluyen, pero no están limitados a: recuentos para el cromosoma de interés dividido por recuentos de otro cromosoma de la misma ejecución experimental, los recuentos para el cromosoma de interés divididos por el número total de recuentos para la ejecución experimental, comparación de recuentos de la muestra de interés frente a una muestra de control separada. Sin perder la generalidad, se puede asumir que las puntuaciones aumentarán a medida que aumente el número de copias. El conocimiento de la fracción fetal se puede usar para establecer límites de "corte" para designar estados de "aneuploidía", "normal" o "marginal" (incierto). Después, se realizan cálculos para estimar el número mínimo de secuencias requeridas para conseguir la sensibilidad adecuada (es decir, probabilidad de identificar correctamente un estado de aneuploidía).

La **Figura 19** es un gráfico de dos poblaciones diferentes de puntuaciones. El eje x es la puntuación y el eje y es la frecuencia. Las puntuaciones en las muestras de cromosomas sin aneuploidía pueden tener una distribución mostrada en la **Figura 19A**. La **Figura 19B** ilustra una distribución hipotética de una población de puntuaciones en muestras con un cromosoma amplificado. Sin perder la generalidad, los gráficos y ecuaciones muestran el caso de una puntuación univariante donde la condición de aneuploidía representa una ampliación del número de copias. Los casos multivariante y/o anomalías de reducción/deleción son extensiones simples o reorganizaciones de las descripciones dadas y se pretende que caigan dentro del ámbito de esta técnica.

La cantidad de "superposición" entre las poblaciones puede determinar como de bien se pueden discriminar los casos normales y de aneuploidía. En general, aumentar la fracción fetal, ff, aumenta la potencia de la discriminación separando los dos centros de población (moviendo "C2", el "Centro de Puntuaciones de Aneuploidía" y aumentando "d", provocando que las poblaciones se superpongan menos. Además, un aumento en el valor absoluto de la magnitud, m, (por ejemplo teniendo cuatro copias del cromosoma en lugar de una trisomía) de la amplificación también aumentará la separación de los centros de población llevando a potencia más alta (es decir, probabilidad más alta de identificar correctamente estados de aneuploidía).

Aumentar el número de secuencias generadas, N, reduce las desviaciones estándar "sdevA" y/o "sdevB", la propagación de las dos poblaciones de puntuaciones, lo que también provoca que las poblaciones se superpongan menos.

Establecer Límites y Estimar el Tamaño de la Muestra

El siguiente procedimiento puede usarse para establecer "c", el valor crítico para designar estados de "aneuploidía", "normal" o "marginal" (incierto). Sin perder la generalidad, se usan a continuación pruebas estadísticas unilaterales.

Primero, se decide una proporción de falsos positivos aceptable, FP (denominada algunas veces "error tipo I" o "especificidad", que es la probabilidad de un falso positivo o designar falsamente una aneuploidía. Por ejemplo, la FP puede ser al menos, o alrededor de 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, o 0.1.

Segundo, se puede determinar el valor de "c" resolviendo la ecuación : FP = integral de c a infinito de (f1(x)dx).

$$\mathrm{FP} = \int_{c}^{\infty} f1(x)dx \qquad \text{(Ecuación 1)}$$

Una vez que se ha determinado un valor crítico, c, se puede estimar el número mínimo de secuencias requeridas para conseguir una acierta TP = proporción de verdaderos positivos. La proporción de verdaderos positivos puede ser, por ejemplo, de alrededor de 0.5, 0.6, 0.7, 0.8, ó 0.9. En una realización, la proporción de verdaderos positivos puede ser 0.8. En otras palabras, N es el número mínimo de secuencias requeridas para identificar aneuploidía 100*TP porciento del tiempo. N = número mínimo de tal forma que TP = integral de c a infinito de f2(x,ff)dx > 0.8. N se determina resolviendo

$$\min_{N} s.t.\{TB \geq \int_{c}^{\infty} f2(x,N)dx\} \qquad \text{(Ecuación 2)}$$

En las pruebas estadísticas clásicas f1 y f2 sin a menudo F, las distribuciones F no centrales (un caso especial de distribuciones t y t no central) aunque no es una condición necesaria para esta solicitud.

Establecer "Niveles" de Límites para Dar Más Control de Errores

Los límites también pueden establecerse en etapas usando los métodos anteriores. Por ejemplo, puede establecerse un límite de alta confianza designando "aneuploidía", expresar ca, usando FP 0.001 y un límite "marginal", expresar cb, usando FP 0.05. En este caso si la Puntuación, S:

(S > ca)	entonces designar "Trisomía"
(cb > S <= ca)	entonces designar "Marginal"
(S < cb)	entonces designar "Normal"

Algunas Generalizaciones Triviales Que Caen Dentro del Ámbito de esta Técnica

Se pueden usar diferentes valores para los límites como TP, FP, etc. Los procedimientos pueden ser ejecutados en cualquier orden. Por ejemplo, uno puede comenzar con N y resolverse para c, etc. Las distribuciones pueden depender de ff de modo que f1(x,N,ff), f2(x,N,ff), y/o otras variables. Las anteriores ecuaciones integrales pueden resolverse con referencia a tablas o por métodos informáticos iterativos. Se puede estimar un parámetro de no centralidad y la potencia se puede leer de tablas estadísticas estándar. La potencia estadística y los tamaños de muestra se pueden derivar del cálculo o estimación de medias cuadráticas esperadas. Se pueden usar las formas cerradas de distribuciones teóricas como f, t, t no central, normal, etc. o estimaciones (kernel u otras) para modelar las distribuciones f1, f2. Se puede usar el establecimiento de límites empírico y la selección de parámetros usando Curvas Características Operativas del Receptor (ROC) y se pueden cotejar con la fracción fetal. Se pueden usar varias estimaciones de la propagación de la distribución (varianza, desviación media absoluta, rango intercuartil, etc.). Se pueden usar varias estimaciones del centro de distribución (media, mediana, etc.). Se pueden usar pruebas estadísticas bilaterales en oposición a unilaterales. Se puede reformular la prueba de hipótesis simple como regresión lineal o no lineal. Se pueden usar métodos combinatorios, simulación (por ejemplo monte carlo), maximización (por ejemplo, expectativa de maximización), iterativos u otros métodos independientemente o en conjunción con lo anterior para establecer la potencia estadística o los límites.

Ejemplo 17

Demostración de Detección de Aneuploidía

Los datos de secuenciación obtenidos de las muestras descritas en los Ejemplos 4 y 5, y mostrados en las figuras 9-13 se analizaron adicionalmente para ilustrar la sensibilidad del método para identificar con éxito aneuploidías en muestras maternas. Las dosis de cromosomas normalizadas para los cromosomas 21, 18, 13, X e Y se analizaron como una distribución relativa a la desviación estándar de la media (eje Y) y mostrada en la Figura 21. El cromosoma normalizador usado se muestra como el denominador (eje X).

La Figura 20 (A) muestra la distribución de dosis de cromosomas en relación a la desviación estándar de la media para la dosis de cromosoma 21 en las muestras no afectadas (normales) (o) y las muestras de trisomía 21 (T21; Δ) cuando se usa el cromosoma 14 como el cromosoma normalizador para el cromosoma 21. La Figura 20 (B) muestra la distribución de dosis de cromosomas en relación a la desviación estándar de la media para la dosis de cromosoma 18 en las muestras no afectadas (o) y las muestras de trisomía 18 (T18; Δ) cuando se usa el cromosoma 8 como el cromosoma normalizador para el cromosoma 18. La Figura 20 (C) muestra la distribución de dosis de cromosomas en relación a la desviación estándar de la media para la dosis de cromosoma 13 en las muestras no afectadas (o) y las muestras de trisomía 13 (T13; Δ) usando la densidad de etiqueta de secuencia media del grupo de cromosomas 3, 4, 5 y 6 como el cromosoma normalizador para determinar la dosis de cromosoma para el cromosoma 13. La Figura 20 (D) muestra la distribución de dosis de cromosomas en relación a la desviación estándar de la media para la dosis del cromosoma X en las muestras femeninas no afectadas (o), las muestras masculinas no afectadas (Δ), y las muestras de monosomía X (XO;+) cuando se usa el cromosoma 4 como el cromosoma normalizador para el cromosoma X. La Figura 20 (E) muestra la distribución de dosis de cromosomas en relación a la desviación estándar de la media para la dosis del cromosoma Y en las muestras masculinas no afectadas (o), las muestras femeninas no afectas (Δ) y las muestras de monosomía X (+) cuando se usa la densidad de etiqueta de secuencia media del grupo de cromosomas 1-22 y X como el cromosoma normalizador para determinar la dosis de cromosoma para el cromosoma Y.

Los datos muestran que la trisomía 21, trisomía 18, trisomía 13 eran claramente distinguibles de las muestras no afectadas (normales). Se identificó fácilmente que las muestras de monosomía X tenían una dosis de cromosoma X que era claramente inferior que las de las muestras femeninas no afectadas (Figura 20 (D)), y que tenían dosis de cromosoma Y que eran claramente inferiores que las de las muestras masculinas no afectadas (Figura 20(E)).

Por lo tanto el método divulgado en la presente es sensible y específico para determinar la presencia o ausencia de aneuploidías cromosómicas en una muestra de sangre materna.

Reivindicaciones

1. Un método para preparar un biblioteca de secuenciación a partir de una muestra materna que comprende una mezcla de moléculas de ácidos nucleicos fetales y maternos, en donde el método comprende los pasos consecutivos de reparación de extremos, adición de colas de dA y ligado por adaptadores de dichos ácidos nucleicos, y en donde dichos pasos consecutivos excluyen purificar los productos reparados en el extremo antes del paso de la adición de colas de dA y excluyen purificar los productos de la adición de colas de dA antes del paso de ligado por adaptadores.

2. El método de la Reivindicación 1, en donde dichos pasos consecutivos se realizan en ausencia de polietilenglicol.

3. El método de la Reivindicación 1 o la Reivindicación 2, en donde dichos pasos consecutivos se realizan en menos de 1 hora.

4. El uso de la biblioteca de secuenciación preparada por el método de cualquiera de las Reivindicaciones 1-3 en un método de secuenciación masivamente paralelo.

5. El uso de la Reivindicación 4, en donde dicho método de secuenciación es un método para determinar una aneuploidía cromosómica fetal en la muestra materna.

6. El uso de la Reivindicación 5, dicho método comprendiendo:

(a) secuenciar al menos una porción de dichas moléculas de ácidos nucleicos en dicha biblioteca de secuenciación, obteniendo de esta manera información de secuencia para una pluralidad de moléculas de ácidos nucleicos fetales y maternos de una muestra materna, en donde la información de secuencia comprende lecturas de secuencias; y
(b) usar la información de secuencia para identificar un número de etiquetas de secuencias mapeadas para al menos un cromosoma normalizador y para un cromosoma aneuploide, comparando las secuencias de las lecturas de secuencias con la secuencia de un genoma de referencia humano para determinar el origen cromosómico de las moléculas de ácidos nucleicos secuenciados;
(c) calcular una dosis de cromosoma para dicho cromosoma aneuploide como:

(i) una proporción del número de etiquetas de secuencias mapeadas identificadas para dicho cromosoma aneuploide y el número de etiquetas de secuencias mapeadas identificadas para el al menos un cromosoma normalizador; o
(ii) una proporción de una proporción de la densidad de etiqueta de secuencia para dicho cromosoma aneuploide y una proporción de la densidad de etiqueta de secuencia para dicho al menos un cromosoma normalizador, en donde la proporción de la densidad de etiqueta de secuencia para dicho cromosoma aneuploide se calcula relacionando el número de etiquetas de secuencias mapeadas identificadas par dicho cromosoma aneuploide en el paso (b) con la longitud de dicho cromosoma aneuploide, y la proporción de la densidad de etiqueta de secuencia para dicho al menos un cromosoma normalizados se calcula relacionando el número de etiquetas de secuencias mapeadas identificadas para dicho al menos un cromosoma normalizador en el paso (b) con la longitud de dicho al menos un cromosoma normalizador; y

(d) comparar dicha dosis con un valor límite, en donde dicho valor límite es un número que sirve como un límite de diagnóstico de una aneuploidía, y determinar de esta manera la presencia o ausencia de aneuploidía fetal, en donde la presencia de una aneuploidía fetal se identifica si la dosis de cromosoma excede el valor límite, en donde:

(i) dicho al menos un cromosoma normalizador es un cromosoma o grupo de cromosomas que en un conjunto de datos de calificación de muestras que comprenden cromosomas presentes en un número de copias conocido y no aneuploide para el cromosoma de interés mostró una variabilidad en el número de etiquetas de secuencias mapeadas para él que se aproximó mejor a la variabilidad en el número de etiquetas de secuencias mapeadas para el cromosoma de interés; y/o
(ii) dicho al menos un cromosoma normalizador es un cromosoma o grupo de cromosomas que proporcionaron la diferencia estadística más grande entre la distribución de dosis de cromosomas para el cromosoma de interés en un conjunto de datos de calificación de muestras que comprenden cromosomas presentes en un número de copias conocido y no aneuploide para el cromosoma de interés y la dosis de cromosoma para el cromosoma de interés en una o más muestras afectadas.

7. El uso de la Reivindicación 5 o la Reivindicación 6, en donde dicha aneuploidía es una aneuploidía cromosómica.

8. El uso de la Reivindicación 5 o la Reivindicación 6, en donde dicha aneuploidía es una aneuploidía parcial.

9. El uso de la Reivindicación 5 o la Reivindicación 6, en donde dicha aneuploidía es una aneuploidía cromosómica elegida de trisomía 8, trisomía 13, trisomía 15, trisomía 16, trisomía 18, trisomía 21, trisomía 22, monosomía X, XXX, XXY y XYY.

10. El uso de la Reivindicación 4, en donde dicho método de secuenciación es un método para determinar la fracción de ácidos nucleicos fetales en la muestra materna.

11. El método o uso de cualquiera de las Reivindicaciones anteriores, en donde dicha muestra materna es un fluido biológico seleccionado de sangre, plasma, suero, orina y saliva.

12. El método o uso de la Reivindicación 11, en donde dicha muestra materna es una muestra de plasma.

13. El método o uso de cualquiera de las Reivindicaciones anteriores, en donde dichas moléculas de ácidos nucleicos fetales y maternos son moléculas de ADN libre células (ADNcf).

14. El uso de la Reivindicación 4, en donde dicha secuenciación:

(i) es secuenciación de próxima generación (NGS);
(ii) es secuenciación masivamente paralela usando secuenciación por síntesis con terminadores de colorante reversibles;
(iii) es secuenciación masivamente paralela usando secuenciación por ligadura;
(iv) comprende una amplificación; o
(v) es secuenciación de moléculas individuales.

