

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 651 612**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.10.2012 PCT/IB2012/002091**

87 Fecha y número de publicación internacional: **25.04.2013 WO13057568**

96 Fecha de presentación y número de la solicitud europea: **18.10.2012 E 12787077 (2)**

97 Fecha y número de publicación de la concesión europea: **22.11.2017 EP 2768978**

54 Título: **Diagnóstico de aneuploidía cromosómica fetal**

30 Prioridad:
18.10.2011 US 201161548632 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.01.2018

73 Titular/es:
**MULTIPLICOM NV (100.0%)
Galileilaan 18
2845 Niel, BE**

72 Inventor/es:
**DEL-FAVERO, JURGEN;
GOOSSENS, DIRK y
HEYRMAN, LIEN**

74 Agente/Representante:
SÁEZ MAESO, Ana

Observaciones :

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 651 612 T3

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

DESCRIPCIÓN

Diagnóstico de aneuploidía cromosómica fetal

Campo de la invención

5 La presente invención se refiere en general a la prueba de diagnóstico de una aneuploidía cromosómica fetal mediante la identificación de una aneuploidía en los cromosomas 13, 18, 21, X y/o Y mediante el ensayo de una muestra materna tal como sangre.

Introducción a la invención

10 La aneuploidía se refiere a una cantidad anormal de cromosomas (o parte de los cromosomas) que es una causa común de defectos de nacimiento. En la aneuploidía, los genes pueden estar presentes en tres copias, "trisomía", o en una sola copia, "monosomía". Estos cambios en el número de cromosomas, que resultan de la ausencia de disyunción de los cromosomas durante la meiosis, tienen efectos dramáticos en las personas afectadas y dan como resultado síndromes bien conocidos. La mayoría de las trisomías y monosomías son letales para el feto y provocan abortos espontáneos o la muerte inmediatamente después del nacimiento. Algunas aneuploidías, sin embargo, son viables y dan lugar a síndromes. Las aneuploidías más frecuentes entre los nacidos vivos son los cromosomas 21, 18, 13 de trisomía y un número distorsionado de cromosomas sexuales. La aneuploidía autosómica más común con la que los bebés pueden sobrevivir es la trisomía 21 (síndrome de Down), que afecta a 1 de cada 800 nacimientos. La trisomía 18 (síndrome de Edwards) afecta a 1 de cada 6.000 nacimientos y la trisomía 13 (síndrome de Patau) afecta a 1 de cada 10.000 nacimientos. La aneuploidía del cromosoma sexual (SCA) afecta a 1 de cada 400 recién nacidos y, por lo tanto, es más común que el síndrome de Down en general. Mientras que SCA incluye una variedad de anomalías de los cromosomas sexuales, con mucho, el SCA más común es la eliminación del cromosoma X (45, síndrome de X-Turner) o la adición de un cromosoma X o Y (47, síndrome XXY-Klinefelter, 47, XYY, 47, XXX). De estas condiciones, solo el síndrome de Turner da como resultado un fenotipo físico fácilmente identificable. Sin embargo, el lenguaje sutil y las dificultades de aprendizaje se han identificado en la mayoría de las formas de SCA. El factor de riesgo más importante para la aneuploidía es la edad materna, ya que la mayoría de los niños con aneuploidía nacen de madres mayores de 35 años, por lo que la prevalencia aumenta a medida que más mujeres eligen o necesitan retrasar la maternidad. Los programas contemporáneos de detección prenatal suelen incluir las aneuploidías cromosómicas fetales comunes 21, 18 y 13. El riesgo de embarazo se evalúa de varias maneras. Para las aneuploidías cromosómicas, se han implementado pruebas de detección no invasivas basadas en ultrasonografía y la medición de marcadores en suero materno para identificar embarazos de alto riesgo en los primeros 3 meses de embarazo (11-14 semanas). El sonograma mide el líquido debajo de la piel a lo largo de la parte posterior del cuello del bebé, llamado translucencia nucal (NT). El sonograma también determinará si el hueso nasal del bebé está presente o ausente. Una muestra de sangre materna se usa para analizar dos marcadores séricos llamados beta-gonadotropina coriónica humana (hCG) y proteína A plasmática asociada al embarazo (PAPP-A), que se encuentran en la sangre de todas las mujeres embarazadas. En los embarazos con aneuploidía hay líquido extra detrás del cuello del bebé y/o los resultados de hCG y PAPP-A son más altos o más bajos que el promedio. Además, el hueso nasal de un bebé puede estar ausente en algunos embarazos con una anomalía cromosómica.

40 La combinación del riesgo relacionado con la edad con la medición NT, datos de hueso nasal y marcadores sanguíneos proporciona una cifra de riesgo para el síndrome de Down y una cifra de riesgo para trisomía 13 o trisomía 18. La tasa de detección del primer trimestre es de aproximadamente 90% con un 5% falso tasa positiva para embarazos en los que el bebé tiene síndrome de Down, y es algo mayor para los embarazos con trisomía 13 o trisomía 18. Se puede realizar una ecografía de translucencia nucal sin medir hCG y PAPP-A. En este caso, sin embargo, la tasa de detección de aneuploidía se reduce a aproximadamente 70%.

45 El diagnóstico prenatal es una parte integral de la práctica obstétrica. Para realizar un diagnóstico genético prenatalmente, se requiere material genético del feto no nacido. Convencionalmente, el ADN fetal se muestrea mediante procedimientos invasivos como la amniocentesis o el muestreo de vellosidades coriónicas. Estos procedimientos están asociados con un riesgo de aborto espontáneo de 0.5 y 1-2% respectivamente. Por lo tanto, es una rutina reservar los procedimientos de diagnóstico invasivo para los embarazos que se estima que tienen un alto riesgo de aneuploidía, que representan alrededor del 5-10% de todas las mujeres examinadas para el riesgo de aneuploidía. Teniendo en cuenta los riesgos relacionados con el procedimiento del diagnóstico prenatal convencional, sería ideal si el análisis genético del feto se pudiera realizar de forma no invasiva. Para realizar el diagnóstico prenatal no invasivo, se requiere una fuente de material genético fetal sin dañar al feto. Un avance importante para este fin fue informado por Lo et al. (1997)¹ y WO98/39474 que describen la existencia de ADN fetal flotante libre en plasma materno. Posteriormente mostraron que el ADN derivado del feto contribuyó con ~10% del ADN que flota libremente en el plasma materno. El ADN fetal se puede detectar en el plasma materno solo semanas después de la concepción y se elimina rápidamente del plasma materno y desaparece unas horas después del parto. Como resultado, el ADN fetal flotante libre en el plasma materno es una fuente prometedora de material genético fetal para el desarrollo de una prueba prenatal no invasiva. Sin embargo, el ADN fetal representa solo una fracción menor del total de ADN flotante libre en plasma con la porción restante de ADN aportada por la madre, derivada principalmente de los glóbulos blancos maternos.

Dado el enorme potencial, se han descrito varias metodologías no invasivas para la detección de aneuploidías en la última década. Un método es centrarse en el análisis de moléculas de ácido nucleico que son específicas del feto en el plasma materno y, por lo tanto, superar la interferencia causada por el ADN materno de fondo. Uno podría tener como objetivo la detección de ARNm expresado en la placenta o de firmas epigenéticas específicas de la placenta que se originan en el cromosoma de interés. En una serie de desarrollos desde el año 2000, se estableció la base del ARN plasmático como herramienta de diagnóstico prenatal. Poon et al. (2000)² demostraron que el ARNm transcrito del cromosoma Y podría detectarse en el plasma de mujeres portadoras de fetos masculinos. Más tarde, se demostró que la placenta es una fuente importante de ARN derivado del feto en el plasma materno usando ARNm de lactógeno placentario humano y ARNm que codifica la subunidad beta de la gonadotropina coriónica humana como ejemplos³.

En 2007, se identificó un ARNm específico de la placenta, transcrito de un gen localizado en el cromosoma 21, PLAC4, utilizando un enfoque basado en microarreglos y se demostró que era detectable en el plasma materno y se despejó después del parto del feto⁴. Para determinar la dosificación del cromosoma 21 utilizando ARNm de PLAC4 en el plasma materno, se utilizó el enfoque de relación alélica ARN-SNP. Este método se basa en la presencia de un SNP en la región codificante del gen PLAC4. Si un feto es heterocigoto para este SNP, posee dos alelos que se distinguen por secuencia de ADN. Si el feto es euploide, la proporción de estos dos alelos SNP es 1:1. Por el contrario, si el feto tiene trisomía 21, entonces la relación alélica ARN-SNP se convertiría en 1:2 o 2:1. Lo et al. (2007)⁴ demostraron que esta estrategia podría aplicarse para determinar no invasivamente el estado de trisomía del cromosoma 21 de un feto. De forma similar, el enfoque de ARN-SNP también se aplicó para la detección no invasiva de trisomía 18 mediante el análisis de la relación alélica de SERPINB2 mRNA⁵.

Sin embargo, la principal limitación del enfoque de proporción alélica de ARN-SNP es que solo los fetos heterocigotos para el SNP analizado pueden diagnosticarse con éxito. Por ejemplo, con el uso del único SNP en PLAC4, se espera que aproximadamente el 45% de los fetos sean heterocigotos y, por lo tanto, diagnosticables con este enfoque. En consecuencia, se necesitan varios marcadores para una cobertura de diagnóstico completa. Con este fin, varios investigadores han descrito nuevos marcadores polimórficos de SNP que pueden analizarse utilizando este enfoque. Un informe preliminar describe diez marcadores con una tasa combinada de heterocigotidad que cubre hasta el 95% de la población general de los Estados Unidos. La evaluación de estos marcadores en ensayos clínicos a gran escala se espera para los próximos años.