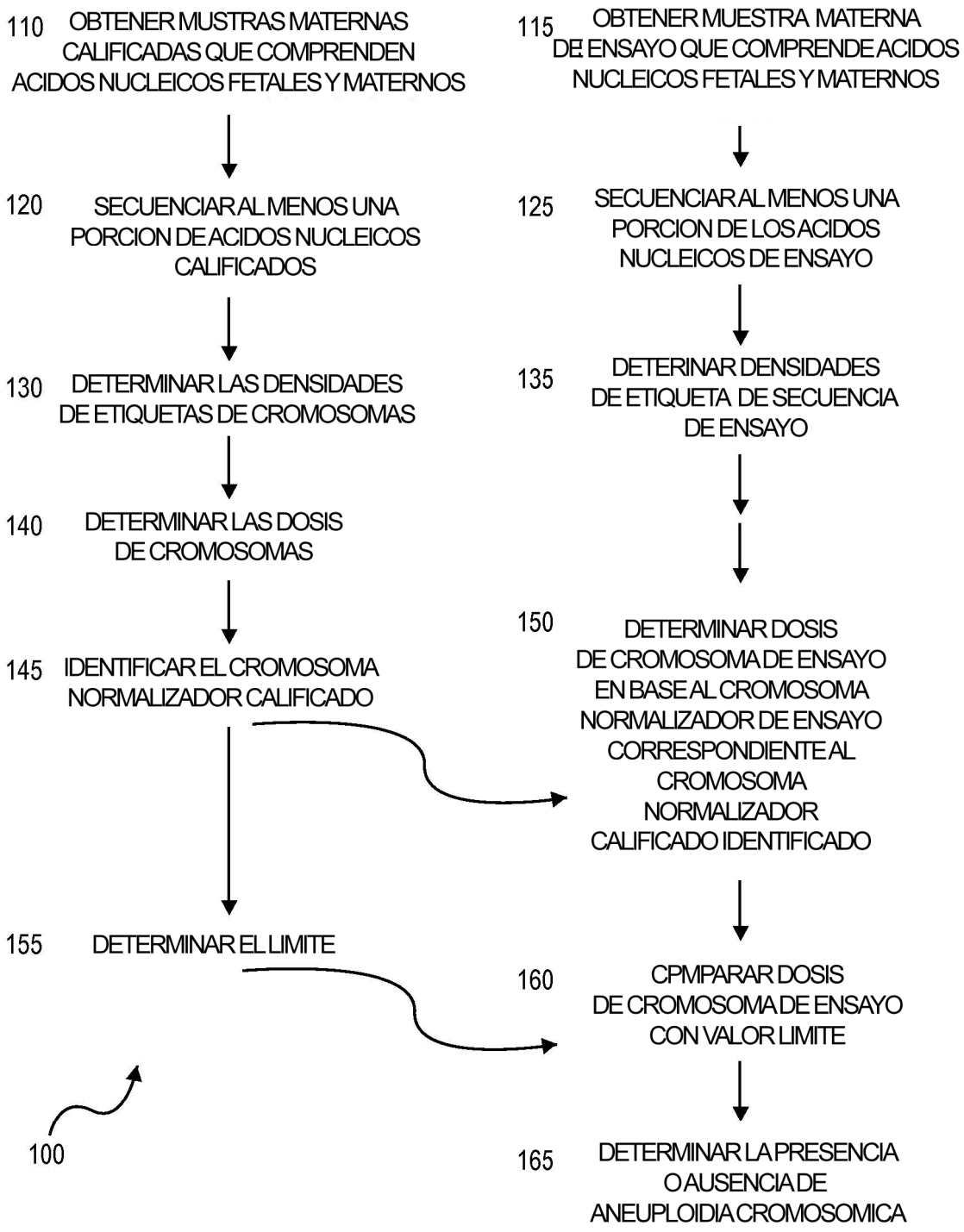
110 OBTENER MUSTRAS MATERNAS CALIFICADAS QUE COMPRENDEN ACIDOS NUCLEICOS FETALES Y MATERNOS
120 SECUENCIAR AL MENOS UNA PORCION DE ACIDOS NUCLEICOS CALIFICADOS
130 DETERMINAR LAS DENSIDADES DE ETIQUETAS DE CROMOSOMAS
140 DETERMINAR LAS DOSIS DE CROMOSOMAS
145 IDENTIFICAR EL CROMOSOMA NORMALIZADOR CALIFICADO
155 DETERMINAR EL LIMITE
100
115 OBTENER MUESTRA MATERNA DE ENSAYO QUE COMPRENDE ACIDOS NUCLEICOS FETALES Y MATERNOS
125 SECUENCIAR AL MENOS UNA PORCION DE LOS ACIDOS NUCLEICOS DE ENSAYO
135 DETERINAR DENSIDADES DE ETIQUETA DE SECUENCIA DE ENSAYO
150 DETERMINAR DOSIS DE CROMOSOMA DE ENSAYO EN BASE AL CROMOSOMA NORMALIZADOR DE ENSAYO CORRESPONDIENTE AL CROMOSOMA NORMALIZADOR CALIFICADO IDENTIFICADO
160 CPMPARAR DOSIS DE CROMOSOMA DE ENSAYO CON VALOR LIMITE
165 DETERMINAR LA PRESENCIA O AUSENCIA DE ANEUPLOIDIA CROMOSOMICA

FIG. 1

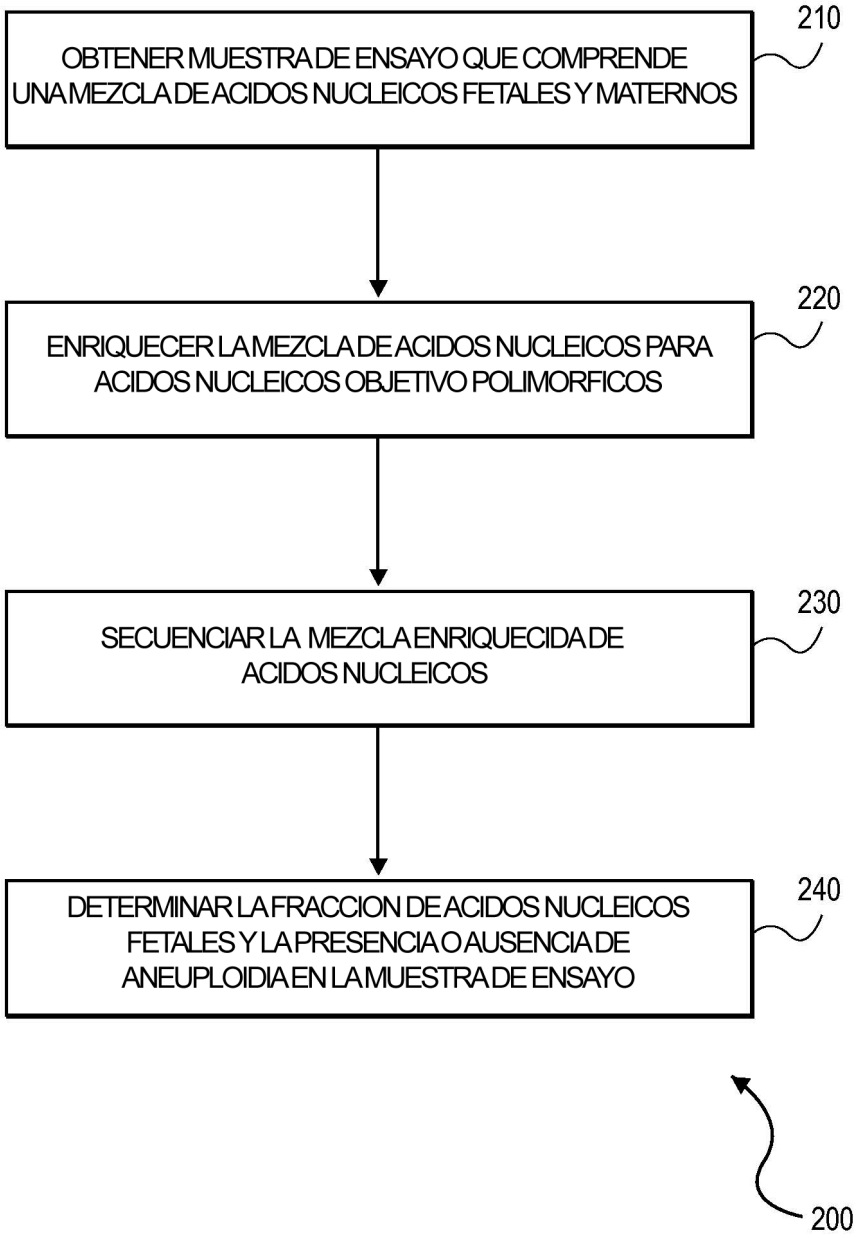
OBTENER MUESTRA DE ENSAYO QUE COMPRENDE UNA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
210
ENRIQUECER LA MEZCLA DE ACIDOS NUCLEICOS PARA ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS
220
SECUENCIAR LA MEZCLA ENRIQUECIDA DE ACIDOS NUCLEICOS
230
DETERMINAR LA FRACCION DE ACIDOS NUCLEICOS FETALES Y LA PRESENCIA O AUSENCIA DE ANEUPLOIDIA EN LA MUESTRA DE ENSAYO
240
200

FIG. 2

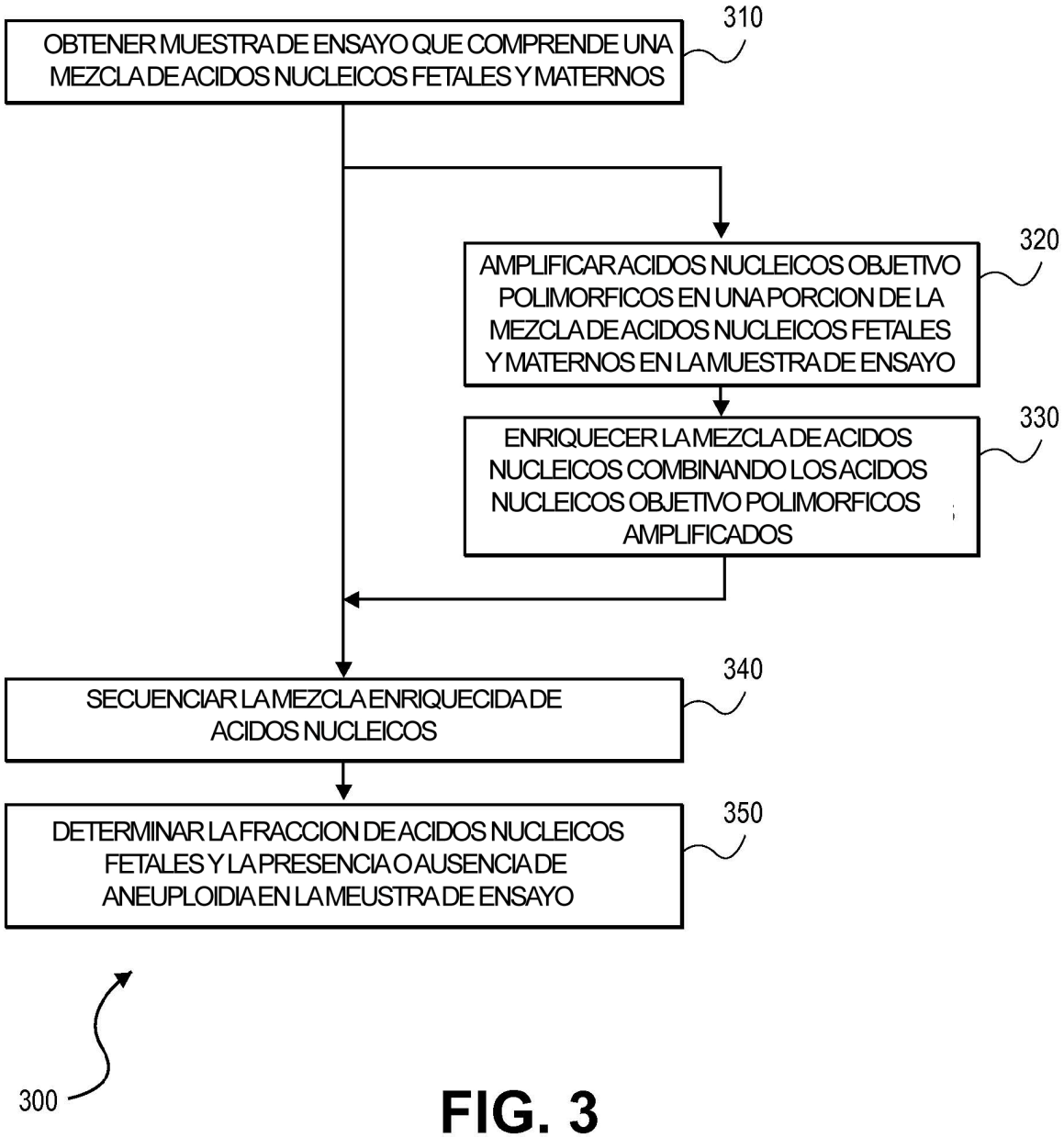
OBTENER MUESTRA DE ENSAYO QUE COMPRENDE UNA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
310
AMPLIFICAR ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS EN UNA PORCION DE LA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS EN LA MUESTRA DE ENSAYO
320
ENRIQUECER LA MEZCLA DE ACIDOS NUCLEICOS COMBINANDO LOS ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS AMPLIFICADOS
330
SECUENCIAR LA MEZCLA ENRIQUECIDA DE ACIDOS NUCLEICOS
340
DETERMINAR LA FRACCION DE ACIDOS NUCLEICOS FETALES Y LA PRESENCIA O AUSENCIA DE ANEUPLOIDIA EN LA MEUSTRA DE ENSAYO
350
300

FIG. 3

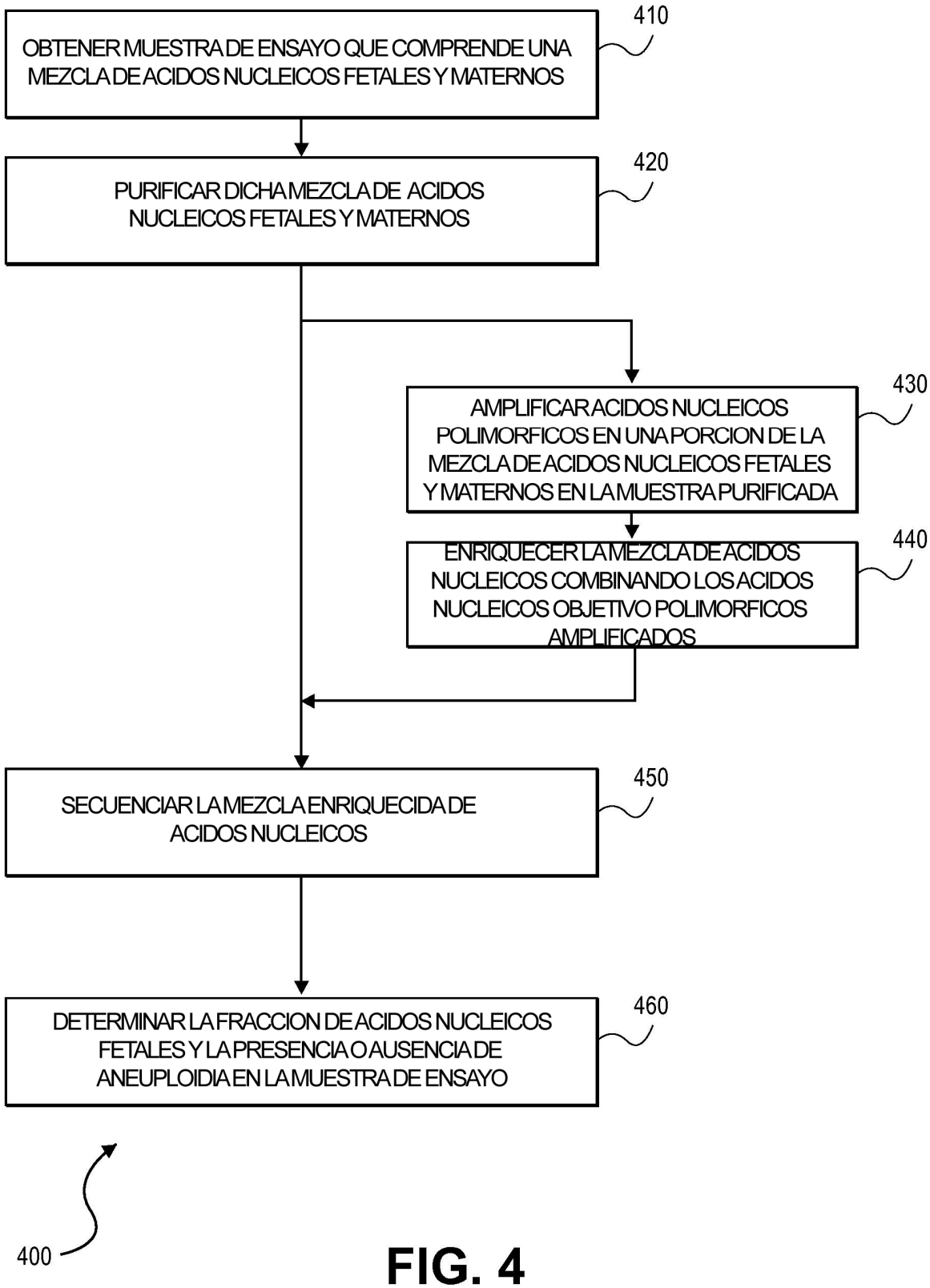
OBTENER MUESTRA DE ENSAYO QUE COMPRENDE UNA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
410
PURIFICAR DICHA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
420
AMPLIFICAR ACIDOS NUCLEICOS POLIMORFICOS EN UNA PORCION DE LA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS EN LA MUESTRA PURIFICADA
430
ENRIQUECER LA MEZCLA DE ACIDOS NUCLEICOS COMBINANDO LOS ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS AMPLIFICADOS
440
SECUENCIAR LA MEZCLA ENRIQUECIDA DE ACIDOS NUCLEICOS
450
DETERMINAR LA FRACCION DE ACIDOS NUCLEICOS FETALES Y LA PRESENCIA O AUSENCIA DE ANEUPLOIDIA EN LA MUESTRA DE ENSAYO
460
400

FIG. 4

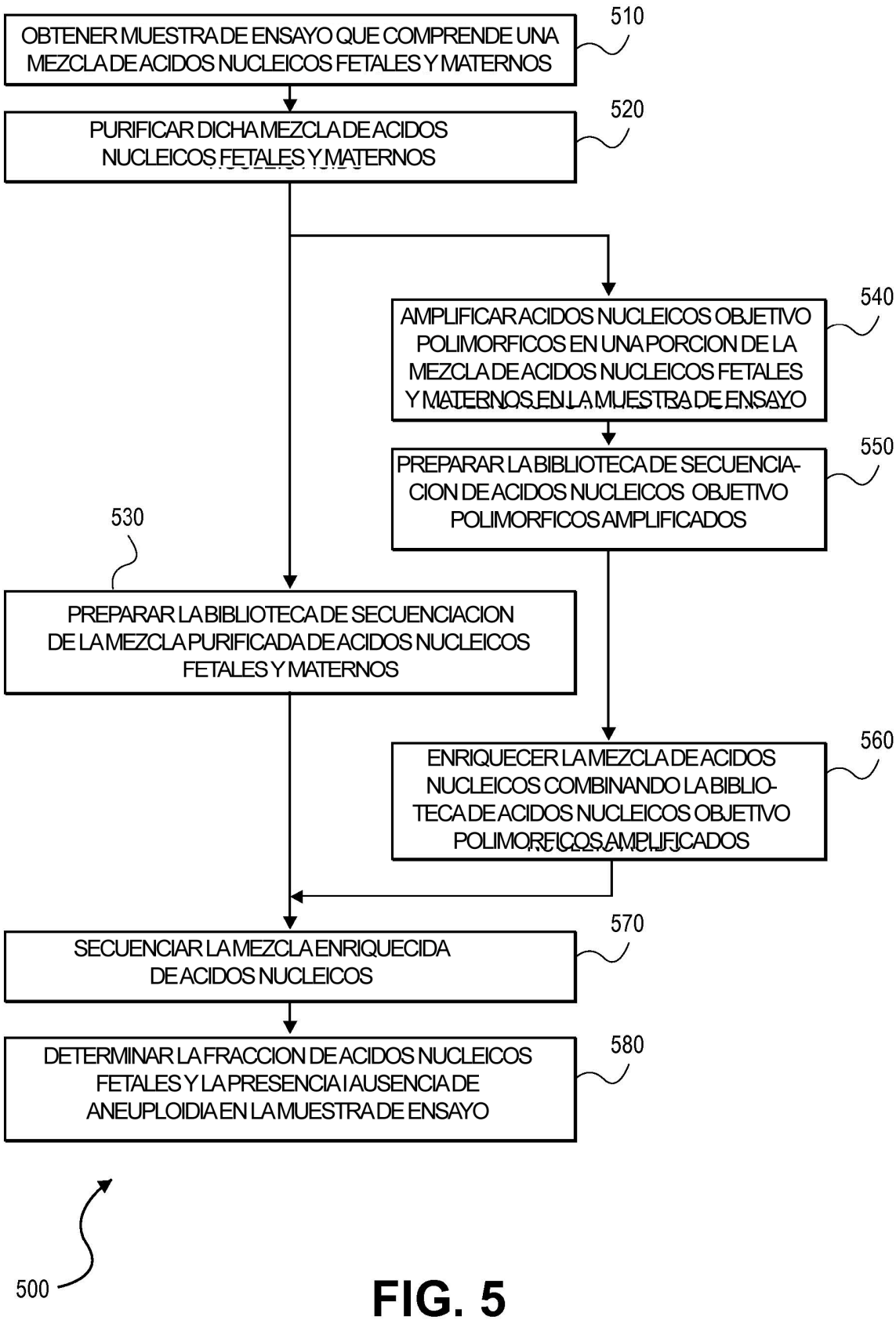
OBTENER MUESTRA DE ENSAYO QUE COMPRENDE UNA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
510
PURIFICAR DICHA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
520
AMPLIFICAR ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS EN UNA PORCION DE LA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS EN LA MUESTRA DE ENSAYO
540
PREPARAR LA BIBLIOTECA DE SECUENCIACION DE ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS AMPLIFICADOS
550
530
PREPARAR LA BIBLIOTECA DE SECUENCIACION DE LA MEZCLA PURIFICADA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
ENRIQUECER LA MEZCLA DE ACIDOS NUCLEICOS COMBINANDO LA BIBLIOTECA DE ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS AMPLIFICADOS
560
SECUENCIAR LA MEZCLA ENRIQUECIDA DE ACIDOS NUCLEICOS
570
DETERMINAR LA FRACCION DE ACIDOS NUCLEICOS FETALES Y LA PRESENCIA I AUSENCIA DE ANEUPLOIDIA EN LA MUESTRA DE ENSAYO
580
500

FIG. 5

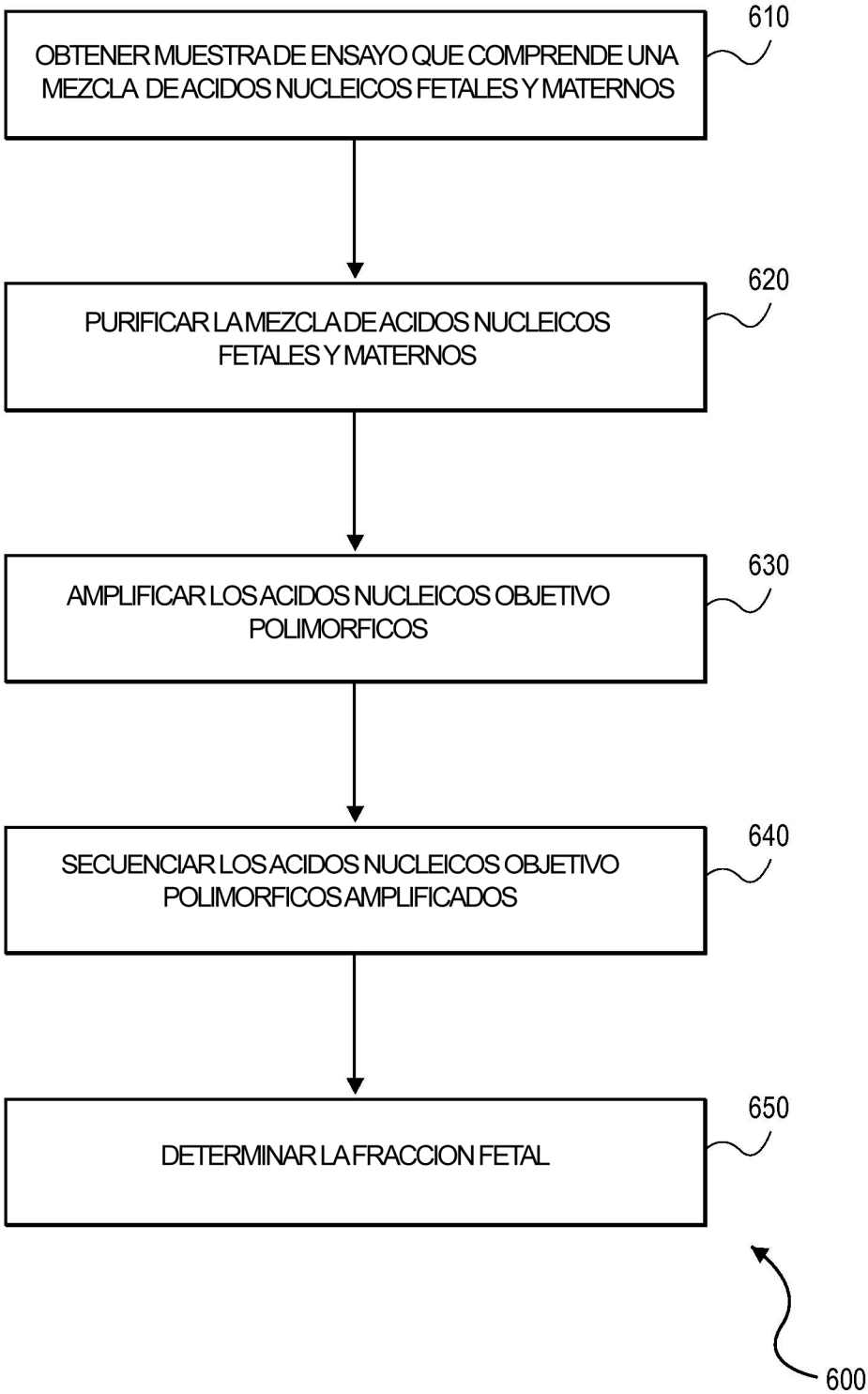
610
OBTENER MUESTRA DE ENSAYO QUE COMPRENDE UNA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
620
PURIFICAR LA MEZCLA DE ACIDOS NUCLEICOS FETALES Y MATERNOS
630
AMPLIFICAR LOS ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS
640
SECUENCIAR LOS ACIDOS NUCLEICOS OBJETIVO POLIMORFICOS AMPLIFICADOS
650
DETERMINAR LA FRACCION FETAL
600

FIG. 6

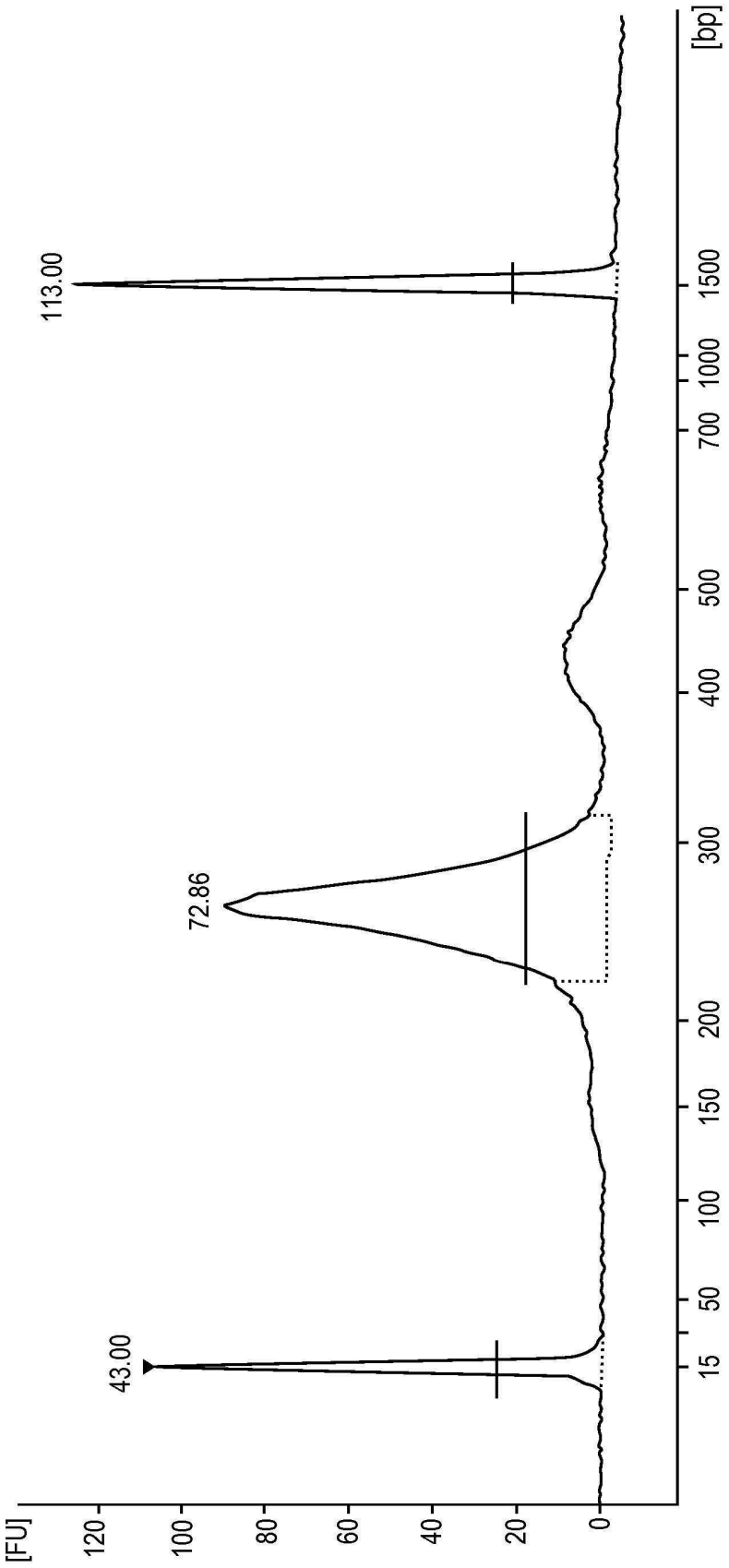
[FU]
120
100
80
60
40
20
0
43.00
72.86
113.00
15
50
100
150
200
300
400
500
700
1000
1500
[bp]

FIG. 7A

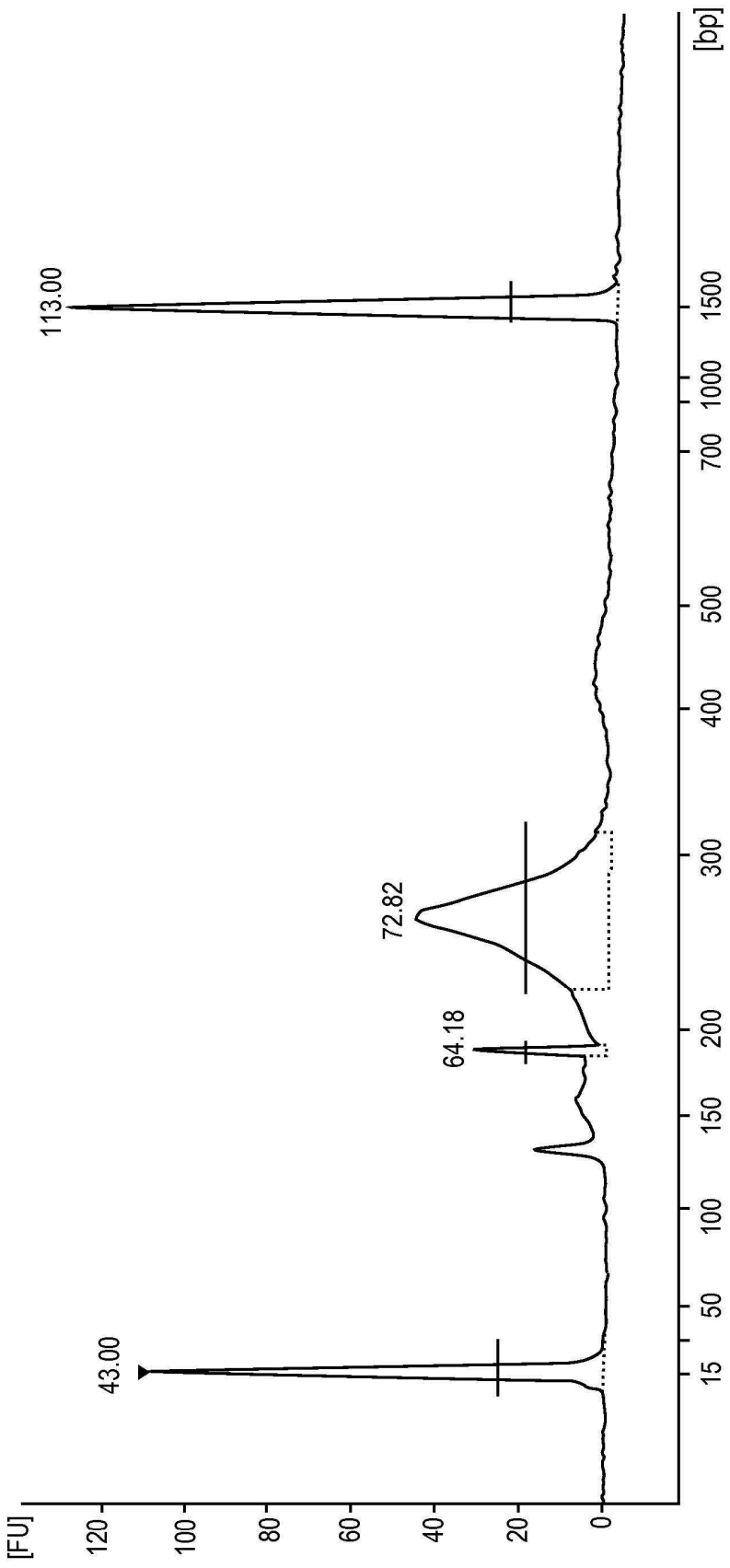
[FU]
120
100
80
60
40
20
0
15
50
100
150
200
300
400
500
700
1000
1500
[bp]
43.00
64.18
72.82
113.00