Se han investigado también las firmas epigenéticas específicas de la placenta, como la metilación del ADN, que se origina en el cromosoma fetal de interés. Como los tejidos en el cuerpo tienen diferentes perfiles de expresión génica, el estado de metilación de ciertos genes también exhibe patrones específicos de tejido. La evidencia muestra que el ADN fetal en el plasma materno se origina en la placenta y que el fondo del ADN materno se deriva de las células sanguíneas de la madre. Por lo tanto, una forma de desarrollar marcadores epigenéticos de ADN fetal es identificar genes cuyo estado de metilación difiera entre los tejidos placentarios y las células sanguíneas de la madre. Chim et al. (2005)⁷ estudiaron el perfil de metilación del promotor SERPINB5 (maspin) y demostraron que estaba hipometilado en los tejidos de la placenta, pero hipermetilado en las células sanguíneas de la madre. Utilizando la PCR específica de metilación, el SERPINB5 hipometilado derivado de placenta podría detectarse y distinguirse de las moléculas hipermetiladas derivadas de la madre en el plasma materno. Esto hizo que SERPINB5, ubicado en el cromosoma 18, fuera el primer marcador de ADN fetal circulante universal que pudiera usarse para todos los embarazos independientemente del sexo y el genotipo fetal. Dado que el gen SERPINB5 se encuentra en el cromosoma 18, permitió el desarrollo de una estrategia que es análoga al enfoque de proporción alélica ARN-SNP, el denominado enfoque de relación alélica epigenética. Por lo tanto, si un feto es heterocigoto para un SNP ubicado en la región promotora de SERPINB5, la medición de la relación de los alelos SNP en la versión hipometilada del gen permite la determinación del estado de trisomía 18 del feto.

Sin embargo, la PCR específica de metilación requiere el uso de una conversión de bisulfito, que altera las citosinas no metiladas a los nucleótidos de uracilo. Pero la conversión de bisulfito degrada hasta el 95% de las moléculas de ADN en una muestra y, por lo tanto, reduce sustancialmente la cantidad de ADN fetal en una muestra de plasma materno y puede dar como resultado una detección falsamente negativa. En consecuencia, los investigadores desarrollaron marcadores epigenéticos fetales que podrían detectarse en el plasma materno sin la necesidad de conversión de bisulfito. Con este fin, Chan et al. (2006)⁸ utilizaron el promotor de RASSF1, ubicado en el cromosoma 3, que está hipermetilado en los tejidos de la placenta pero hipometilado en las células sanguíneas de la madre. En consecuencia, las secuencias de RASSF1 hipometiladas derivadas de las células de la sangre materna se pueden eliminar del plasma materno usando digestión con enzimas de restricción sensibles a la metilación. De hecho, después de la digestión con enzimas de restricción, las secuencias RASSF1 fetales podrían detectarse en el plasma materno antes del parto, pero desaparecieron por completo del plasma materno dentro de las 24 horas después del parto. Chan et al. (2006)⁹ utilizaron el patrón de metilación diferencial del promotor RASSF1 como el control positivo para la detección del ADN fetal en un grupo sanguíneo prenatal fetal rhesus D no invasivo tipificado para 54 mujeres con mal pronóstico RhD temprano.

El enfoque de relación alélica ARN-SNP y el enfoque de metilación del ADN se dirigen a subconjuntos de moléculas de ácido nucleico presentes en el plasma materno de una manera molecular. Una alternativa es usar métodos físicos que den como resultado el enriquecimiento relativo del ADN fetal presente en el plasma materno.

Recientemente⁹, se demostró que la longitud del ADN fetal flotante libre en el plasma materno es -20 pb más corto que el ADN libre flotante derivado de la madre. Por lo tanto, los métodos de fraccionamiento de tamaño tales como la electroforesis en gel permiten el fraccionamiento de tamaño del ADN del plasma y el enriquecimiento de los fragmentos de ADN fetal más cortos. Este enfoque se ha utilizado con éxito para enriquecer el ADN fetal flotante libre. Si bien este enfoque ha demostrado empíricamente ser útil para la detección cualitativa de mutaciones causantes de enfermedades, por ejemplo las que causan beta-talasemia, aún no se sabe si el grado de enriquecimiento podría ser suficiente para la detección de aneuploidías cromosómicas fetales que requieren medición cuantitativa de la dosis de cromosomas. Dhallan et al. (2006)¹⁰ informaron de otro enfoque para el enriquecimiento del ADN fetal en el plasma materno. Lanzaron la hipótesis de que una porción significativa del ADN libre flotante derivado de la madre en el plasma materno es liberado por los glóbulos blancos maternos después de la flebotomía. Por lo tanto, se propuso que si las células sanguíneas nucleadas maternas pudieran ser fijadas, usando formaldehído, entonces esta dilución del ADN fetal en el plasma materno podría evitarse. Dhallan¹⁰ demostró el beneficio de este enfoque para el diagnóstico prenatal no invasivo de la trisomía 21 que muestra una proporción media de ADN fetal del 34% en un experimento que comprende 60 mujeres embarazadas. Sin embargo, otros grupos no pudieron replicar los efectos beneficiosos del tratamiento con formaldehído.

Las metodologías mencionadas anteriormente se basan en la suposición de que la baja concentración fraccional de ADN fetal en el plasma materno hace que resulte desafiante perseguir la detección directa de aneuploidías cromosómicas fetales. Esto se basa en la precisión limitada de los métodos convencionales para la detección de ADN fetal circulante, por ejemplo, mediante PCR en tiempo real.

La reciente disponibilidad de técnicas de conteo de moléculas individuales permite la detección de aneuploidía fetal sin la necesidad de restringir el análisis a ácidos nucleicos específicos del feto en el plasma materno. La PCR digital y la secuenciación masivamente paralela son métodos de recuento de moléculas individuales, que permiten la cuantificación de ácidos nucleicos mediante el recuento de moléculas y tienen una precisión analítica superior en comparación con los métodos convencionales de detección basados en PCR. La PCR digital se refiere a la realización de PCR múltiples en paralelo en las que cada PCR contiene típicamente una molécula diana única o ninguna. A través del recuento del número de reacciones positivas al final de la amplificación, es posible determinar el número de moléculas diana de entrada. Por lo tanto, pueden cuantificar con precisión pequeños incrementos en la cantidad total (materna vs. fetal) de moléculas de ADN derivadas del cromosoma aneuploide. De hecho, Lo et al. (2007)¹¹ demostraron que la detección del estado de aneuploidía es posible incluso cuando el ADN trisómico está presente como una fracción menor (10%). Cuanto menor es la concentración de ADN fetal, menor es el incremento esperado en la cantidad de ADN cromosómico aneuploide. Para la PCR digital, la precisión cuantitativa mejora al aumentar el número de análisis de PCR realizados. Lo et al. (2007)¹¹ demostraron que la detección precisa de la trisomía 21 fetal en una muestra de plasma materno que contiene un 25% de ADN fetal requiere la realización de aproximadamente 8000 PCR digitales, lo que requiere el uso de plataformas automatizadas en el entorno clínico. Dichas plataformas automatizadas que usan microfluidos están disponibles (por ejemplo, Fluidigm) pero son caras. Varios grupos demostraron que la detección no invasiva de trisomía 21 fetal podría lograrse con el uso de secuenciación masivamente paralela o de próxima generación (por ejemplo, WO2009/013496). Secuenciadores masivamente paralelos permiten el análisis de secuencias de nucleótidos de millones a miles de millones de moléculas de ADN en cada ciclo. Por lo tanto, además de la identidad, se puede obtener una distribución de frecuencia de las moléculas de ADN en la muestra analizada. Dado que el ADN que flota libremente en el plasma materno está fragmentado en la naturaleza, puede usarse directamente para identificar el origen cromosómico de cada molécula de ADN y determinar la proporción de moléculas derivadas de un cromosoma potencialmente aneuploide. Varios grupos mostraron que la proporción de moléculas de ADN del cromosoma 21 en plasma de mujeres embarazadas con un feto con trisomía 21 era elevada en comparación con la de los embarazos euploides. Este enfoque fue muy preciso para la detección directa de la trisomía fetal 21 del plasma materno en pequeñas cohortes de embarazos.

Recientemente, se realizaron dos estudios de validación clínica aplicando el método descrito anteriormente. En un estudio se analizaron 449 muestras de las cuales 39 eran trisómico para el cromosoma 21¹². En un segundo estudio se analizaron muestras de sangre de 1.014 en los embarazos de riesgo recogidos en 13 ubicaciones de las clínicas de los Estados Unidos antes de que se sometieran a un procedimiento prenatal invasivo¹³. De estas, 119 muestras se sometieron a secuenciación de ADN masivamente paralela. Cincuenta y tres muestras secuenciadas se clasificaron correctamente como con un cariotipo fetal anormal. Ambos estudios de validación clínica mostraron una excelente sensibilidad y especificidad. Estos datos demuestran que la secuenciación del ADN plasmático es un método viable para la detección no invasiva de la trisomía fetal 21 y justifica la validación clínica en un estudio multicéntrico más amplio.

Por otro lado, se ha demostrado que la medición de las cantidades proporcionales de secuencias derivadas de cromosomas con contenidos de GC más altos o más bajos que el cromosoma 21 no era tan robusta. Por lo tanto, las medidas para los cromosomas 18 y 13 son menos precisas y sufren de sesgo cuantitativo usando protocolos de trisomía 21. Por lo tanto, para lograr una detección no invasiva confiable de trisomía 18 y trisomía 13, los protocolos de secuenciación y análisis de datos que son menos susceptibles a los efectos cromosómicos del contenido GC deben desarrollarse y validarse aún más. Un estudio reciente resolvió parcialmente el problema anterior utilizando un genoma de referencia enmascarado no repetido y un enfoque bioinformático para corregir el sesgo del contenido de GC en los datos de secuenciación¹⁴. Usando este enfoque, todos los fetos de trisomía 13 (25 de 25) se

detectaron a una especificidad de 98.9% y 92% (34 de 37) de los fetos de trisomía 18 con 98.0% de especificidad. Estos datos indican que con un análisis bioinformático apropiado, el diagnóstico prenatal no invasivo de trisomía 13 y la trisomía 18 mediante la secuenciación del ADN del plasma materno no es tan confiable como la trisomía 21.

5 Además, el coste de la secuencia masiva paralela es alto y el rendimiento es bajo. Solo se pueden analizar un puñado de casos por ejecución, lo que lleva varios días. Se necesita más trabajo para desarrollar protocolos más rentables con un mayor rendimiento.

10 Recientemente, se utilizó el enriquecimiento de dianas para obtener un enfoque de secuenciación paralelo masivo más eficiente y rentable¹⁵. Este estudio investigó la aplicabilidad en el enriquecimiento de regiones genómicas seleccionadas del ADN plasmático y el rendimiento cuantitativo de este enfoque. El experimento mostró que la cobertura de la secuencia media de las muestras enriquecidas era ~200 veces mayor que la de las muestras no enriquecidas y, lo que es más importante, que las moléculas de ADN materno y fetal se enriquecieron de manera uniforme. Además, al usar datos de SNP, los autores pudieron demostrar que la cobertura de los alelos específicos del feto dentro de la región objetivo aumentó del 3.5% al 95.9%. En general, la secuenciación dirigida del ADN plasmático materno permite la detección eficiente e imparcial de los alelos fetales y es un método poderoso para medir la proporción de ADN fetal en una muestra de plasma materno. Con base en este único documento científico, el enriquecimiento de los objetivos es muy prometedor, ya que puede reducir sustancialmente el coste de la secuenciación. Al mismo tiempo, requiere un paso extra de enriquecimiento que agregará un coste adicional a la prueba final y también retrasará la prueba, ya que un protocolo de enriquecimiento típico demora alrededor de 24-36 horas en completarse.

20 El documento WO 2007/147079 describe sistemas, aparatos y métodos para detectar la presencia de células fetales cuando se mezclan con una población de células maternas en una muestra y para probar anomalías fetales, por ejemplo aneuploidía.

25 El documento WO 2007/147074 divulga un aparato y métodos para enriquecer componentes o células de una muestra y realizar análisis genéticos, tales como gen tipificado de SNP para proporcionar resultados de diagnóstico para trastornos o afecciones fetales.

Resumen de la invención

30 La presente invención proporciona una prueba diagnóstica de ADN no invasiva para la detección de aneuploidía de los cromosomas 21 y/o 18 y/o 13 y/o X y/o Y combinando la amplificación basada en PCR múltiple de secuencias de ADN específicas (es decir, dianas) que contienen al menos un SNP en combinación con tecnologías de secuenciación.

En otro aspecto, la invención proporciona una prueba diagnóstica de ADN no invasiva para la detección de aneuploidía de los cromosomas 21 y 18 y 13 y X e Y combinando la amplificación basada en PCR múltiple de secuencias de ADN específicas (es decir, dianas) que contienen al menos un SNP en combinación con tecnologías de secuenciación.

35 En resumen, la presente invención se dirige a un método de detección diferencial de un conjunto predeterminado de secuencias diana en una mezcla de material genético materno y fetal.

40 Por tanto, los métodos y materiales descritos en este documento aplican técnicas para analizar numerosos ácidos nucleicos preparados a partir de una muestra biológica (preferiblemente suero o plasma) que contiene ADN libre flotante que es una mezcla de ADN tanto de la madre como del feto y permite la detección de diferencia entre fetos euploides y triploides. A diferencia de los actuales métodos masivos de secuenciación paralela, basados en muestras genómicas completas o enriquecidas, que no alcanzan una sensibilidad y especificidad suficientes, en particular, para el cromosoma 13, la presente invención proporciona un ensayo de diagnóstico no invasivo con una especificidad y sensibilidad cerca del 100% (respectivamente 99.99% de especificidad y 99.5% de sensibilidad) para la detección simultánea de cromosoma 13, 18, 21, X e Y aneuploidies. Sin limitar la invención a una teoría o explicación particular, una razón por la que la PCR múltiple no se consideró antes en el desarrollo de pruebas de aneuploidía diagnósticas no invasivas es la presencia de regiones ricas en GC particularmente en el cromosoma 13. Otra razón es que el uso de multiplex-PCR fue desaconsejado por uno de los principales inventores (es decir, Dennis Lo) en el documento US2010/0112590. De hecho, en la última aplicación de los párrafos 116-117, se recomienda aplicar ensayos independientes de locus en lugar de ensayos dependientes de locus, tales como, por ejemplo, la amplificación dirigida llevada a cabo por los métodos de la presente invención.

50 Por tanto, en un aspecto, la invención proporciona un método para determinar la presencia o ausencia de aneuploidía fetal en el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y en ADN flotante libre preparado a partir de una muestra biológica materna, amplificando un conjunto seleccionado de secuencias de ADN diana en una reacción de PCR múltiple cuantitativa (es decir, amplificando el ADN plantilla de manera que el ADN amplificado está reproduciendo la relación de ADN plantilla original), donde las secuencias de ADN amplificadas tienen longitudes por debajo de 140 pares de bases, llevando a cabo secuenciación de ADN de dicho conjunto amplificado de secuencias de ADN diana para determinar la secuencia de dichas secuencias de ADN, usando los datos de secuencia obtenidos para comparar una cantidad de secuencias

amplificadas derivadas de al menos un primer cromosoma en dicha mezcla de ADN materno y fetal para una cantidad de secuencias de ADN amplificadas derivadas de al menos un segundo cromosoma en dicha mezcla de ADN materno y fetal, donde dicha al menos una primera se presume que el cromosoma es euploide en el feto, en donde se sospecha que al menos un segundo cromosoma es aneuploide en el feto, determinando de ese modo la presencia o ausencia de dicha aneuploidía fetal.

En otro aspecto, la invención proporciona un método para determinar la presencia o ausencia de aneuploidía fetal en el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y en ADN libre flotante preparado a partir de una muestra biológica materna, amplificando un conjunto seleccionado de secuencias de ADN diana en una reacción de PCR múltiple cuantitativa donde cada secuencia de ADN amplificada comprende al menos un SNP, y en donde las secuencias de ADN amplificadas tienen longitudes por debajo de 140 pares de bases que se consideran informativas en caso de que la mujer embarazada sea heterocigótica para este SNP, llevando a cabo secuenciación de ADN de dicho conjunto amplificado de secuencias de ADN diana para determinar la secuencia de dichas secuencias de ADN, usando los datos de secuencia obtenidos para comparar una cantidad de secuencias amplificadas que llevan un SNP informativo derivado de al menos un primer cromosoma en dicha mezcla de ADN derivado de la madre y el feto a una cantidad de secuencias de ADN amplificadas que llevan una derivación SNP informativa a partir de al menos un segundo cromosoma en dicha mezcla de ADN derivado de la madre y del feto, en el que se supone que al menos un primer cromosoma es euploide en el feto, donde se sospecha que dicho al menos un segundo cromosoma es aneuploide en el feto, determinar la presencia o ausencia de dicha aneuploidía fetal y/o determinar en dichas secuencias de ADN determinadas las relaciones alélicas de los SNP informativos en los que una ración alélica distorsionada es indicativa de la presencia de una aneuploidía cromosómica fetal en el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y en dicha mujer embarazada.

Leyendas de las Figuras

Figura 1: Cocientes de Dosificación (DQ) del feto trisómico en comparación con el feto euploide. El área sombreada gris indica los porcentajes esperados de ADN fetal.

Figura 2: Número de SNP necesarios para obtener como mínimo un número dado de SNP informativos, representados por Frecuencia del Alelo Menor (MAF). Los cálculos se realizan con una probabilidad mínima del 99%.

Figura 3: Gráfica de los recuentos de lectura normalizados esperados frente a los observados para el cromosoma 21 en un síndrome de Down (trisomía 21), muestras de ADN (cuadradas) y 4 muestras de ADN euploide (círculos).

Figura 4: Gráfica de recuentos de lectura normalizados esperados frente a los observados para dos muestras de ATP (que representan el 20% de trisomía 21) muestras de ADN (cuadrados) y 4 muestras de ADN euploide (círculos).

Figura 5: Representación esquemática de la primera reacción de PCR múltiple del procedimiento de ensayos MASTR. Los cebadores inversos y directos son cebadores específicos de amplicón. Tag1 y Tag2 son secuenciaciones universales que se utilizan en la segunda reacción de PCR del procedimiento de ensayo MASTR para incorporar

Figura 6: Representación esquemática de la segunda reacción de PCR del procedimiento MASTR. En este paso, las secuencias MID (códigos de barras) y los adaptadores A y B (para PCR en emulsión 454) se incorporan en los amplicones resultantes de la primera reacción de PCR.

Descripción detallada de la invención

La técnica anterior ha demostrado la viabilidad de la secuenciación paralela masiva como una plataforma de análisis para la prueba de aneuploidía basada en ADN flotante libre. Sin embargo, los protocolos actuales dan como resultado pruebas costosas y de bajo rendimiento cuando se usan como herramienta de diagnóstico molecular. La razón principal de esto es el hecho de que las pruebas actuales se basan en la secuenciación genómica del ADN flotante que resulta en la producción de grandes conjuntos de datos de secuenciación de los cuales solo una pequeña fracción (~5%) se usa para determinar el estado de ploidía del feto. Con este amplio enfoque del genoma, es obligatorio utilizar una parte sustancial de la capacidad de un secuenciador paralelo masivo, lo que da como resultado la secuenciación de un número limitado de individuos por ejecución, que tarda varios días en completarse. Además, se generan grandes conjuntos de datos de secuenciación por individuo que dificultan el almacenamiento y análisis de datos eficientes.

La presente invención ofrece una solución para este problema utilizando un enfoque basado en PCR múltiple para amplificar varias regiones cromosómicas seleccionadas. Las regiones cromosómicas seleccionadas se amplifican en una reacción de PCR múltiple de uno o más cromosomas que se presume son aneuploides y se amplifican en un conjunto seleccionado de regiones cromosómicas, preferiblemente en la misma reacción de PCR múltiple, de uno o

más cromosomas que se presume son euploides. Los cromosomas que se supone que son euploides se designan aquí adicionalmente como un "cromosoma de referencia".