FIG. 7B

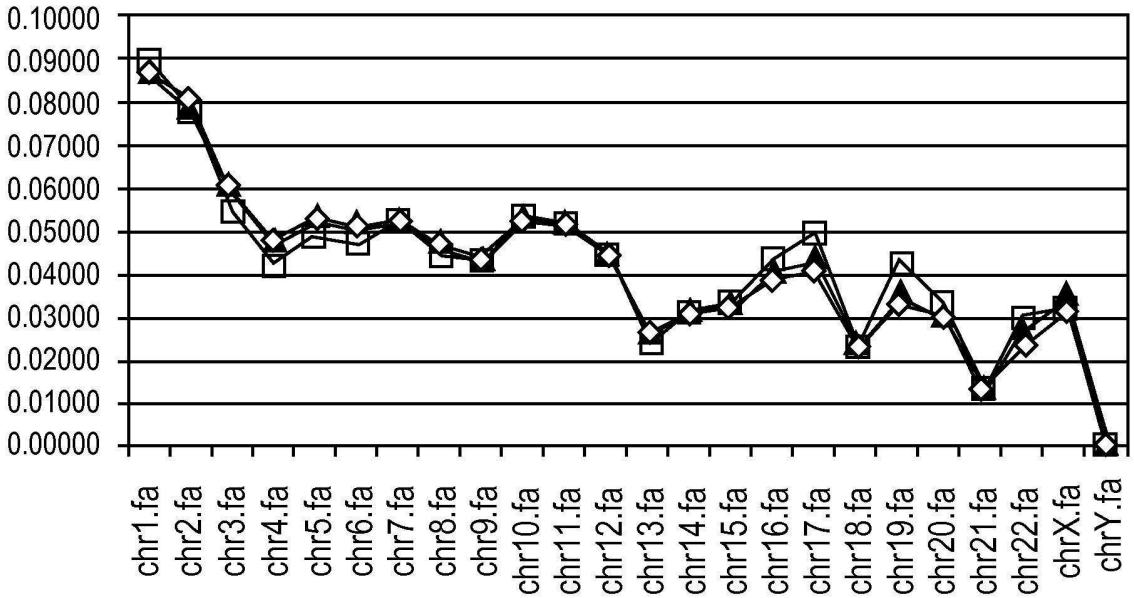
0.10000
0.09000
0.08000
0.07000
0.06000
0.05000
0.04000
0.03000
0.02000
0.01000
0.00000
chr1.fa
chr2.fa
chr3.fa
chr4.fa
chr5.fa
chr6.fa
chr7.fa
chr8.fa
chr9.fa
chr10.fa
chr11.fa
chr12.fa
chr13.fa
chr14.fa
chr15.fa
chr16.fa
chr17.fa
chr18.fa
chr19.fa
chr20.fa
chr21.fa
chr22.fa
chrX.fa
chrY.fa

FIG. 8

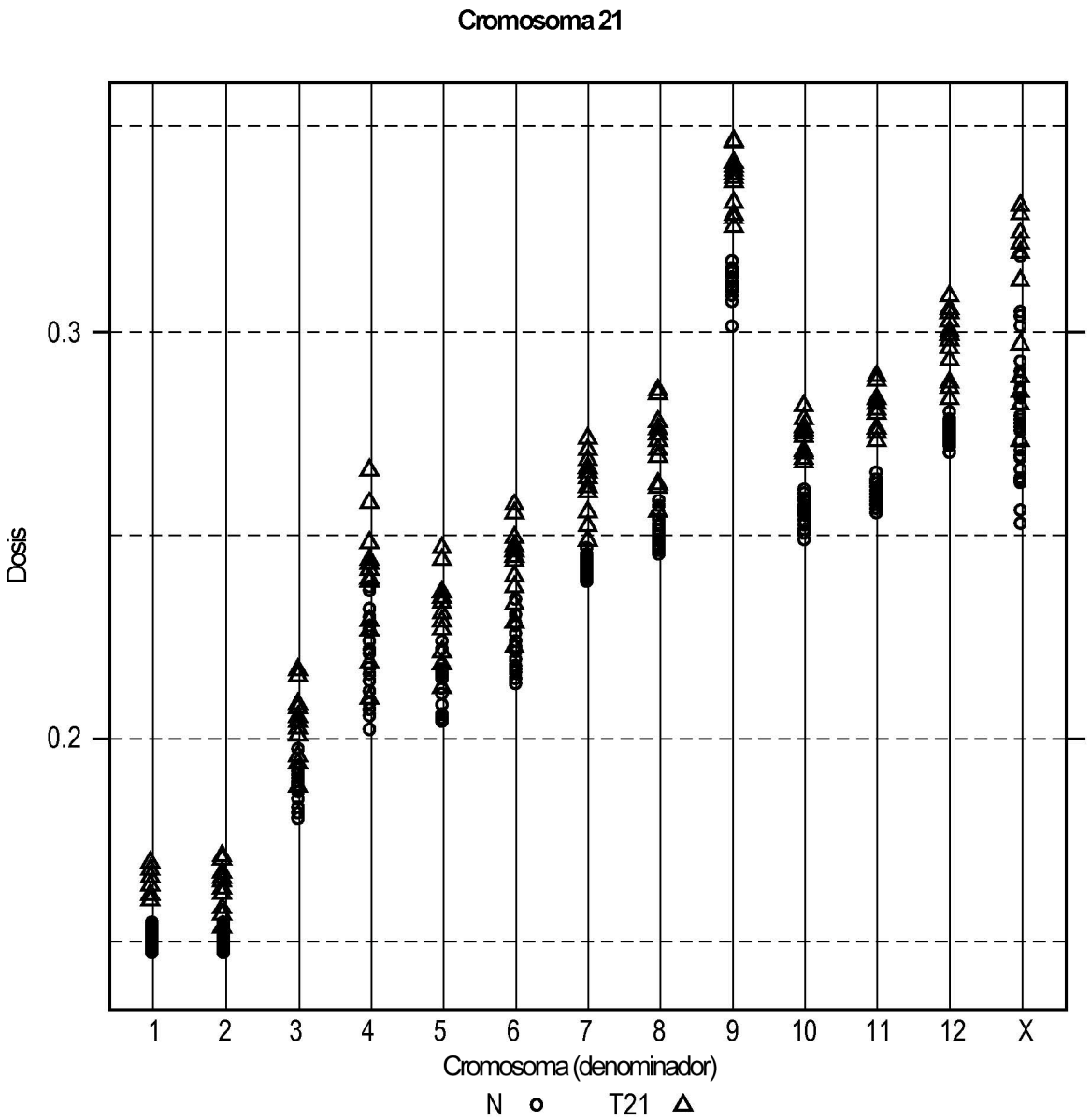
Cromosoma 21
Dosis
0.3
0.2
1
2
3
4
5
6
7
8
9
10
11
12
X
Cromosoma (denominador)
N
T21

FIG. 9A

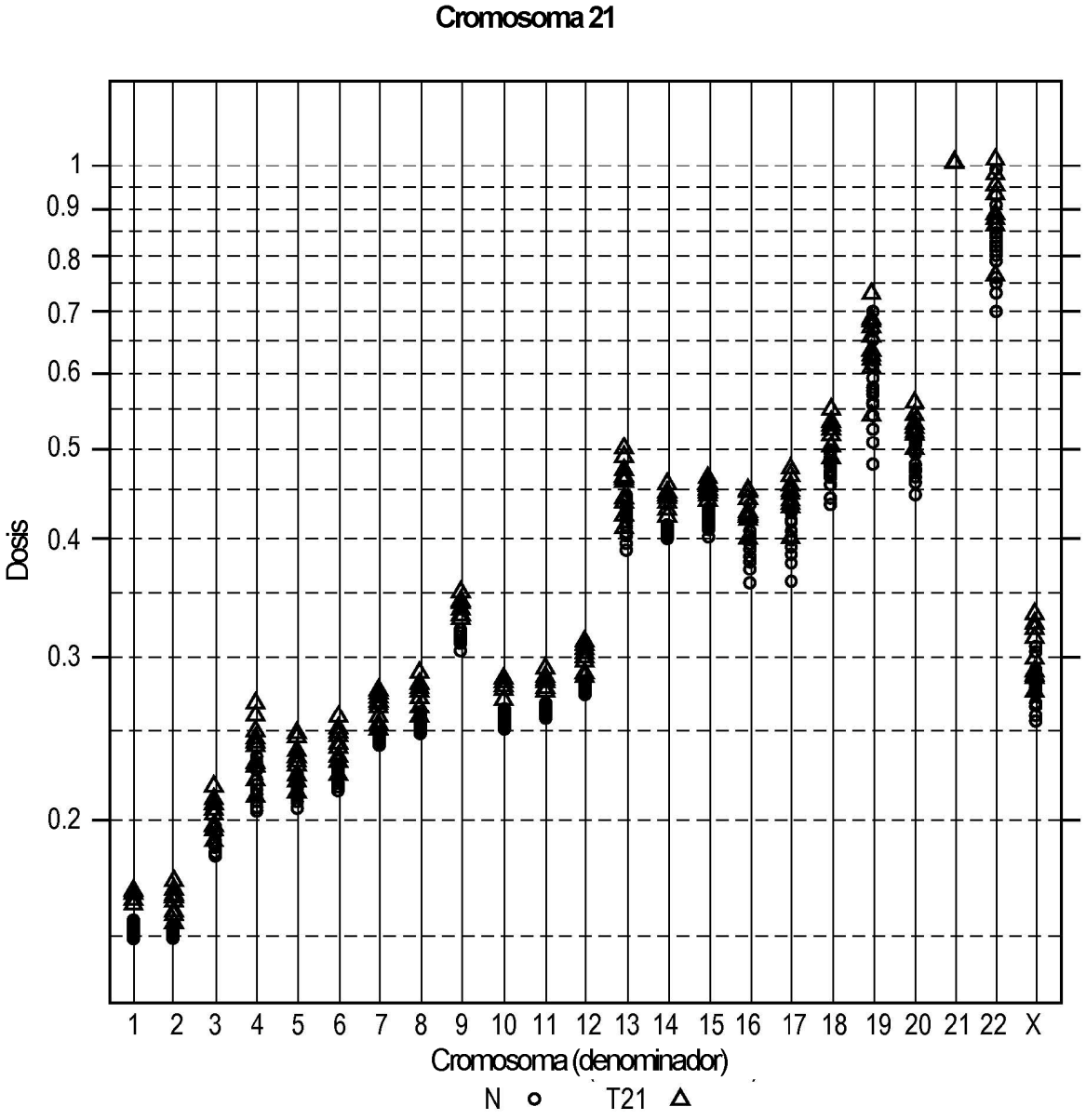
Cromosoma 21
Dosis
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Cromosoma (denominador)
N
T21

FIG. 9B

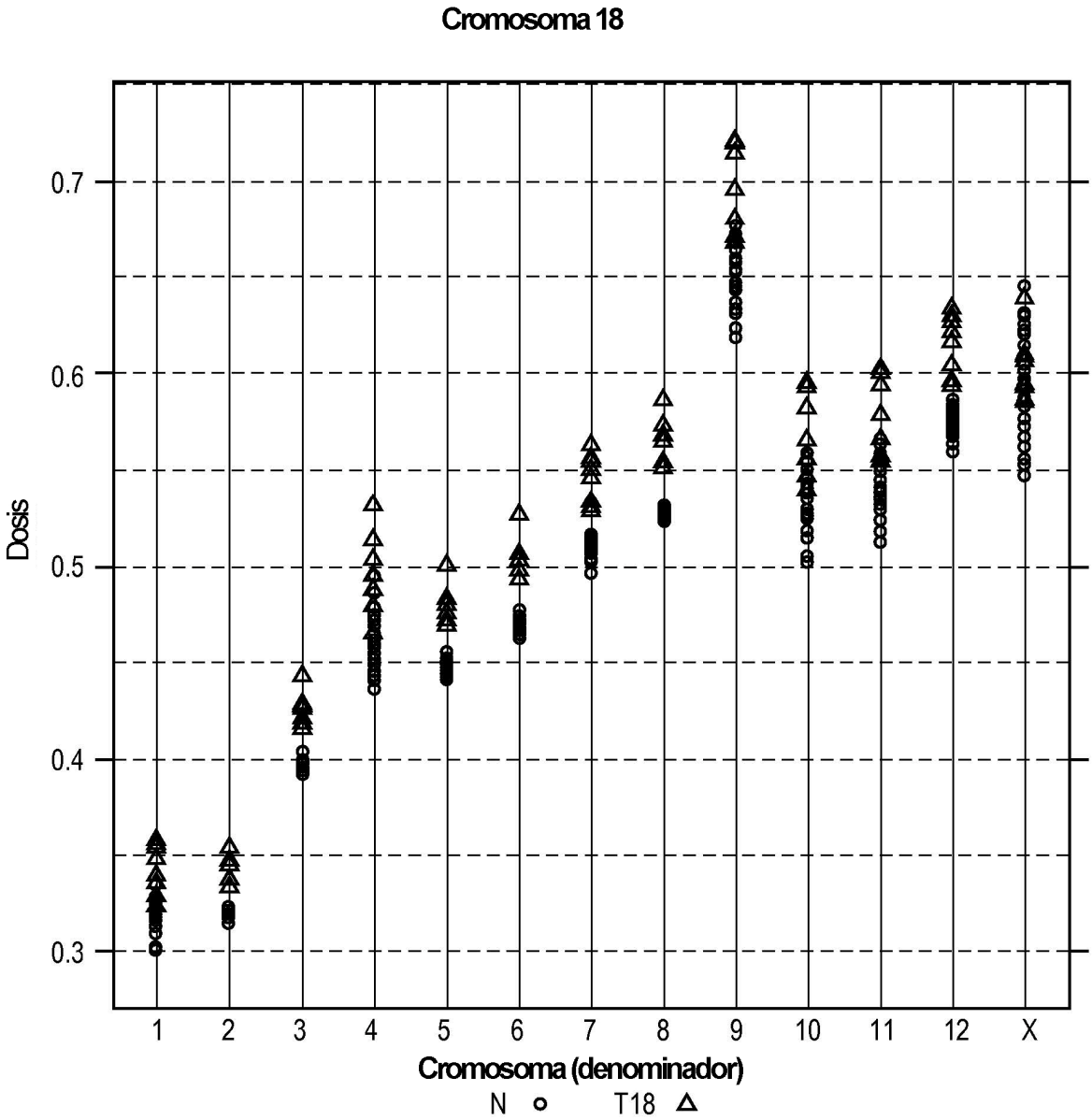
Cromosoma 18
Dosis
0.7
0.6
0.5
0.4
0.3
1
2
3
4
5
6
7
8
9
10
11
12
X
Cromosoma (denominador)
N
T18

FIG. 10A

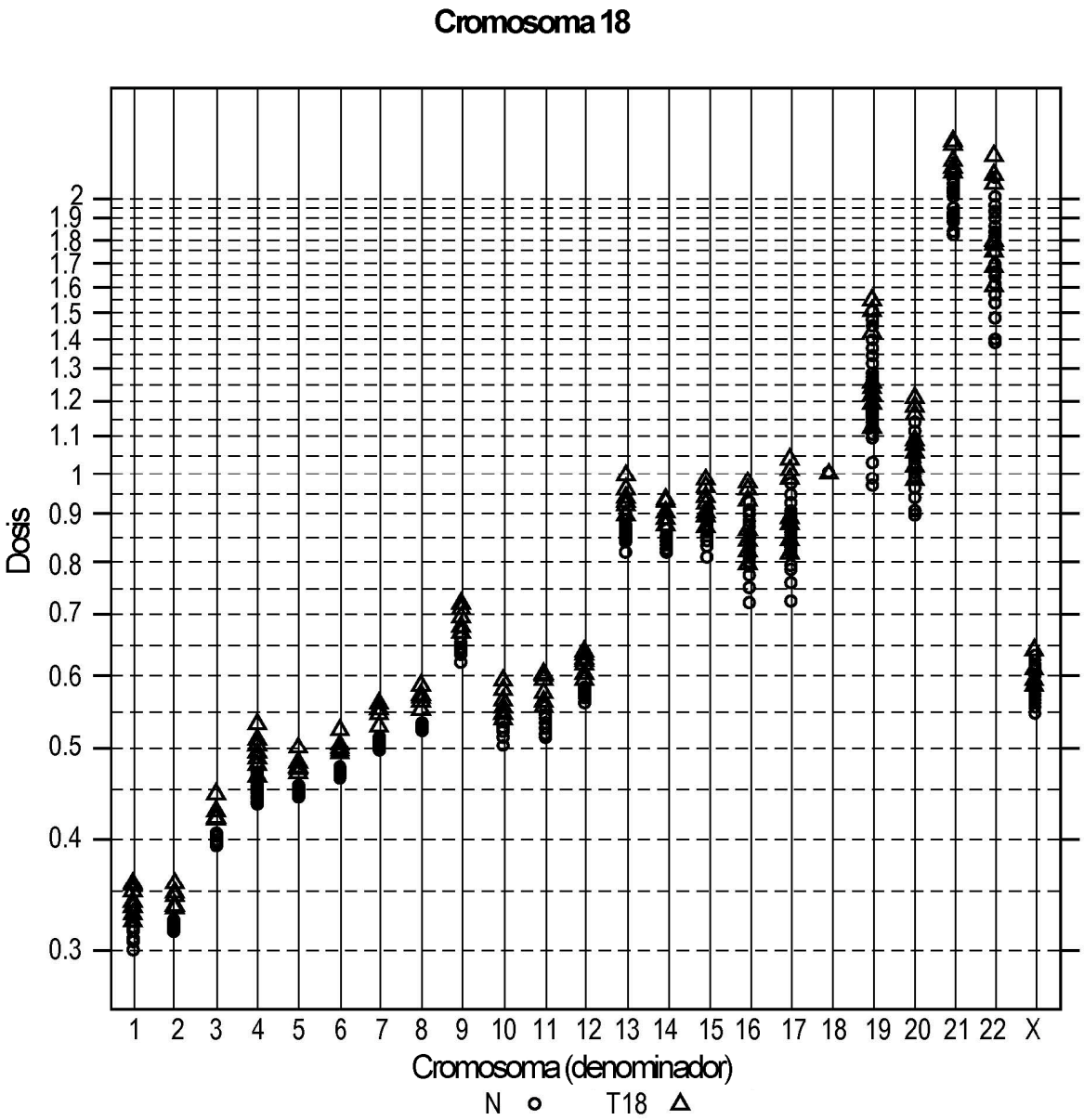
Cromosoma 18
Dosis
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Cromosoma (denominador)
N
T18