Por consiguiente, la presente invención proporciona en una primera realización un método para la detección de una aneuploidía cromosómica fetal en una mujer embarazada que comprende i) preparar ADN flotante libre de una muestra biológica de dicha mujer embarazada, ii) amplificar un conjunto seleccionado de secuencias de ADN diana de uno o más cromosomas que se supone que son aneuploides y que amplifican un conjunto seleccionado de secuencias de ADN diana de uno o más cromosomas que se supone que son euploides en al menos una reacción de PCR múltiple cuantitativa en la que al menos una secuencia de ADN amplificada comprende al menos un SNP que se considera informativo si la mujer gestante es heterocigota para este SNP donde las secuencias de ADN amplificadas tienen longitudes inferiores a 140 pares de bases, iii) secuenciación de las secuencias de ADN diana amplificadas y iv) cálculo de la suma de recuentos de lectura para todas las secuencias de ADN amplificadas de una sospecha de aneuploidía cromosómica seguida de normalización, contra la suma de los recuentos de lectura para todo la secuencias de ADN amplificadas de un cromosoma de referencia para determinar mediante métodos estadísticos una puntuación establecida indicativa de la presencia de una aneuploidía cromosómica fetal y/o determinar las relaciones alélicas de los SNP informativos en los que una relación alélica distorsionada es indicativa de la presencia de una aneuploidía cromosómica fetal en dicha mujer embarazada, en donde dicha aneuploidía cromosómica fetal es el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y.

El término "muestra biológica" como se usa en el presente documento se refiere a cualquier muestra que se toma de un sujeto que es del sexo femenino embarazada o una mujer embarazada y que contiene una o más moléculas de ácido nucleico de interés.

Por consiguiente, una muestra biológica comprende, por ejemplo, sangre, esputo, orina, fluido cerebroespinal (CSF), lágrimas, plasma, suero, saliva o fluido de lavado transcervical.

El término "ácido nucleico" o "polinucleótido" se refiere a un ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN) y a un polímero del mismo en forma monocatenaria o bicatenaria. A menos que se limite específicamente, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares a los del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia particular de ácido nucleico también abarca implícitamente variantes modificadas conservativamente de la misma (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, SNP y secuencias complementarias, así como la secuencia explícitamente indicada. Específicamente, las sustituciones de codones degenerados se pueden lograr generando secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con restos de base mixta y/o desoxinosina. El término ácido nucleico se usa de forma intercambiable con gen, ADNc, ARNm, ARN no codificante pequeño, ARN micro (miARN), ARN que interacciona con Piwi y ARN de horquilla corta (ARNhc) codificado por un gen o locus.

El término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica. Puede incluir regiones que preceden y siguen a la región de codificación (líder y final) así como secuencias intermedias (intrones) entre segmentos de codificación individuales (exones). El término "reacción", como se usa en el presente documento, se refiere a cualquier proceso que implica una acción química, enzimática o física que es indicativa de la presencia o ausencia de una secuencia de polinucleótido particular de interés. Un ejemplo de una "reacción" es una reacción de amplificación tal como una reacción en cadena de la polimerasa (PCR), preferiblemente una reacción de PCR múltiple. Otro ejemplo de una "reacción" es una reacción de secuenciación, ya sea por síntesis o por ligadura. El término "secuencia de ácido nucleico clínicamente relevante" como se usa en el presente documento puede referirse a una secuencia de polinucleótidos que corresponde a un segmento de una secuencia genómica más grande cuyo desbalance potencial está siendo probado o a la secuencia genómica más grande en sí misma. Los ejemplos incluyen los cromosomas 18, 13, 21, X e Y. Otros ejemplos más incluyen secuencias genéticas mutadas o polimorfismos genéticos o variaciones en el número de copias que un feto puede heredar de uno o ambos de sus padres. El término "secuencia de ácido nucleico de fondo", como se usa en el presente documento, puede referirse a secuencias de ácido nucleico que se originan en la madre o que se originan a partir del cromosoma no evaluado para aneuploidía en un análisis particular.

El término "ADN de flotación libre" es ADN que se deriva de ADN genómico; el ADN de flotación libre es de hecho ADN genómico degradado y se produce en el espacio extracelular. Como tal, el ADN que flota libremente se puede aislar de los fluidos corporales (por ejemplo, suero, plasma, esputo). El término "datos cuantitativos" como se usa en el presente documento significa datos que se obtienen a partir de una o más reacciones y que proporcionan uno o más valores numéricos. El término "parámetro" tal como se usa en el presente documento significa un valor numérico que caracteriza un conjunto de datos cuantitativos y/o una relación numérica entre conjuntos de datos cuantitativos. Por ejemplo, una relación (o función de una relación) entre una primera cantidad de una primera secuencia de ácido nucleico y una segunda cantidad de una segunda secuencia de ácido nucleico es un parámetro.

5 El término "valor de corte" como se usa en este documento significa un valor numérico cuyo valor se usa para arbitrar entre dos o más estados (por ejemplo, enfermo y no enfermo) de clasificación para una muestra biológica. Por ejemplo, si un parámetro es mayor que el valor de corte, se realiza una primera clasificación de los datos cuantitativos (por ejemplo, estado patológico); o si el parámetro es menor que el valor de corte, se realiza una clasificación diferente de los datos cuantitativos (por ejemplo, estado no enfermo).

El término "desbalance" como se usa en la presente memoria significa cualquier desviación significativa como se define por al menos un valor de corte en una cantidad de la secuencia de ácido nucleico clínicamente relevante a partir de una cantidad de referencia.

10 El término "aneuploidía cromosómica" como se usa en el presente documento significa una variación en la cantidad cuantitativa de un cromosoma de la de un genoma diploide. La variación puede ser una ganancia o una pérdida. Puede involucrar todo un cromosoma o una región de un cromosoma. De acuerdo con la presente invención, las aneuploidías cromosómicas se derivan del cromosoma 13, 18, 21, X e Y.

15 El término "secuenciación aleatoria" como se usa en la presente memoria se refiere a la secuenciación por la cual los fragmentos de ácido nucleico secuenciados no se han identificado específicamente ni se han dirigido antes del procedimiento de secuenciación. Los cebadores específicos de secuencia para localizar loci génicos específicos no son necesarios cuando se aplica una secuencia aleatoria. Los grupos de ácidos nucleicos secuenciados varían de muestra a muestra e incluso de análisis a análisis para la misma muestra. En la secuenciación aleatoria, las identidades de los ácidos nucleicos secuenciados solo se revelan a partir de la secuencia de salida generada en contraste con la secuenciación de las secuencias de nucleótidos amplificadas por PCR múltiple.

20 Las realizaciones de esta divulgación proporcionan métodos, sistemas y aparatos para determinar si existe un aumento o disminución (estado enfermo) de una región cromosómica clínicamente relevante en comparación con un estado no enfermo. Esta determinación puede realizarse usando un parámetro de una cantidad de una región cromosómica clínicamente relevante en relación con otras regiones cromosómicas no clínicamente relevantes (regiones de fondo) dentro de una muestra biológica. Las moléculas de ácido nucleico de la muestra biológica se secuencian, de modo que se secuencia una fracción del genoma, y la cantidad se puede determinar a partir de los resultados de la secuenciación. Se eligen uno o más valores de corte para determinar si existe un cambio comparado con una cantidad de referencia (es decir, un desbalance), por ejemplo, con respecto a la relación de cantidades de dos regiones cromosómicas (o conjuntos de regiones).

30 El cambio detectado en la cantidad de referencia puede ser cualquier desviación (hacia arriba o hacia abajo) en la relación de la secuencia de ácido nucleico clínicamente relevante con las otras secuencias no clínicamente relevantes. Por lo tanto, el estado de referencia puede ser cualquier proporción u otra cantidad (por ejemplo, distinta de una correspondencia 1-1), y un estado medido que signifique un cambio puede ser cualquier relación u otra cantidad que difiera de la cantidad de referencia determinada por uno o más valores de corte.

35 La región cromosómica clínicamente relevante (también llamada secuencia de ácido nucleico clínicamente relevante o sospecha de cromosoma aneuploide o región cromosómica) y la secuencia de ácido nucleico de fondo pueden provenir de un primer tipo de células y de uno o más segundos tipos de células. Por ejemplo, las secuencias de ácidos nucleicos fetales que se originan en células fetales/placentarias están presentes en una muestra biológica, como el plasma materno, que contiene un fondo de secuencias de ácidos nucleicos maternos que se originan en células maternas. Las secuencias de ácidos nucleicos maternos y fetales se derivan de ADN flotante. En una realización, el valor de corte se determina basándose al menos en parte en un porcentaje del primer tipo de células en una muestra biológica. Obsérvese que el porcentaje de secuencias fetales en una muestra puede determinarse mediante cualquier loci derivado del feto y no se limita a medir las secuencias de ácido nucleico clínicamente relevantes.

45 En otra realización, los métodos de la invención usan productos químicos estabilizadores de células (por ejemplo, células sanguíneas) en la preparación de los ácidos nucleicos presentes en la muestra biológica que se recibe de la mujer embarazada. De hecho, uno de los mayores desafíos técnicos en el uso de ADN fetal flotante de la sangre materna es la baja fracción de ADN fetal presente en la muestra. Esta fracción es típicamente entre 10 y 20% en el primer trimestre del embarazo (semana 11-14), que corresponde con la etapa en la que se realiza mejor una prueba de aneuploidía de ADN. Esta baja fracción de ADN fetal es incluso para métodos de recuento molecular desafiantes con respecto a la sensibilidad y especificidad de la prueba. Por lo tanto, es importante maximizar la proporción de ADN flotante fetal/materno. La presente invención proporciona diferentes soluciones para este problema.

55 En una realización particular, la interrupción de las células sanguíneas nucleadas se evita durante la recolección, el almacenamiento o el transporte del material biológico, en particular una muestra de sangre materna antes del aislamiento del plasma. Esto es importante para evitar la dilución de ADN fetal que da como resultado una disminución de la proporción de ADN flotante fetal/materno. Se encuentran disponibles varios tubos comerciales de

recolección de sangre estabilizadora de células que estabilizan las células sanguíneas durante al menos 14 días a temperatura ambiente, lo que permite la recolección, el transporte y el almacenamiento de muestras de manera conveniente (disponible, por ejemplo, en www.streck.com).

5 En aún otra realización particular, se usa un fraccionamiento por tamaño en los métodos de la invención para preparar ácidos nucleicos maternos y fetales.