FIG. 10B

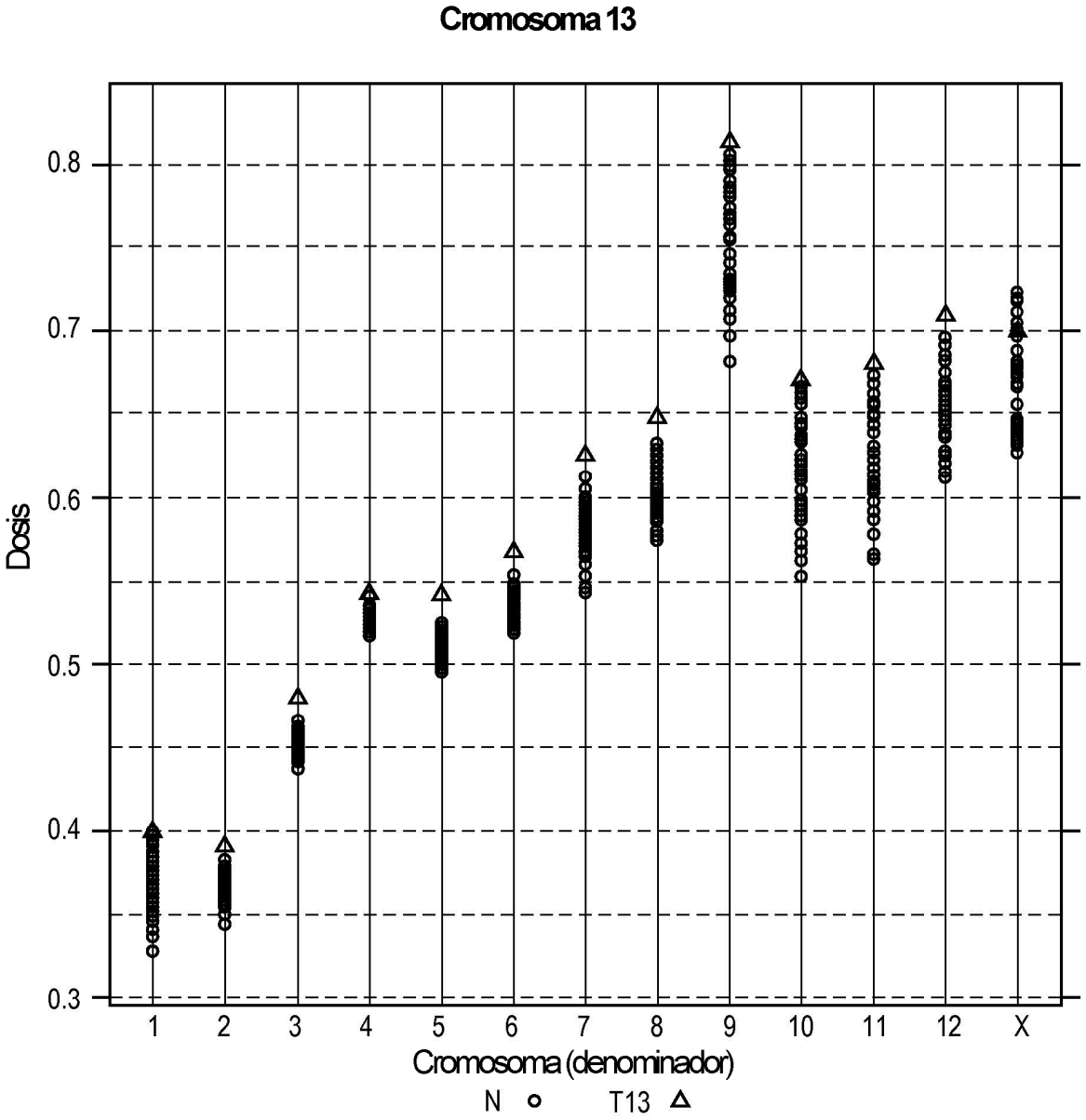
Cromosoma 13
Dosis
0.8
0.7
0.6
0.5
0.4
0.3
1
2
3
4
5
6
7
8
9
10
11
12
X
Cromosoma (denominador)
N
T13

FIG. 11A

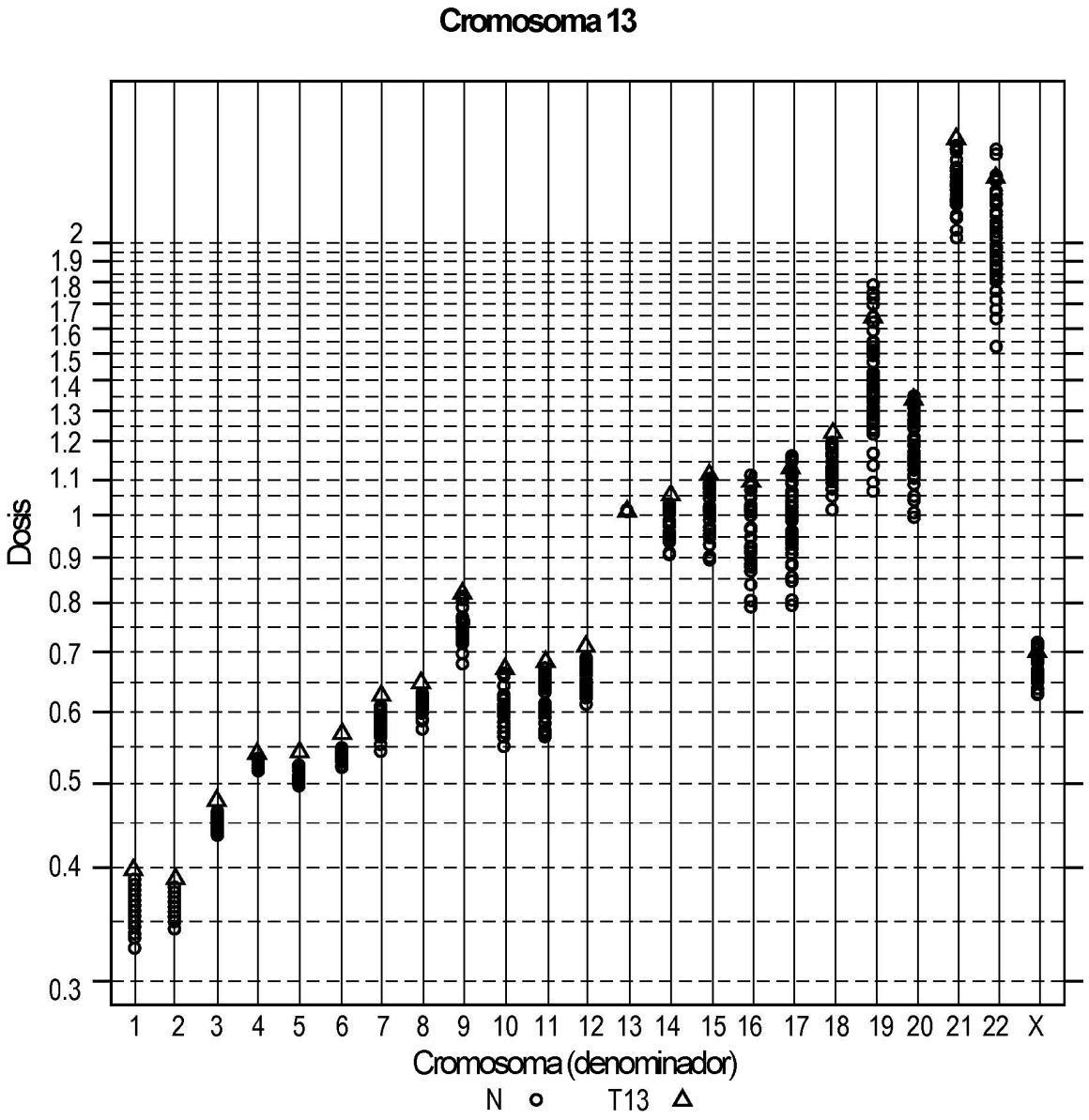
Cromosoma 13
Dosis
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Cromosoma (denominador)
N
T13

FIG. 11B

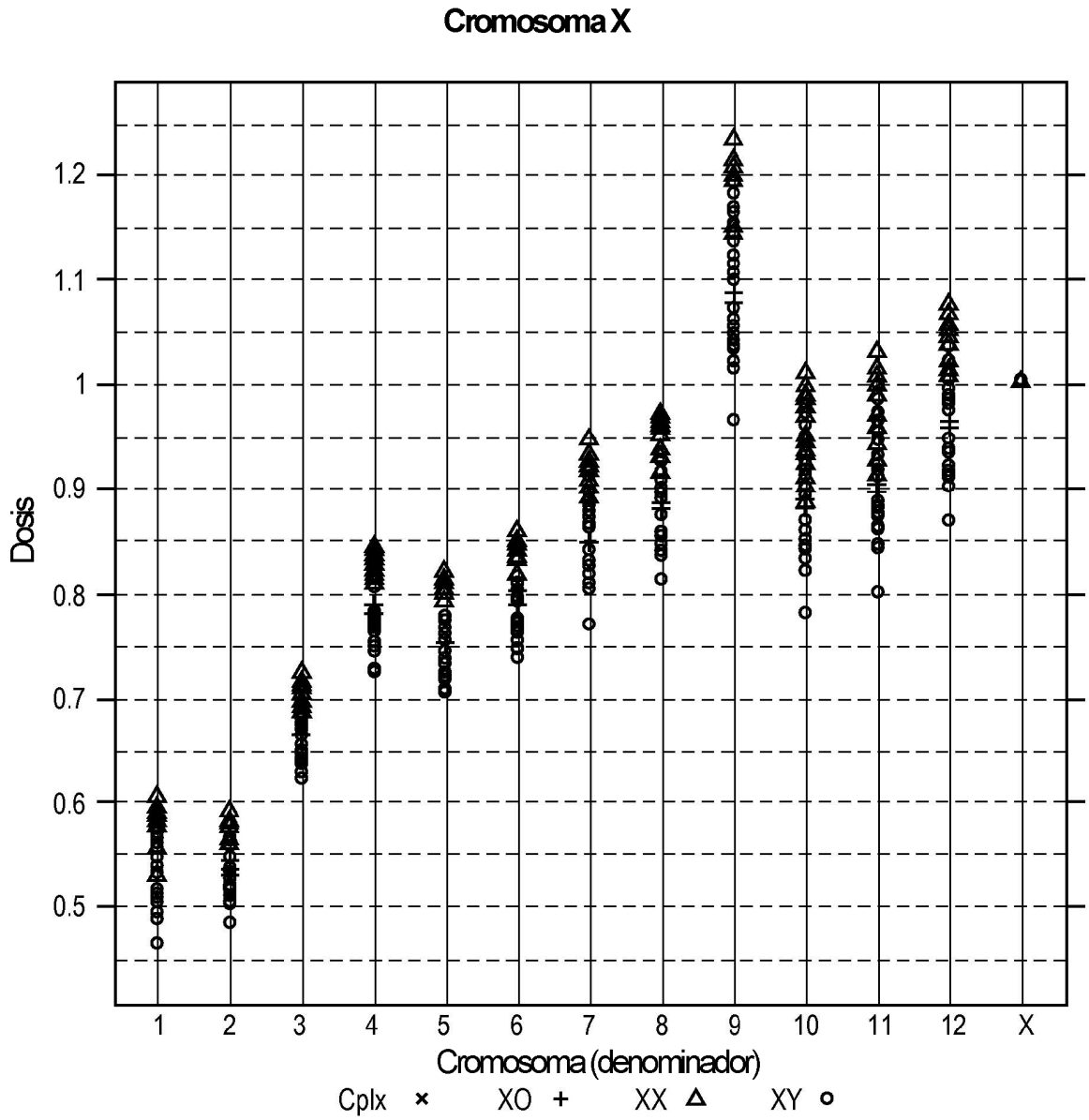
Cromosoma X
Dosis
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
1
2
3
4
5
6
7
8
9
10
11
12
X
Cromosoma (denominador)
Cplx
XO
XX
XY

FIG. 12A

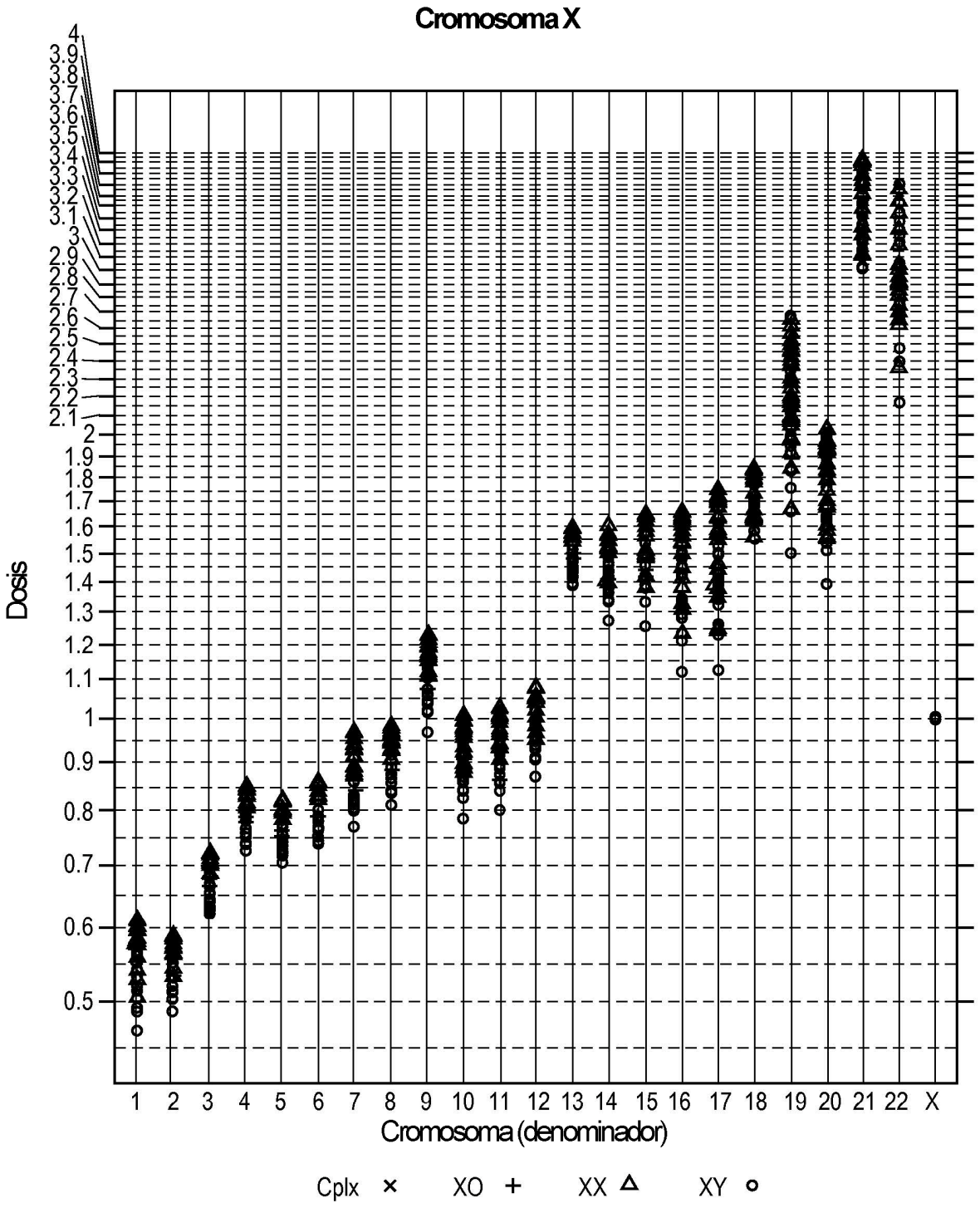
Cromosoma X
Dosis
4
3.9
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Cromosoma (denominador)
Cplx × XO + XX △ XY o