10 De hecho, la técnica anterior muestra que el ADN flotante fetal y materno tiene diferentes distribuciones de tamaño. El ADN fetal flotante libre es generalmente 20 pb más corto que el ADN flotante libre de la madre y esta observación se puede usar para enriquecer aún más la fracción de ADN fetal flotante si esta fracción de menor tamaño se separa específicamente de la fracción materna. Una forma de lograr esto es mediante electroforesis en gel. En una realización particular, se usa un dispositivo de fraccionamiento de tamaño basado en electroforesis en gel comercializado por Sage Science (www.sagescience.com). Este dispositivo es un sistema totalmente automatizado que permite una selección de tamaño ajustado y una alta tasa de recuperación. Además, elimina por completo el riesgo de contaminación cruzada ya que todas las muestras se separan entre sí durante todo el proceso de fraccionamiento.

15 En una realización particular, las secuencias de ADN amplificadas obtenidas en la reacción de PCR múltiple cuantitativa en los métodos de la invención tienen un tamaño entre 80 y 140 pares de bases.

En vista de las distribuciones de tamaño de las poblaciones de ADN libre flotante fetal y materna, es esencial mantener las longitudes de la secuencia de ADN amplificada por debajo de 140 pb para asegurar la amplificación eficiente de la fracción de ADN de flotación libre fetal más corta.

20 Las longitudes de secuencia de ADN amplificadas preferidas están entre 80 y 140 pares de bases.

En otra realización más, las secuencias de ADN amplificadas obtenidas en una única reacción de PCR múltiple están entre 30 y 80.

En otra realización más, las secuencias de ADN amplificadas obtenidas en una única reacción de PCR múltiple están entre 60 y 80.

25 En otra realización más, las secuencias de ADN amplificadas obtenidas en una única reacción de PCR múltiple están entre 70 y 80.

Preferentemente, solo se aplica una reacción de PCR múltiple cuantitativa para practicar los métodos de la invención.

30 En otra realización más, el contenido de GC de las secuencias de ADN diana (es decir, las secuencias de ADN que se amplifican con la reacción de PCR múltiple cuantitativa) está entre 30% y 70%. Nuestros datos experimentales señalan que un rango de 40% -60% de GC es óptimo para una sensibilidad y especificidad cercanas al 100% de los métodos de la invención.

35 Una etapa esencial en los métodos de la presente invención es la secuenciación de las secuencias de ADN diana amplificadas. Como un gran número de lecturas de secuenciación, del orden de cientos de miles a incluso posiblemente cientos de millones o miles de millones teóricamente pueden generarse a partir de cada muestra en cada ejecución, las lecturas secuenciadas resultantes forman un perfil representativo de la mezcla de especies de ácidos nucleicos en la muestra biológica original. Sin embargo, la persona experta en la técnica sabrá cuántas carreras realizar en función de la etapa del embarazo (que se correlaciona con la cantidad de ADN fetal flotante en la muestra biológica) y en función del origen de la muestra biológica derivada de una mujer embarazada. El aspecto más importante es que se obtiene un alto grado de confianza estadística. Para mejorar la confianza estadística, es preferible realizar un gran número de lecturas, preferiblemente entre 10.000 y 100.000 o más lecturas, dependiendo del porcentaje de ADN fetal presente en la mezcla. Una medida comúnmente utilizada de significación estadística cuando se desea un resultado altamente significativo es $p < 0.01$, es decir, un intervalo de confianza del 99% basado en una prueba de cuadrados-chi o t.

45 En una realización preferida, se usan métodos de secuenciación paralelo masivo. En realizaciones particulares, la secuenciación se realiza usando secuenciación masivamente paralela. Secuenciación paralela masiva, como por ejemplo en la plataforma 454 (Roche) (Margulies, M. et al., 2005 Nature 437, 376-380), Illumina Genome Analyzer (o plataforma Solexa) o SOLiD System (Applied Biosystems) o los Helicos la verdadera tecnología de secuenciación de ADN de molécula única (Harris T D et al., 2008 Science, 320, 106-109), la molécula única, la tecnología en tiempo real (SMRT™) de Pacific Biosciences y la secuenciación de nanoporos (Soni G V and Meller A. 2007 Clin Chem 53:

1996-2001), permiten la secuenciación de muchas moléculas de ácido nucleico aisladas de una muestra en altos órdenes de multiplexación de forma paralela. Cada una de estas plataformas secuencian de moléculas de fragmentos de ácidos nucleicos clonalmente expandidas o incluso no amplificadas.

5 Una ventaja importante del conjunto limitado de secuencias de nucleótidos amplificadas que se generan mediante los métodos de la presente invención es que se pueden usar secuenciadores paralelos masivos emergentes de bajo coste y menor capacidad tales como el 454 junior (Roche), PGM (Life Technologies) o MiSeq (Illumina). La combinación de los métodos de la invención y los secuenciadores de gama baja da como resultado un tiempo de respuesta rápido por prueba, ya que estas plataformas típicamente toman solo unas pocas horas por ejecución de secuencia. Además, el menor coste también es una mejora importante con respecto a los métodos utilizados en la técnica anterior.

10 En una realización particular, los datos de secuenciación en paralelo masiva se analizan calculando la suma de recuentos de lectura para todas las secuencias de ADN amplificadas de una aneuploidía cromosómica sospechosa (por ejemplo, todas las secuencias de ADN amplificadas derivadas del cromosoma 21 y/o cromosoma 13 y/o cromosoma 18 y/o el cromosoma X y/o el cromosoma Y) se cuentan (es decir, el número de veces que está presente una secuencia cromosómica amplificada específica en la muestra biológica). La suma de los recuentos de lectura para las secuencias de ADN amplificadas derivadas de un cromosoma aneuploide particular sospechoso (por ejemplo cromosoma 13 o 18 o 21 o X o Y) se normaliza contra la suma de recuentos de lectura para las secuencias de ADN amplificadas derivadas de un cromosoma de referencia (es decir, un cromosoma para el que no se informa aneuploidía). Por lo tanto, la PCR múltiple permite el cálculo de Cocientes de Dosificación (DQ) al comparar (recuento de lecturas de la región diana, es decir, el cromosoma aneuploide sospechoso o la región cromosómica)/(control región lectura conteo, es decir, el cromosoma de referencia o la región cromosómica) relaciones entre la mujer embarazada y el feto. Los DQ en función del porcentaje de ADN fetal se representan en la Figura 1.

15 Un elemento esencial de los métodos de la presente invención es que las secuencias de ADN diana amplificadas están reflejando proporciones idénticas de las cantidades de ácidos nucleicos flotantes maternos y fetales en la muestra biológica y, por lo tanto, los métodos requieren amplificación cuantitativa. En base a los ensayos de PCR multiplex y las condiciones de PCR utilizadas para amplificar las muestras (número limitado de ciclos), hemos demostrado previamente que el ADN plantilla se amplifica cuantitativamente¹⁶. Si hay una distribución normal entre los dos recuentos de lectura, entonces se obtiene una puntuación (por ejemplo, un puntaje Z o un cociente de dosificación). Un puntaje Z de 1 significa que no hay aneuploidía para el cromosoma sospechado de aneuploidía. Una puntuación Z superior a 1, preferentemente superior a 2, más preferentemente superior a 3, es una indicación de la presencia de una aneuploidía del cromosoma. Se entiende que los puntajes Z se determinan para todos los cromosomas sospechosos de aneuploidía para los cuales se obtiene un conjunto seleccionado de secuencias diana de ADN mediante los métodos de la invención. La normalización y el cálculo del puntaje Z es asistido por el uso de métodos estadísticos. Los métodos estadísticos útiles que se pueden usar en el contexto de la presente invención incluyen el método de verosimilitud de tipo Bayesiano, el ensayo de relación de probabilidad secuencial (SPRT), el descubrimiento falso, el intervalo de confianza y la característica operativa del receptor (ROC).

20 En aún otra realización particular, los datos de secuenciación en paralelo masiva de las secuencias de ADN diana amplificadas se analizan basándose en la determinación de las relaciones alélicas de los SNP informativos en los que una relación distorsionada es indicativa de la presencia de aneuploidía cromosómica fetal en la mujer embarazada. La proporción alélica está distorsionada para SNP informativos en cromosomas aneuploides. Esta distorsión se puede medir cuando la madre es heterocigótica para un SNP dado (denominado en este documento "SNP informativo"). Por lo tanto, el análisis de secuencia del ensayo MASTR dará como resultado una serie de SNP informativos que se pueden usar para determinar el estado de ploidía fetal además de la determinación del estado de ploidía fetal por recuento molecular como se describió anteriormente. La Figura 2 muestra el resultado de un cálculo del número de SNP informativos con un 99% de probabilidad siempre que la frecuencia de alelos minoritarios (MAF) sea de 0.25 a 0.50. En base a este cálculo, se muestra en la Figura 2 que con un mínimo de MAF de 0.25 al menos 7 SNP informativos están presentes en un conjunto de 35 secuencias de ADN diana amplificadas, mientras que 10 SNP informativos se identifican para SNP con un MAF de 0.50.

25 En otra realización particular, los datos de secuenciación paralelos masivos de las secuencias de ADN diana amplificadas se analizan basándose en la determinación de las relaciones alélicas de los SNP informativos en los que una relación distorsionada es indicativa de la presencia de una aneuploidía cromosómica fetal en la mujer embarazada en combinación con el cálculo de la suma de recuentos de lectura para todas las secuencias de ADN amplificadas de una aneuploidía cromosómica sospechosa (por ejemplo, todas las secuencias de ADN amplificadas derivadas del cromosoma 21 y/o el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma X y/o el cromosoma Y) (es decir, el número de veces que una secuencia cromosómica amplificada específica está presente en la muestra biológica).

5 En otra realización más basada en llevar a cabo los métodos de la invención, se determina si existe una aneuploidía cromosómica fetal para uno o más cromosomas aneuploides sospechosos, donde los cromosomas son el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y. En una realización, la clasificación es un sí o un no definitivo. En otra realización más, una clasificación puede ser inclasificable o incierta. En otra realización más, la clasificación puede ser una puntuación que se interpretará en una fecha posterior, por ejemplo, por un médico.

10 En realizaciones particulares, los enfoques bioinformáticos, computacionales y estadísticos utilizados para determinar si una muestra biológica obtenida de una mujer embarazada concebida con un cromosoma aneuploide o región cromosómica o feto euploide podrían compilarse en un producto de programa informático utilizado para determinar parámetros a partir del resultado de la secuenciación. El funcionamiento del programa informático implicaría la determinación de una cantidad cuantitativa del cromosoma potencialmente aneuploide, así como las cantidades de uno o más de los otros cromosomas. Se determinará un parámetro y se lo comparará con los valores de corte apropiados para determinar si existe una aneuploidía cromosómica fetal para el cromosoma potencialmente aneuploide.