FIG. 12B

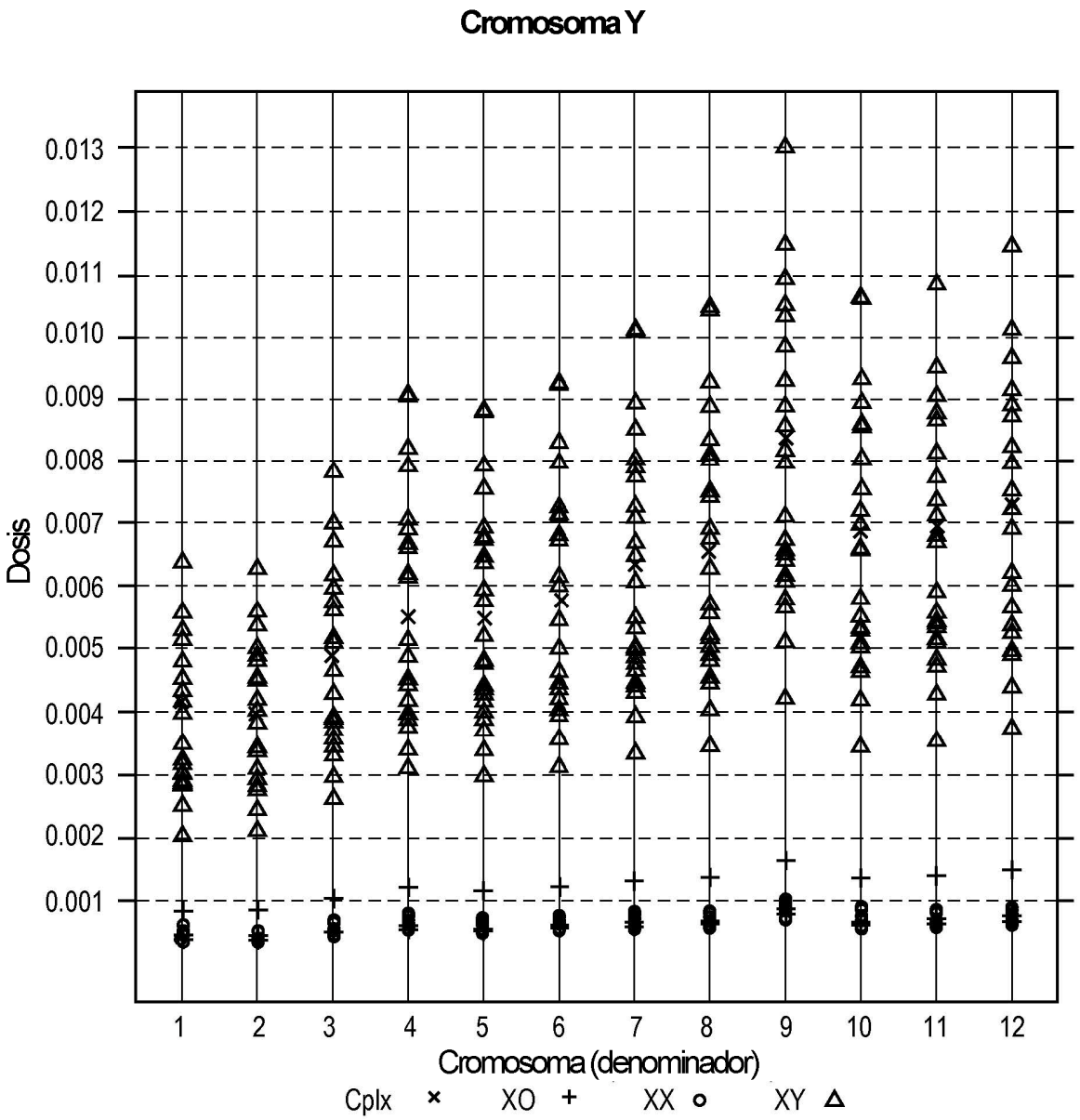
Cromosoma Y
Dosis
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
1
2
3
4
5
6
7
8
9
10
11
12
Cromosoma (denominador)
Cplx
XO
XX
XY

FIG. 13A

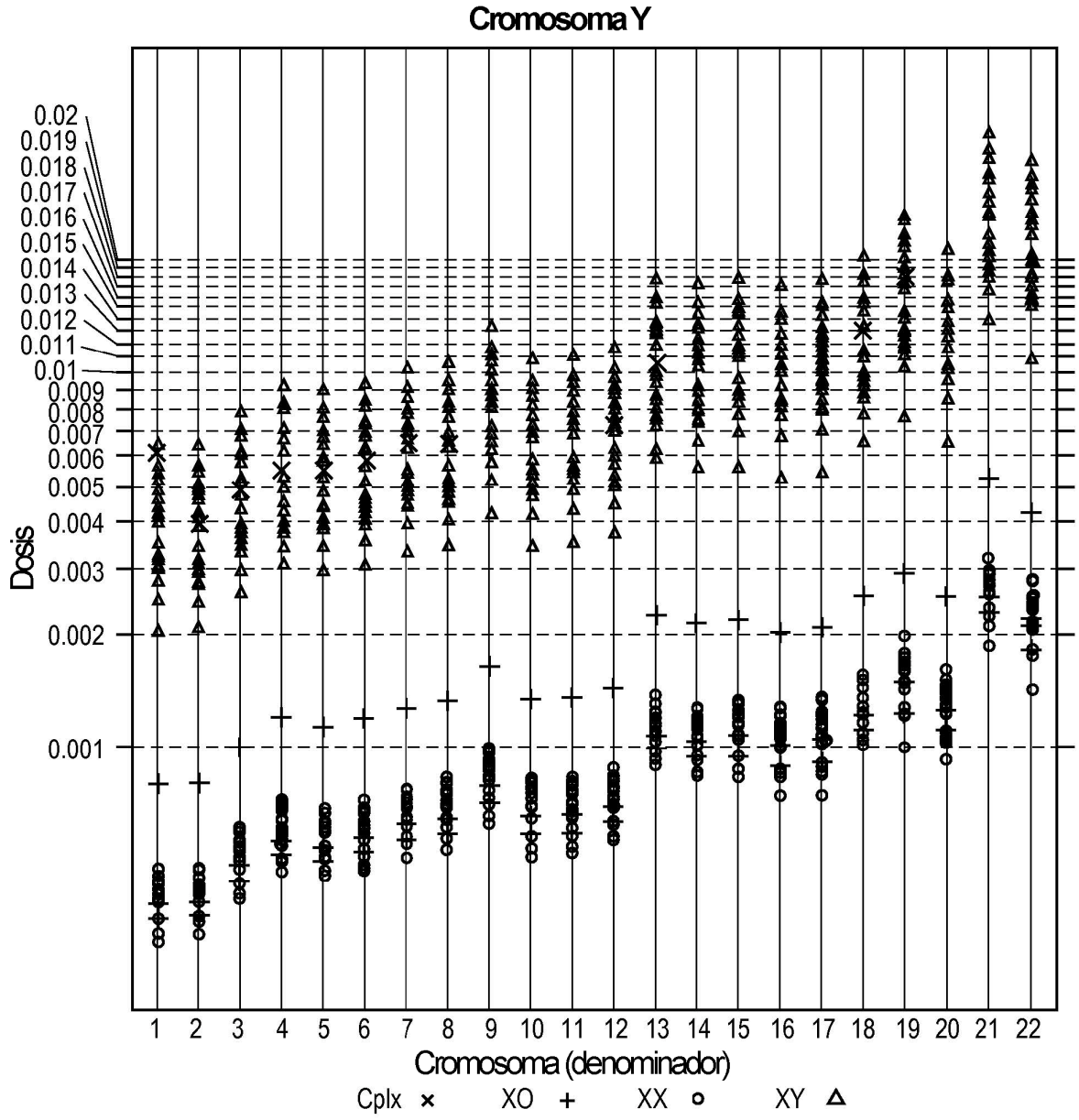
Cromosoma Y
Dosis
0.02
0.019
0.018
0.017
0.016
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Cromosoma (denominador)
Cplx
XO
XX
XY

FIG. 13B

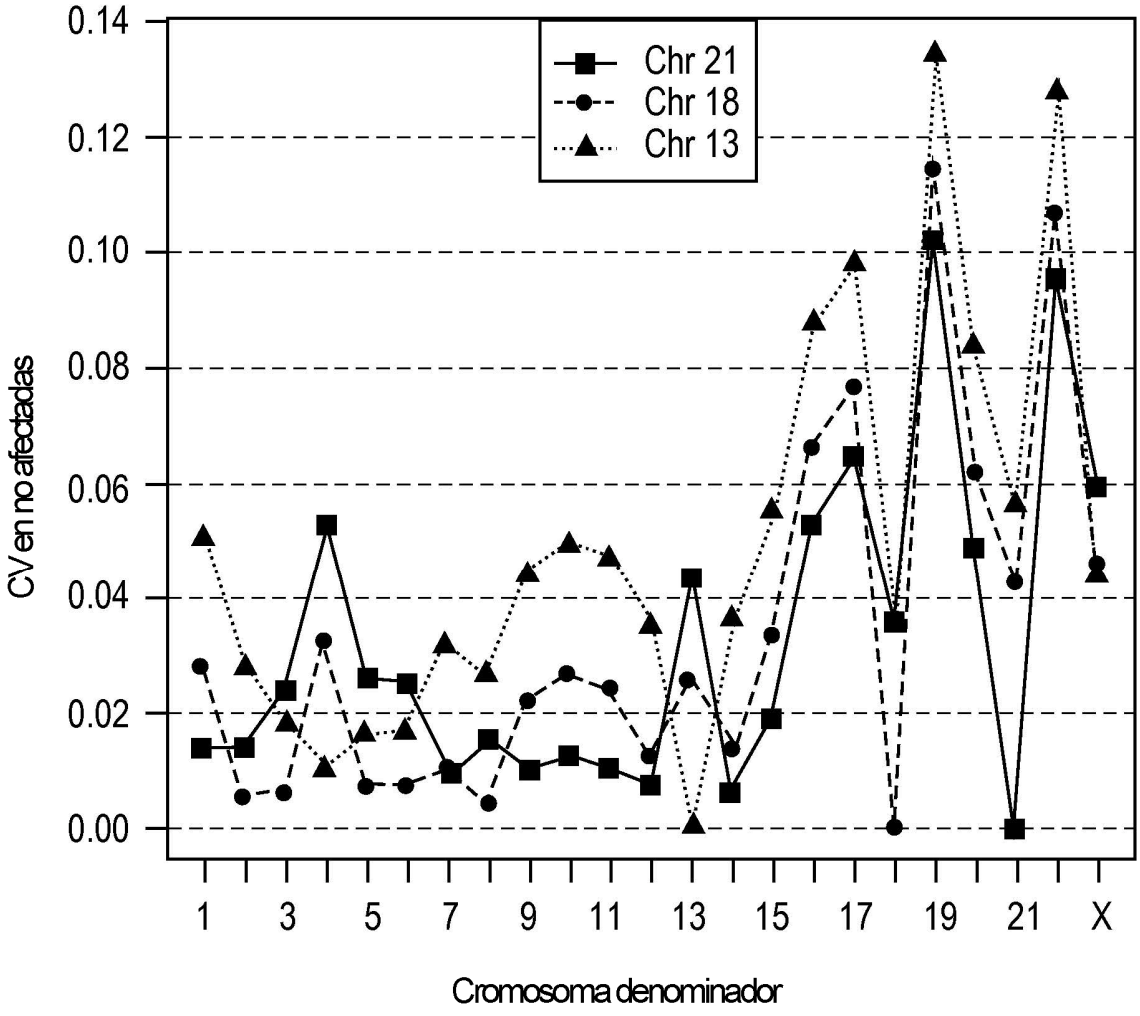
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
CV en no afectadas
Chr 21
Chr 18
Chr 13
1
3
5
7
9
11
13
15
17
19
21
X
Cromosoma denominador

FIG. 14

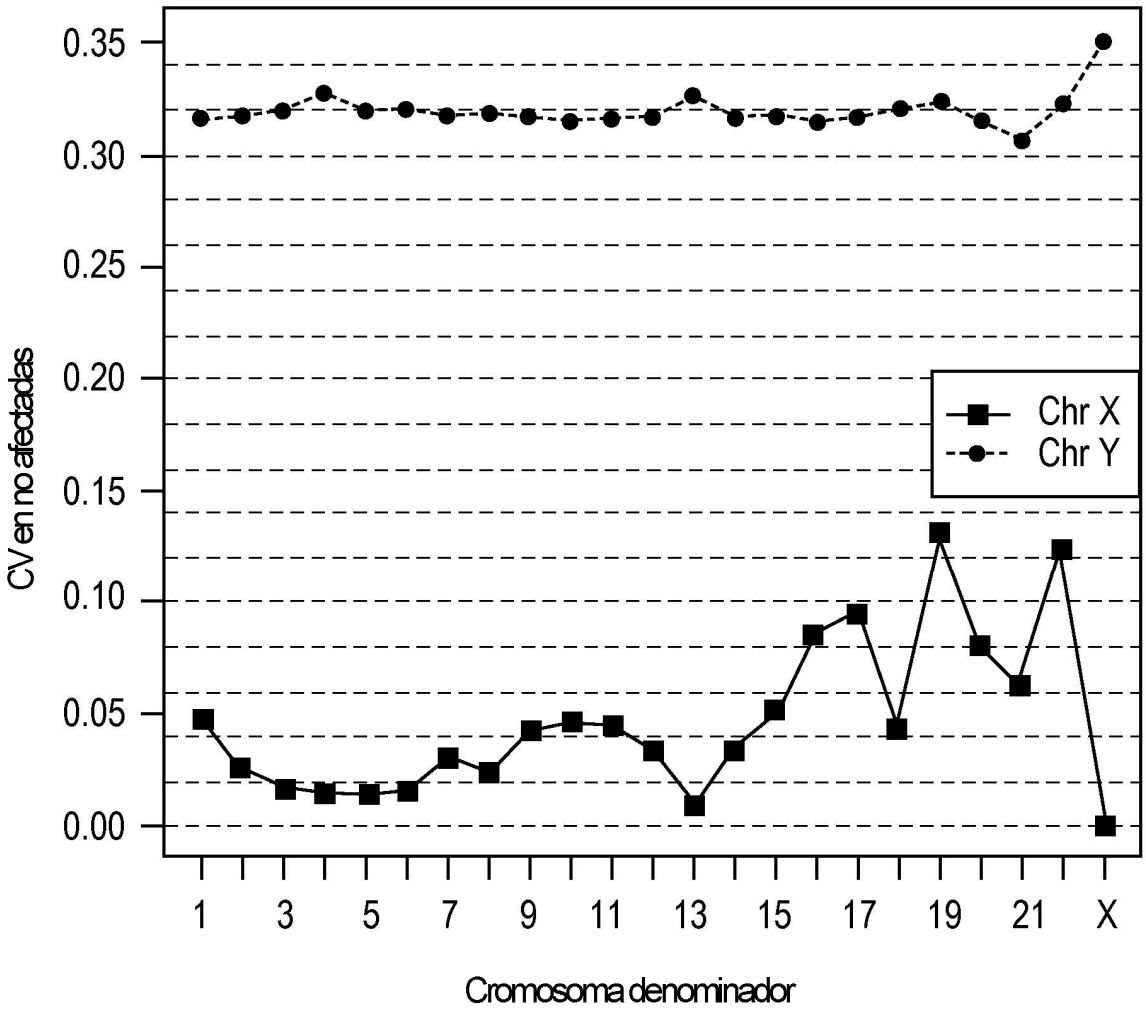
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
CV en no afectadas
Chr X
Chr Y
1
3
5
7
9
11
13
15
17
19
21
X
Cromosoma denominador

FIG. 15

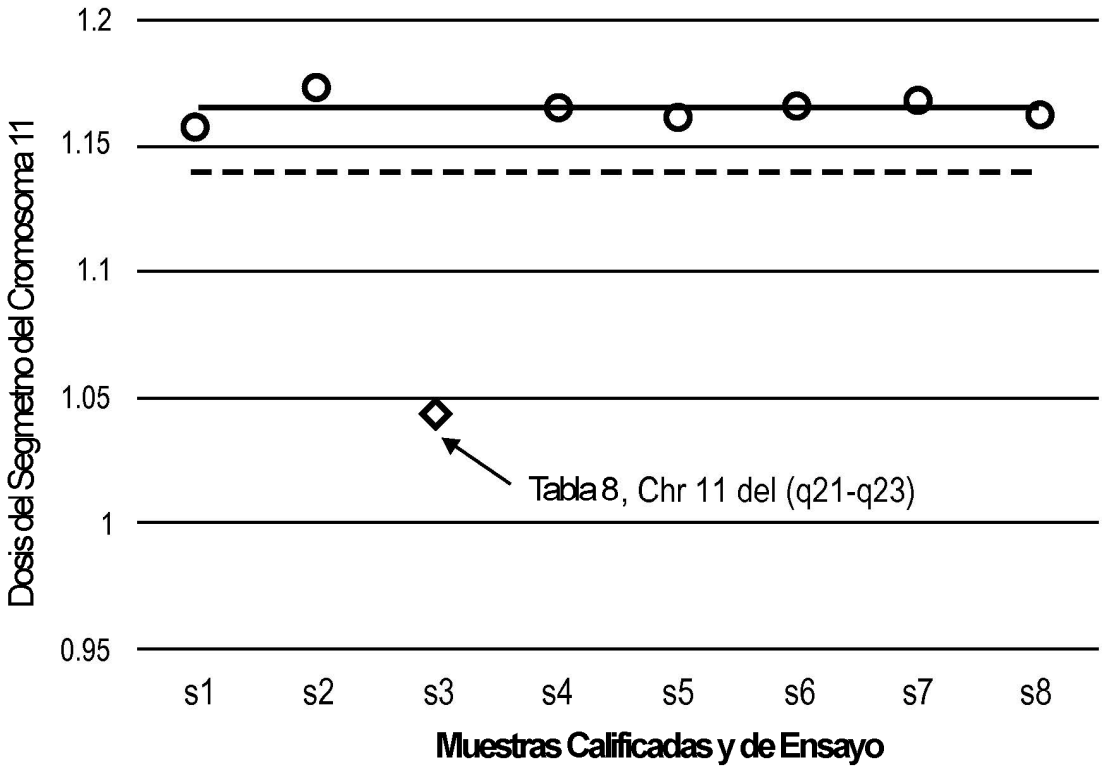
1.2
1.15
1.1
1.05
1
0.95
Dosis del Segmento del Cromosoma 11
Tabla 8, Chr 11 del (q21-q23)
s1
s2
s3
s4
s5
s6
s7
s8
Muestras Calificadas y de Ensayo

FIG. 16

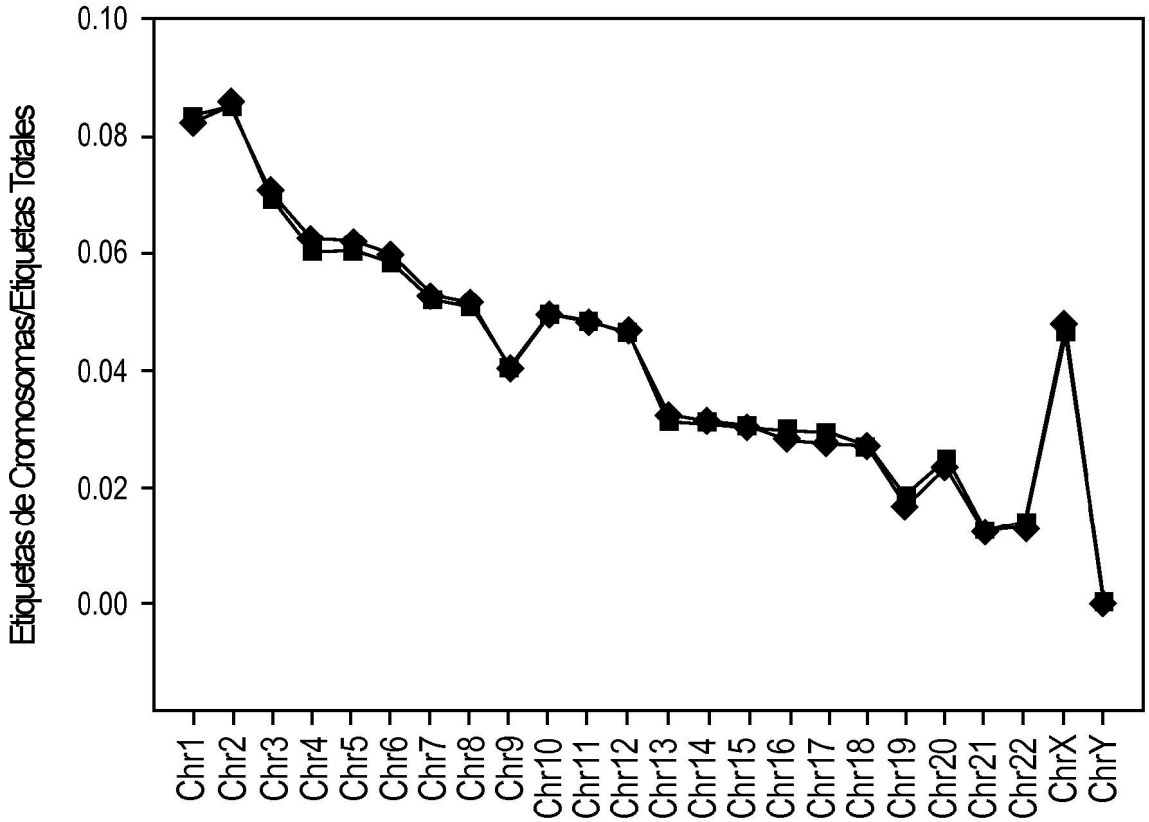
Etiquetas de Cromosomas/Etiquetas Totales
0.10
0.08
0.06
0.04
0.02
0.00
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY

FIG. 17

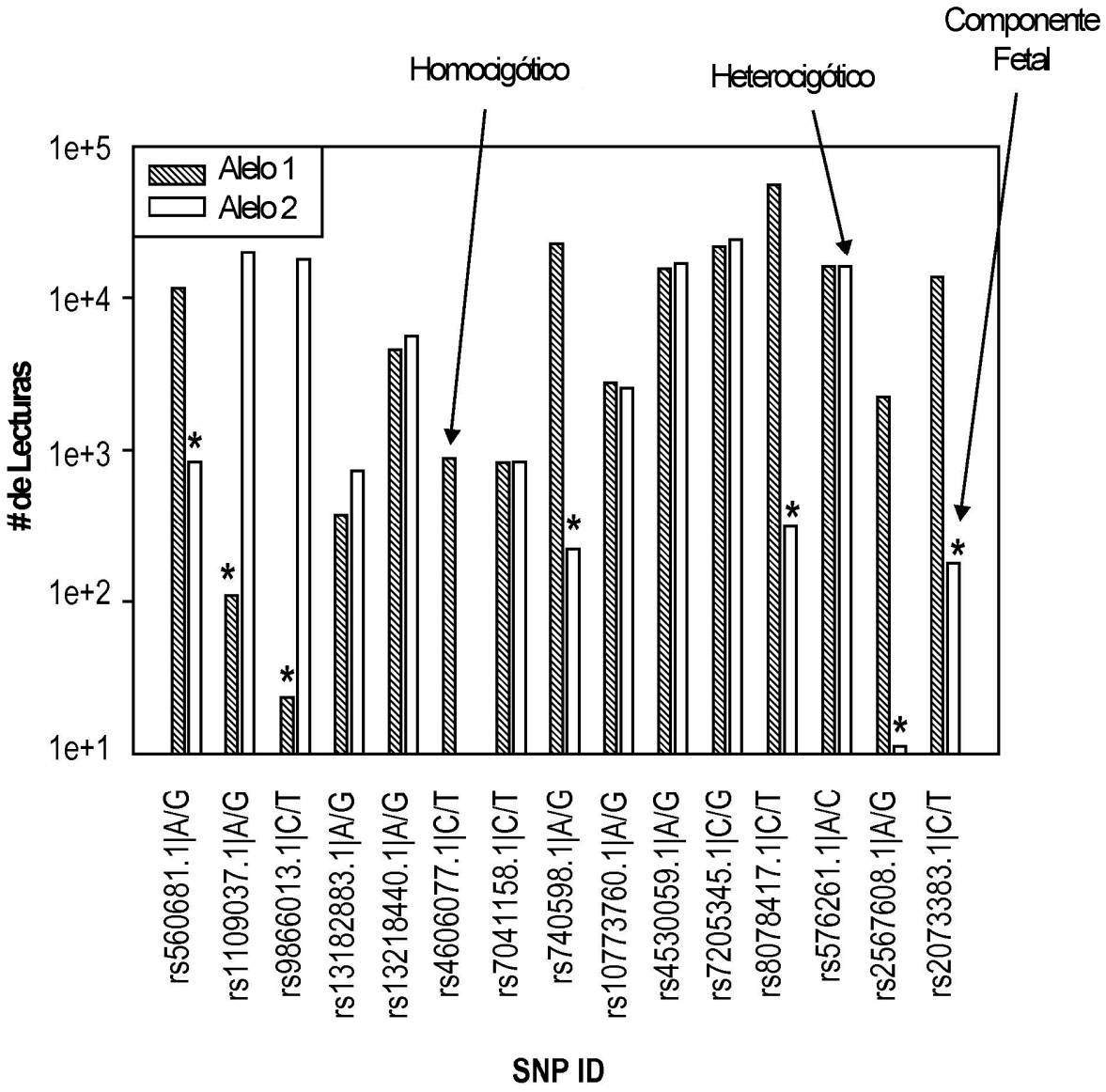
Homocigótico
Heterocigótico
Componente Fetal
Alelo 1
Alelo 2
1e+5
1e+4
1e+3
1e+2
1e+1
de Lecturas
rs560681.1|A/G
rs1109037.1|A/G
rs9866013.1|C/T
rs13182883.1|A/G
rs13218440.1|A/G
rs4606077.1|C/T
rs7041158.1|C/T
rs740598.1|A/G
rs10773760.1|A/G
rs4530059.1|A/G
rs7205345.1|C/G
rs8078417.1|C/T
rs576261.1|A/C
rs2567608.1|A/G
rs2073383.1|C/T
SNP ID

FIG. 18

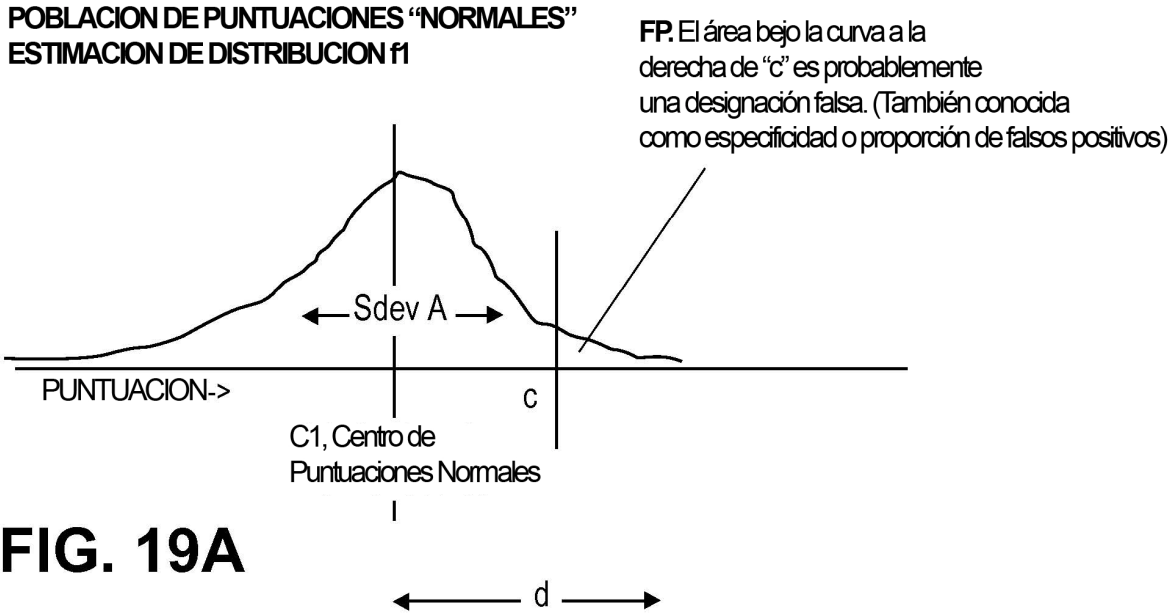

FIG. 19A

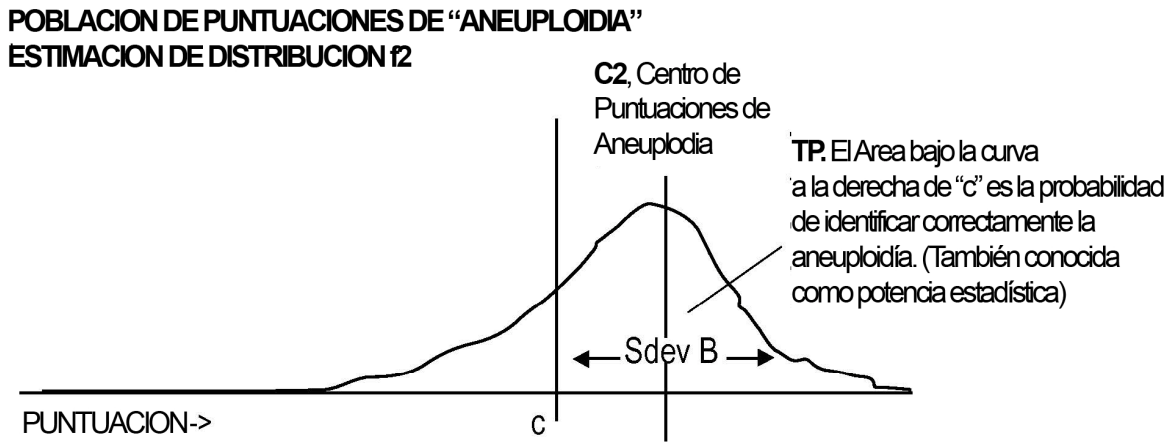

FIG. 19B

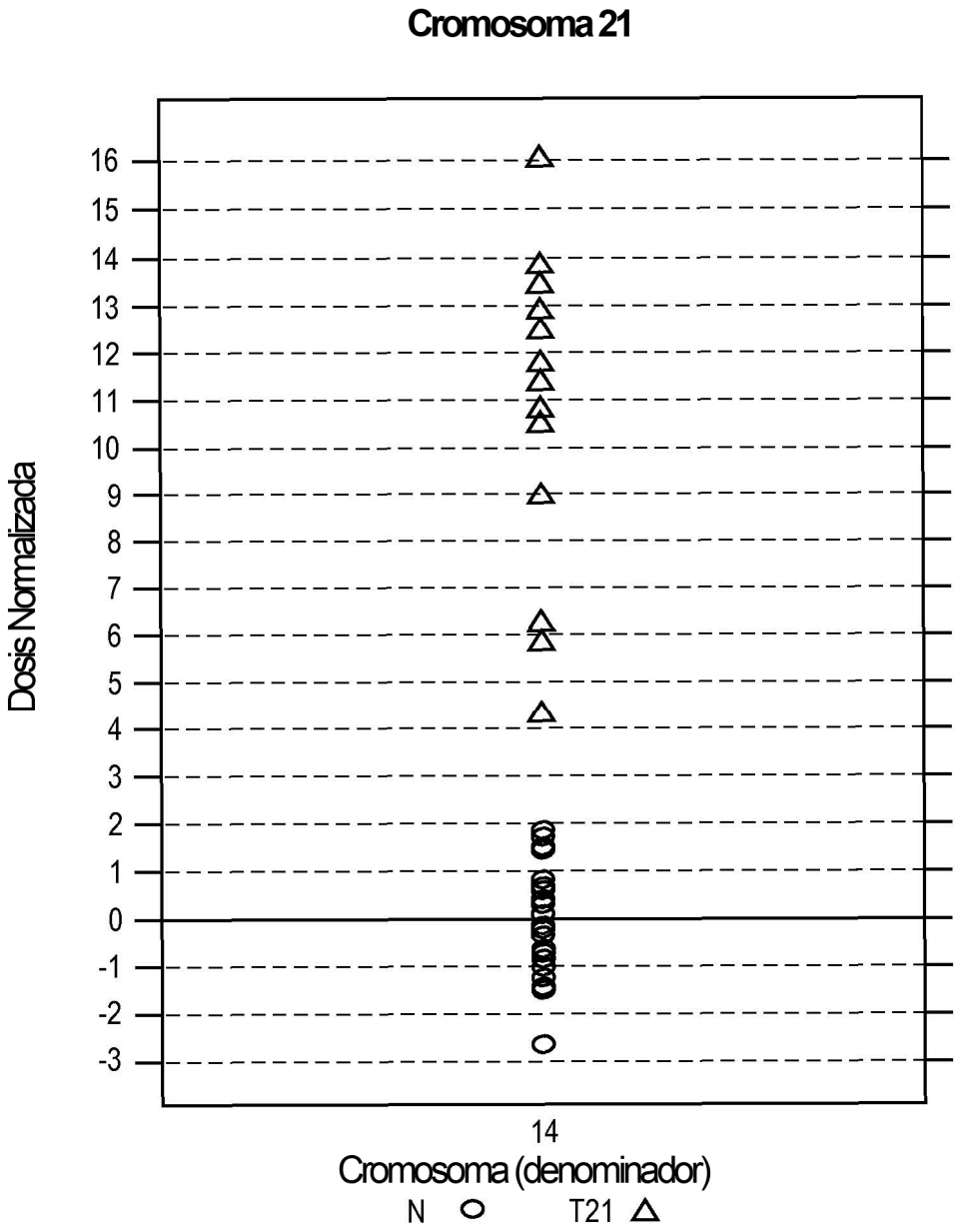
Cromosoma 21
Dosis Normalizada
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
14
Cromosoma (denominador)
N
T21