15 También se describe un kit de diagnóstico para llevar a cabo el método. Tal kit de diagnóstico comprende al menos un conjunto de cebadores para amplificar ácidos nucleicos fetales maternos y diana en los que estos ácidos nucleicos diana se derivan del cromosoma 13 y/o cromosoma 18 y/o cromosoma 21 y/o cromosoma X y/o cromosoma Y. Preferentemente, el kit comprende cebadores para amplificar ácidos nucleicos diana derivados de los cromosomas 13, 18, 21, X e Y. Además, el kit de diagnóstico comprende un conjunto de cebadores que pueden identificar secuencias de ADN diana de un cromosoma de referencia o una referencia parte cromosómica. Se entiende que dicho cromosoma de referencia o parte del mismo es un cromosoma euploide. Euploide se refiere a la cantidad normal de cromosomas. Otros reactivos que pueden incluirse opcionalmente en el kit de diagnóstico son instrucciones y una polimerasa y reguladores para llevar a cabo la reacción de PCR multiplex de polimerasa cuantitativa.

25 Ejemplos

Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar la invención reivindicada.

1. Diagnóstico prenatal de trisomía fetal 21

30 Las muestras de ADN usadas en los presentes ejemplos son muestras preparadas mezclando una muestra de ADN diploide derivada de una mujer (que representa el ADN materno) con una muestra de muestra de ADN masculina euploide para el cromosoma 21 (denominado embarazo euploide artificial o AEP) o con un triploide de muestra de ADN masculino para el cromosoma 21 (conocido como embarazo por trisomía artificial o ATP). Cada muestra artificial estaba compuesta por una mezcla de 80% de ADN materno y 20% de ADN masculino. Además, se incluyó en el análisis una muestra de ADN derivada de un individuo con síndrome de Down, que tenía 3 copias del cromosoma 21.

35 Las mediciones se realizaron en 4 muestras de AEP, 2 muestras de ATP y 1 muestra de ADN del síndrome de Down. Para cada medición, se usaron aproximadamente 50 ng de ADN en un procedimiento de amplificación por PCR de ensayo MASTR estándar en 2 pasos (ver Materiales y métodos). El ensayo MASTR del cromosoma 21 fetal está compuesto por 20 pares de cebadores derivados del cromosoma 21 y 10 pares de cebadores derivados del cromosoma 18. Los amplicones resultantes de cada muestra de ADN individual amplificada con MASTR contenían un código de barras específico. Los amplicones de código de barras resultantes de cada muestra de ADN se mezclaron equimolarmente y se sometieron al protocolo de PCR de emulsión 454 Junior según lo descrito por el fabricante. Después de la PCR en emulsión, las perlas se aislaron y se cargaron en un 454 Junior de acuerdo con el protocolo del fabricante. Se realizaron un total de dos ejecuciones 454 Junior para obtener lecturas suficientes para alcanzar una cobertura por amplicón de entre 300 y 500.

45 Dado que la muestra de ADN del síndrome de Down contiene 3 copias del cromosoma 21, debe proporcionar un 50% más de lecturas del cromosoma 21 que las muestras AEP. Para calcular esto, se realizaron los siguientes pasos de cálculo en la muestra Down y en las muestras AEP:

(i) Los recuentos de lectura para cada cromosoma 18 y 21 amplicones se dividieron por el número total de recuentos de lectura derivados del cromosoma 18

50 (ii) Para cada amplicón de cromosomas 18 y 21, se calculó el recuento de lectura promedio sobre las diferentes muestras de AEP

(iii) Para cada amplicón del cromosoma 18 y 21, (i) se dividió por (ii)

(iv) Para cada cromosoma y cada muestra, se calculó el valor promedio de (iii)

(v) La relación normalizada observada cromosoma 21/cromosoma 18 se calculó dividiendo promedios calculados en (iv) por AEP y ATP

5 La Figura 3 muestra un gráfico del número observado (calculado como anteriormente) y esperado (es decir, valores teóricos) de lecturas para amplicones del cromosoma 21 de la muestra de ADN descendente.

Estos datos muestran que se puede hacer una distinción clara entre una muestra de ADN euploide normal y una muestra de ADN del cromosoma 21 (es decir, síndrome de Down).

10 Para evaluar la viabilidad de distinguir entre una muestra euploide (representada por las muestras artificiales AEP) y una muestra artificial de aneuploidía del cromosoma 21 que contiene 20% de ADN derivado de trisomía del cromosoma 21, los cálculos anteriores se realizaron en las muestras ATP relativas a las muestras AEP.

15 Una presencia del 20% de ADN de trisomía en las muestras de ATP debería dar como resultado un aumento del 10% en el recuento de amplicones del cromosoma 21 en comparación con las muestras de AEP. De hecho, utilizando los cálculos anteriores, la figura 4 muestra una clara distinción entre las muestras de AEP y ATP que reflejan un aumento de aproximadamente 10% en el cromosoma 21 en las dos muestras de ATP.

Materiales y métodos

1. Secuencias de cebador usadas en los ejemplos

Tabla 1: Lista de 30 pares de cebadores que componen el ensayo MASTR de detección de aneuploidía del cromosoma 21

Amplicón	Cromosoma	Avance	Retroceso
NITT_089	cr 18	AAGACTCGGCAGCATCTCCATTTGGAGTTAG CTTGACTTTGG	GCGATCGTCACTGTTCTCCAGAGATGGTATTA GGAAGGTTTGGT
NITT_092	cr 18	AAGACTCGGCAGCATCTCCACACTTTCTCCT AACACCCTTGG	GCGATCGTCACTGTTCTCCAGTGGGTGTCCTT AGGGGTCT
NITT_096	cr 18	AAGACTCGGCAGCATCTCCATCAGCACTCCC TCCATGA	GCGATCGTCACTGTTCTCCACTCAAAGAAATG GAAGAGAATACAAAA
NITT_097	cr 18	AAGACTCGGCAGCATCTCCACCTGCATCTTG ACACAGTCCG	GCGATCGTCACTGTTCTCCAGGCATCCAGGAG GAGAAAA
NITT_093	cr 18	AAGACTCGGCAGCATCTCCAGGATGGTCAC AGTGGGTCA	GCGATCGTCACTGTTCTCCAGAAGAGGGGAG AGTAGAGGTTAAA
NITT_085	cr 18	AAGACTCGGCAGCATCTCCATAAGCAAACAG CAGCACAAAA	GCGATCGTCACTGTTCTCCAGAGGGAATCTGT AATCCACATGA
NITT_094	cr 18	AAGACTCGGCAGCATCTCCACCAGAGTGGA ATTGCTGAGAC	GCGATCGTCACTGTTCTCCACTCCTTCTCTTTC TTCTTCTTCT AAGC

ES 2 651 612 T3

Amplificación	Crom	Avance	Retroceso
NITT_084	cr 18	AAGACTCGGCAGCATCTCCATGCAGATGGA GGACATCGT	GCGATCGTCACTGTTCTCCATTAATTTGCTCT TGGTATACTTCTTG
NITT_083	cr 18	AAGACTCGGCAGCATCTCCATGGTCCAGTTG GAGGGTCT	GCGATCGTCACTGTTCTCCATCACAGATGACAT GGAAAAAAGC
NITT_091	cr 18	AAGACTCGGCAGCATCTCCATAAAAGTGCCT TTGAACTCTGACTA	GCGATCGTCACTGTTCTCCACCCATGTGAAAT CGCATAGTT
NITT_050	cr 21	AAGACTCGGCAGCATCTCCACATCCAGGAC CTACCATCTTG	GCGATCGTCACTGTTCTCCACAACGCTGGCAT TCAAAA
NITT_010	cr 21	AAGACTCGGCAGCATCTCCACCTTCTCACTC ACCTCTTTCTTG	GCGATCGTCACTGTTCTCCAGTGTGCAGAGGA GAGACATGA
NITT_011	cr 21	AAGACTCGGCAGCATCTCCATGTGTGTGTGT TCTCTACCTTGG	GCGATCGTCACTGTTCTCCATATGAGTAGGTG TCTGGTGTATGAAAA
NITT_047	cr 21	AAGACTCGGCAGCATCTCCAGCAAATCTGGT ACTGGGTATGA	GCGATCGTCACTGTTCTCCATCAGATAGTATG GATAAAGGCAATGA
NITT_006	cr 21	AAGACTCGGCAGCATCTCCACAATAATCAGA CTTTGCCTTGG	GCGATCGTCACTGTTCTCCACATTAAGGGTCTT AGGGTGGTAAA
NITT_049	cr 21	AAGACTCGGCAGCATCTCCATCCTGTTGGG GAAATTGG	GCGATCGTCACTGTTCTCCAGAAGATAGAGTTT CTCCTGCATCA
NITT_017	cr 21	AAGACTCGGCAGCATCTCCAGGAACAGGTG CACACATCA	GCGATCGTCACTGTTCTCCACAGCACTGTCCA GCACTTG
NITT_007	cr 21	AAGACTCGGCAGCATCTCCACAGCTGTAACC TGCTGAGAAAA	GCGATCGTCACTGTTCTCCAGTCTTAATTCTGC TCAGGAAAAGC
NITT_009	cr 21	AAGACTCGGCAGCATCTCCAGAACAGCATT CTCCTCCTAGT	GCGATCGTCACTGTTCTCCATTGAACCATAAAT GTCAGCTCTTG
NITT_070	cr 21	AAGACTCGGCAGCATCTCCACCTCACATGTC TGTGCATTA AAA	GCGATCGTCACTGTTCTCCATTCCCTCTTCACA TTCTGCTC
NITT_071	cr 21	AAGACTCGGCAGCATCTCCAGACACAACATC AGAGGCAATCT	GCGA TCGTCACTGTTCTCCACTCTTCAAACAGAGAAA ACTTAGATGA

ES 2 651 612 T3

Amplificación	Cromosoma	Avance	Retroceso
NITT_016	cr 21	AAGACTCGGCAGCATCTCCATCAGGGTAGA GAATCAGAATTGG	GCGATCGTCACTGTTCTCCACAGAGATCAACC GGAGAAAGTAAA
NITT_076	cr 21	AAGACTCGGCAGCATCTCCACCACGGATCC ACTGCATA	GCGATCGTCACTGTTCTCCAGTTCTCTGTAAGT GAAAGCATCCTAAA
NITT_057	cr 21	AAGACTCGGCAGCATCTCCATCTGGTCTAAA TAAAGTCTTCACATCA	GCGATCGTCACTGTTCTCCAGAGGTAGGAAAT GCACCATCA
NITT_020	cr 21	AAGACTCGGCAGCATCTCCACAGAGGCCAT GCCAGTAGT	GCGATCGTCACTGTTCTCCAGAAAGTCTGGGA GGTTGAAGC
NITT_039	cr 21	AAGACTCGGCAGCATCTCCATGCCATCAGAA CCCGTAAA	GCGATCGTCACTGTTCTCCACACACAGAAGCA CAGGAAAAATC
NITT_059	cr 21	AAGACTCGGCAGCATCTCCACCTTCTCTGCC TCCATTCTAGT	GCGATCGTCACTGTTCTCCATAGTGTCCGATAA TGAAGAACAGTAAA
NITT_053	cr 21	AAGACTCGGCAGCATCTCCATGGCTAAGCA CATACCCTTAAA	GCGATCGTCACTGTTCTCCACTGACACAAATG AAGGCCAAAA
NITT_044	cr 21	AAGACTCGGCAGCATCTCCATCTCCATTCT TCTGCTCTTAGT	GCGATCGTCACTGTTCTCCAGAACTCACTCTG GAAGCAATGA
NITT_072	cr 21	AAGACTCGGCAGCATCTCCAGAAAGCTGGG CGTATTGG	GCGATCGTCACTGTTCTCCAGAACATTCTGAAC ATCTGGAA TGA

2. Principio del ensayo MASTR

5 Los pares de cebadores se probaron primero en reacciones de PCR simples en 20 ng de ADN genómico usando 10 pmol por cebador; los otros parámetros fueron iguales a los de la PCR multiplex. Las reacciones de PCR múltiple se realizaron con 50 ng de ADN genómico en una reacción de 25 µl que contenía tampón de PCR Taq Titanium™ (Clontech, Palo Alto, CA) con una concentración final de 0.25 mM para cada dNTP (Invitrogen, Carlsbad, CA) y un total de 0.125 µl de ADN polimerasa Taq Titanium™ (Clontech). Las concentraciones de cebador se optimizaron y variaron entre 0.05 pmol/ml y 0.2 pmol/ml de concentración final.

10 El ensayo final múltiple (ensayo MASTR) se utilizó para amplificar todas las muestras de ADN. La primera reacción de PCR se realizó en 50 ng de ADN con los siguientes ajustes: desnaturalización inicial de la muestra 10 min a 95°C seguida de 20 ciclos cada uno que consta de: 45 seg a 95°C, 45 seg a 60°C y 2 min a 68°C que finaliza con un paso de extensión final de 10 min a 72°C (consulte la Figura 5).

15 Los fragmentos de PCR resultantes se diluyeron 1000 veces seguido de una segunda etapa de PCR para incorporar el código de barras individual. Las condiciones de PCR de este paso son idénticas a las condiciones del primer paso de PCR (ver Figura 6).

Los amplicones con código de barras resultantes se mezclan equimolarmente y se usan en una reacción de PCR en emulsión según lo descrito por el fabricante (Roche Diagnostics).

Referencias

5 ¹ Lo Y, Corbetta N, Chamberlain P, Rai V, Sargent I, Redman C, and Wainscoat J (1997) Presence of fetal DNA in maternal plasma and serum. *The Lancet* 350: 485-487

² Poon L, Leung T, Lau T, Lo Y (2000) Presence of fetal RNA in maternal plasma. *Clin Chem* 46: 1832-1834

³ Ng E, Tsui N, Lau T, Leung T, Chiu R, Panesar N, et al. (2003) mRNA of placental origin is readily detectable in maternal plasma. *PNAS* 100: 4748-4753

10 ⁴ Lo Y, Tsui N, Chiu R, Lau T, Leung T, Heung M, et al. (2007) Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. *Nat Med* 13:218-23

⁵ Tsui N,2, Wong B, Leung T, Lau T, Chiu R and Lo Y (2009) Non-invasive prenatal detection of fetal trisomy 18 by RNA-SNP allelic ratio analysis using maternal plasma SERPINB2 mRNA: a feasibility study. *Prenat Diagn* 29: 1031-1037

15 ⁶ Yang Y, Ding J, Lee M, Loria O, Mohsenian F, Tang M, et al. (2008) Identification of mRNA-SNP markers for a noninvasive prenatal trisomy 21 (T21) test. *Prenat Diagn* 2008: 28-S12

⁷ Chim S, Tong Y, Chiu R, Lau T, Leung T, Chan L, et al. (2005) Detection of the placental epigenetic signature of the maspin gene in maternal plasma. *PNAS* 102: 14753- 14758

20 ⁸ Chan K, Ding C, Gerovassili A, Yeung S, Chiu R, Leung T et al. (2006) Hypermethylated RASSF1A in Maternal Plasma: A Universal Fetal DNA Marker that Improves the Reliability of Noninvasive Prenatal Diagnosis. *Clin Chem* 52: 2211-2218

⁹ Lo D, Chan A, Sun H, Chen E, Jiang P, Lun F et al. (2010) Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. *Sci Transl Med* 2:6

¹⁰ Dhallan R, Guo X, Emche S, Damewood M, Bayliss P, Cronin M et al. (2007) A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. *Lancet* 369: 474-481

25 ¹¹ Lo D, Lun F, Chan A, Tsui Y, Chong C, Lau T, et al. (2007) Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *PNAS* 104:13116-131121

¹² Ehrich M, Deciu C, Zwielfelhofer T; Tynan J, Cagasan L, Tim R et al. (2011) Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. *Am J Obstet Gynecol* 204:205.e1-11

30 ¹³ Sehnert A, Rhees B, Comstock D, de Feo E, Heilek G,1 Burke J and Raval P (2011) Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood. *Clin Chem* 57: 1042-1047

¹⁴ Chen E, Chiu R, Sun H, Akolekar R, Chan A, Leung T et al. (2011) Noninvasive Prenatal Diagnosis of Fetal Trisomy 18 and Trisomy 13 by Maternal Plasma DNA Sequencing. *PLoS ONE* 6: e21791

35 ¹⁵ Liao G, Lun F, Zheng Y, Chan A, Leung T, Lau T et al. (2011) Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles. *Clin Chem* 57: 92-101

REIVINDICACIONES

1. Un método para la detección de una aneuploidía cromosómica fetal en una mujer embarazada que comprende
- i) preparar ADN libre flotante a partir de una muestra biológica de dicha mujer embarazada,
 - 5 ii) amplificar un conjunto seleccionado de secuencias de ADN diana de uno o más cromosomas que se presume que son aneuploides y amplificar un conjunto seleccionado de secuencias de ADN diana de uno o más cromosomas que se presume que son euploides en al menos una reacción de PCR múltiplex cuantitativa en la que al menos una secuencia de ADN amplificada comprende al menos un SNP para el que la mujer es heterocigótica en la que las secuencias de ADN amplificadas tienen longitudes inferiores a 140 pares de bases,
 - 10 iii) secuenciar las secuencias de ADN diana amplificadas y
 - 15 iv) calcular la suma de recuentos de lectura para todas las secuencias de ADN amplificadas de una aneuploidía cromosómica sospechosa seguida de normalización, contra la suma de recuentos de lectura para todas las secuencias de ADN amplificadas de un cromosoma euploide de referencia para determinar mediante métodos estadísticos una puntuación establecida indicativa de la presencia de una aneuploidía cromosómica fetal y/o determinación de las proporciones alélicas de los SNP informativos en los que una relación alélica distorsionada es indicativa de la presencia de una aneuploidía cromosómica fetal en dicha mujer embarazada
- en el que dicha aneuploidía cromosómica fetal es el cromosoma 13 y/o el cromosoma 18 y/o el cromosoma 21 y/o el cromosoma X y/o el cromosoma Y
2. Un método de acuerdo con la reivindicación 1, en el que dicha muestra biológica es sangre materna, plasma, orina, fluido cerebroespinal, suero, saliva o es fluido de lavado transcervical.
- 20 3. Un método de acuerdo con las reivindicaciones 1 o 2, en el que se usan productos químicos estabilizadores de células en la preparación de dicho ADN libre flotante.
4. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que el contenido de GC de las secuencias de ADN diana está entre 20 y 70%.
- 25 5. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 4, en el que se usa un fraccionamiento por tamaño en la preparación del ADN libre flotante.
6. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 5, en el que las secuencias de ADN amplificadas se obtienen en una sola reacción de PCR.
7. Un método de acuerdo con la reivindicación 6, en el que en dicha única reacción de PCR múltiple se obtienen más de 40 secuencias de ADN amplificadas.
- 30 8. Un método de acuerdo con la reivindicación 6, en el que en dicha única reacción de PCR múltiple se obtienen más de 60 secuencias de ADN amplificadas.