FIG. 20A

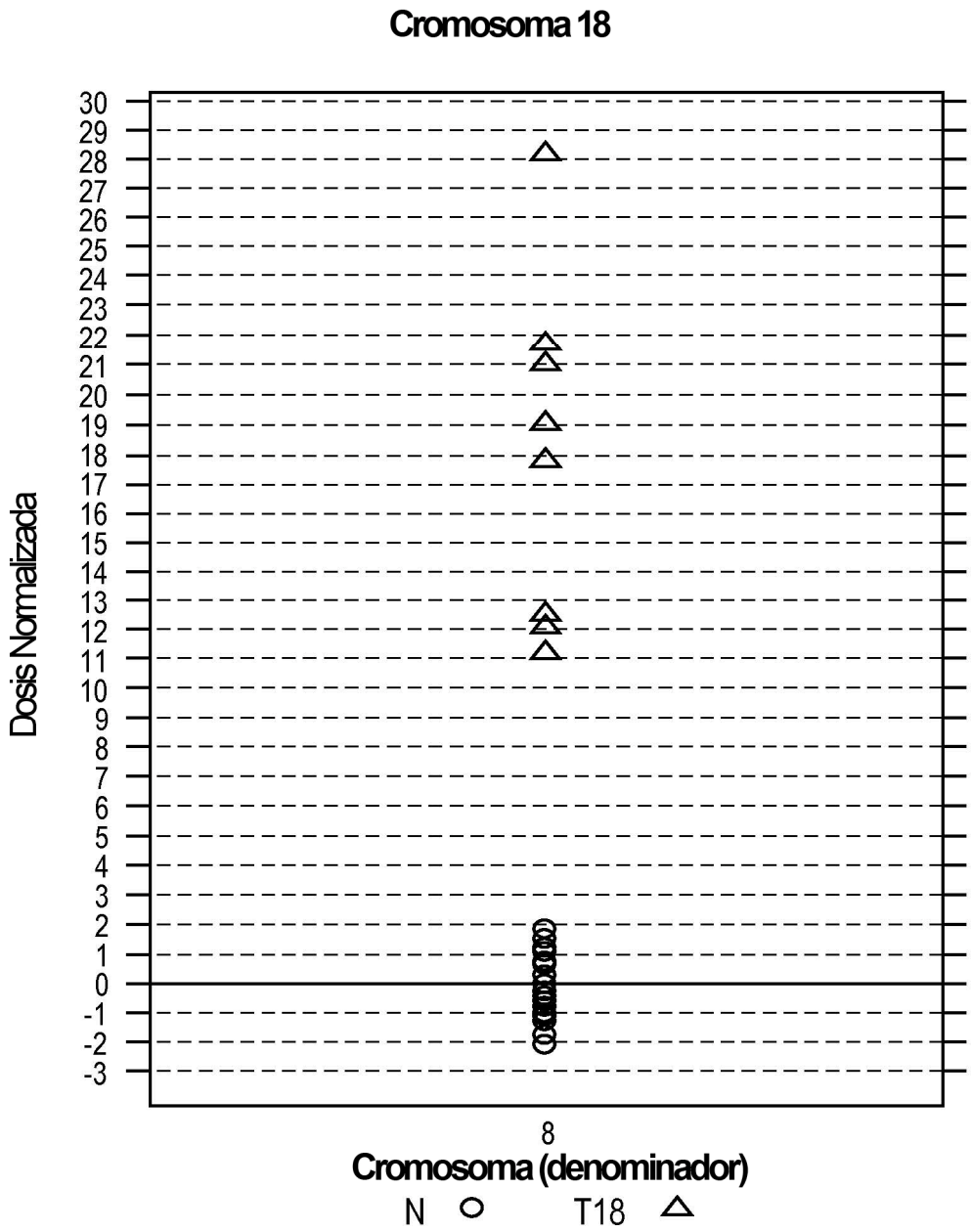
Cromosoma 18
Dosis Normalizada
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
8
Cromosoma (denominador)
N
T18

FIG. 20B

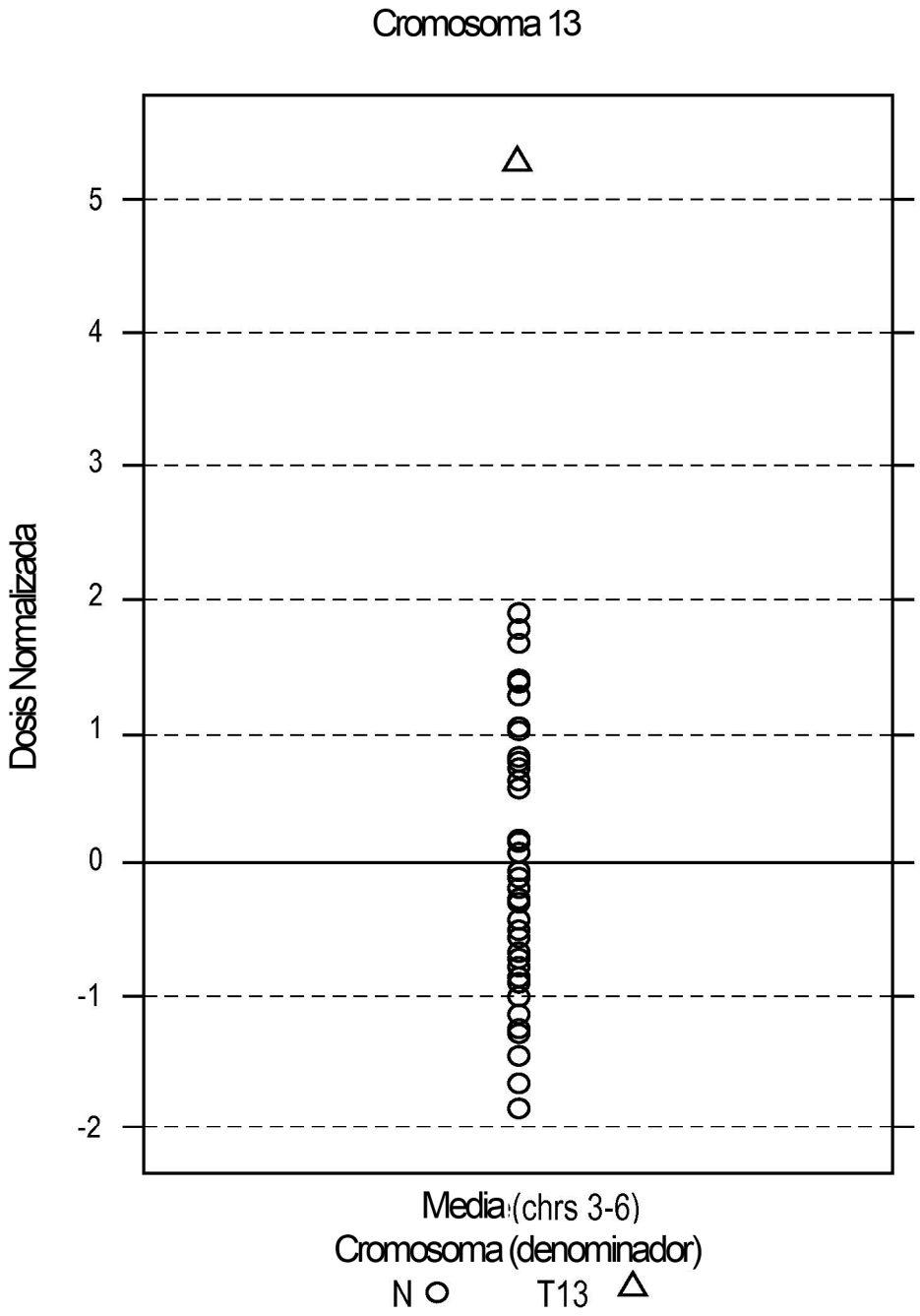
Cromosoma 13
Dosis Normalizada
5
4
3
2
1
0
-1
-2
Media (chrs 3-6)
Cromosoma (denominador)
N
T13

FIG. 20C

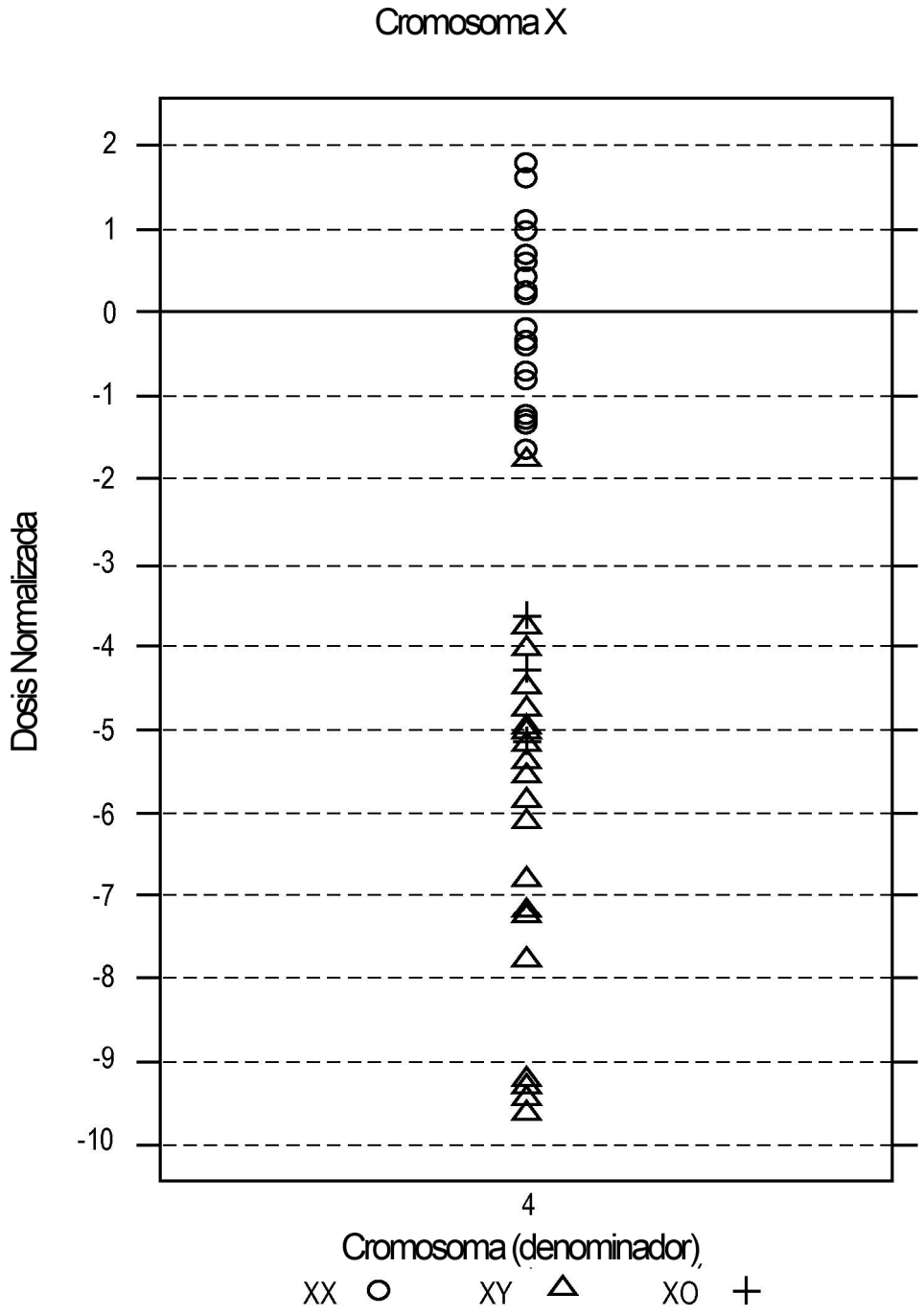
Cromosoma X
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
Dosis Normalizada
4
Cromosoma (denominador)
XX
XY
XO

FIG. 20D

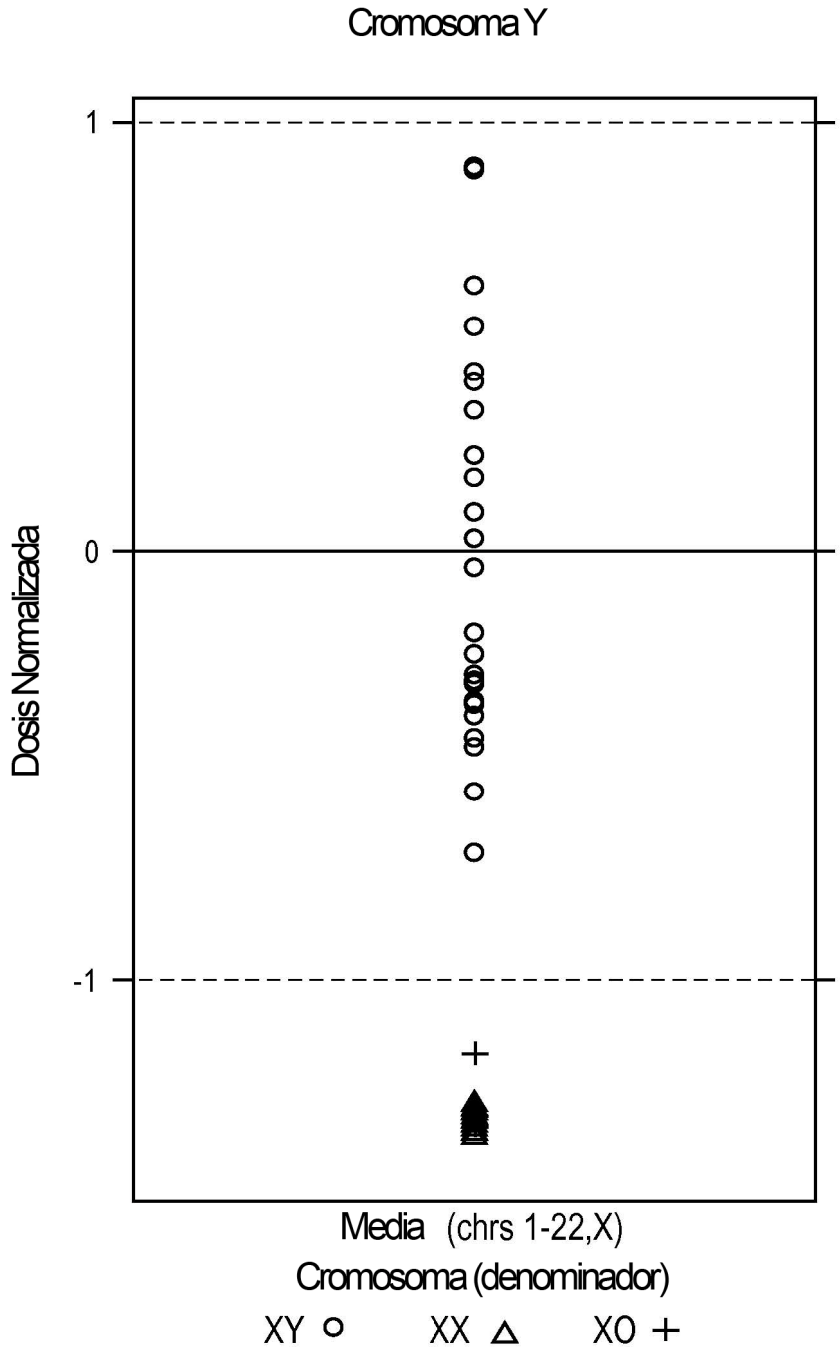
Cromosoma Y
1
0
-1
Dosis Normalizada
Media (chrs 1-22,X)
Cromosoma (denominador)
XY
XX
XO

FIG. 20E