Figura 1

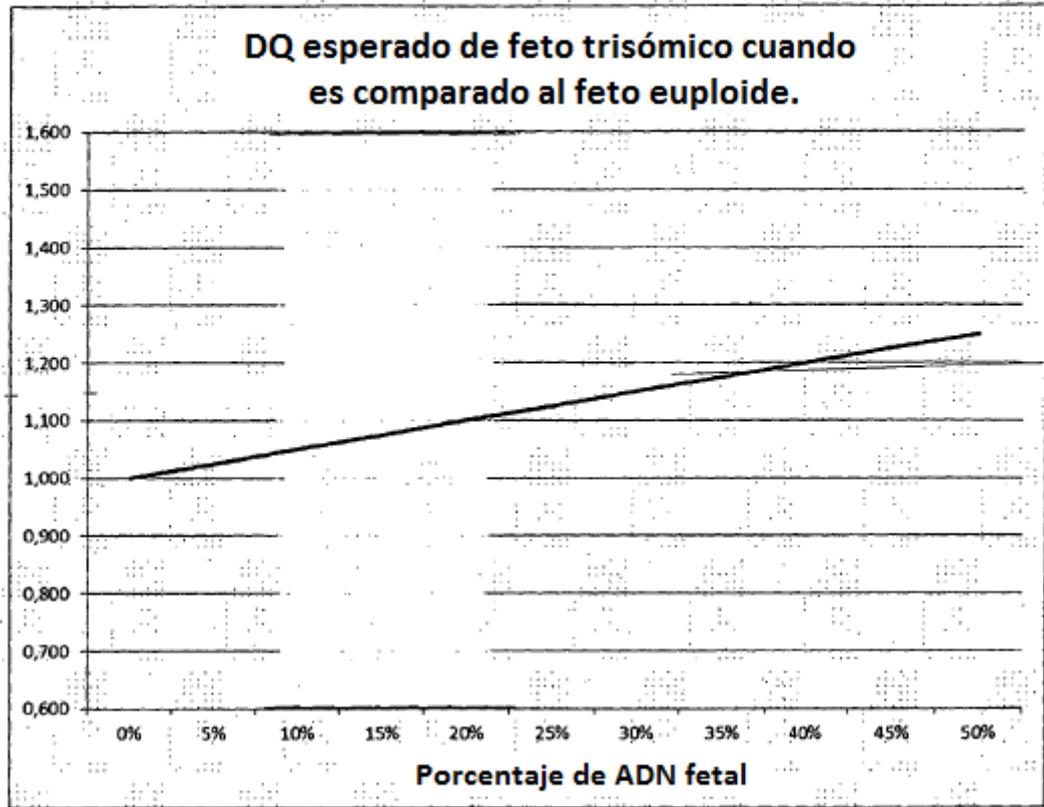


Figura 2

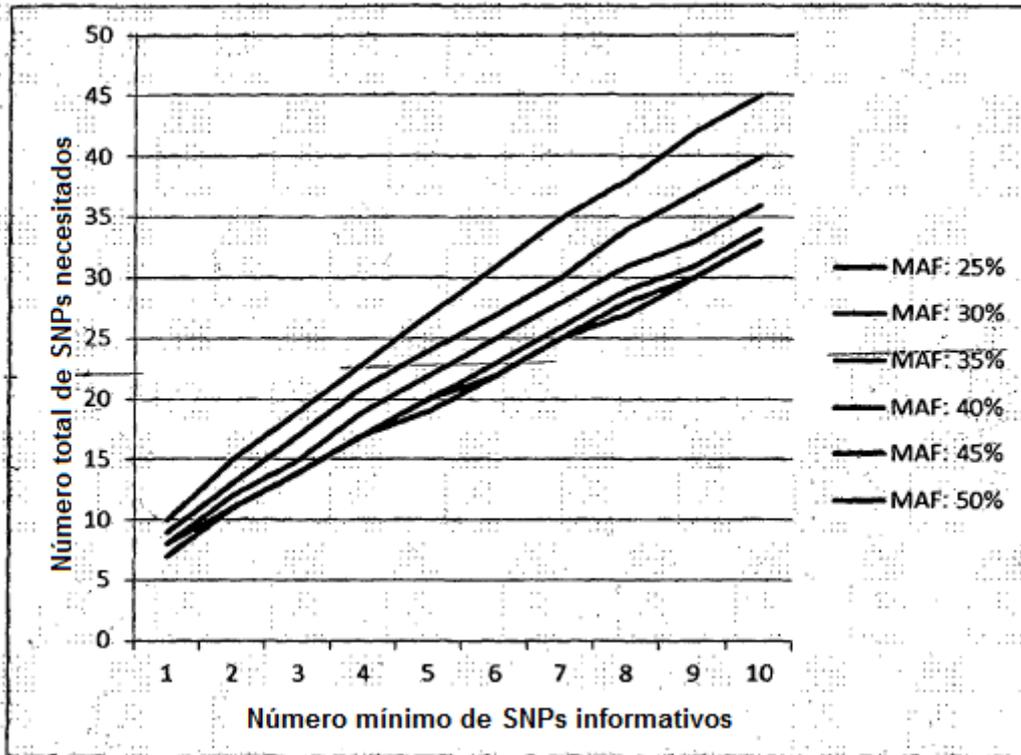


Figura 3

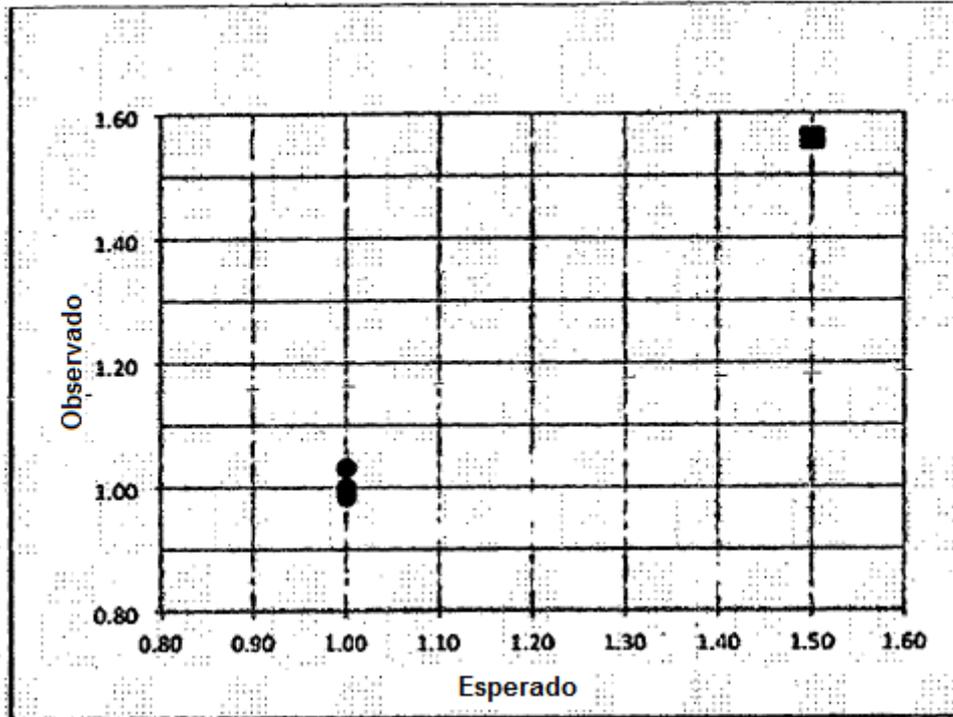


Figura 4

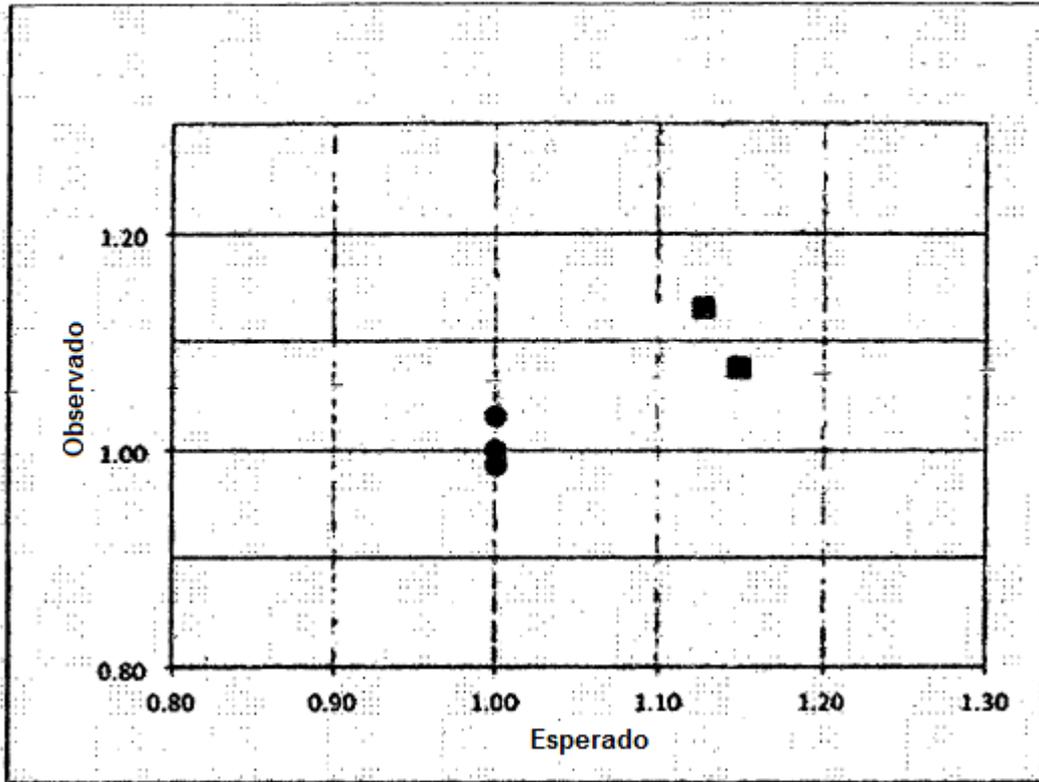


Figura 5

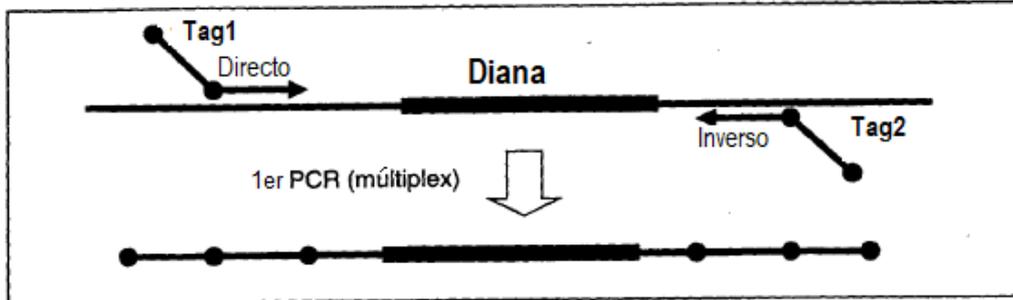


Figura 6

