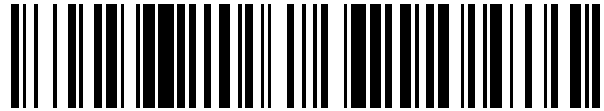


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 564 656**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.10.2010 E 10768959 (8)**

97 Fecha y número de publicación de la concesión europea: **23.12.2015 EP 2494065**

54 Título: **Medios y métodos para el diagnóstico no invasivo de la aneuploidía cromosómica**

30 Prioridad:

26.10.2009 US 272722 P

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

28.03.2016

73 Titular/es:

**LIFECODEXX AG (100.0%)
Jakob-Stadler Platz 7
78467 Konstanz, DE**

72 Inventor/es:

**BENZ, MARCUS;
HOFMANN, WERA;
POHL, THOMAS y
VON KALLE, CHRISTOPH**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 564 656 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

Medios y métodos para el diagnóstico no invasivo de la aneuploidía cromosómica

- 5 La presente invención se refiere a un método de detección prenatal no invasivo en una muestra biológica obtenida de una mujer gestante. En particular, la invención se refiere a un método para el diagnóstico prenatal no invasivo (DPNI) mediante enriquecimiento y cuantificación de secuencias seleccionadas de ácido desoxirribonucleico libres de células en una muestra materna de sangre, con el fin de detectar anomalías cromosómicas fetales, por ejemplo aneuploidías cromosómicas fetales.
- 10 En muchos países el diagnóstico prenatal de aneuploidías fetales, así como de trastornos mendelianos de genes individuales, tales como la fibrosis quística, típicamente constituyen una parte de los programas de salud pública. La atención obstétrica incluye en particular los ensayos prenatales no invasivos, tales como el cribado en el primer trimestre o en el segundo trimestre para estratificar según riesgo las mujeres gestantes previamente a los procedimientos diagnósticos invasivos. Dichos ensayos de cribado habitualmente incluyen un análisis de sangre que estima algunos marcadores bioquímicos y un examen de ultrasonidos para medir la translucencia nucal fetal con una precisión de hasta 90% (Malone et al., N. Eng. J. Med. 353:2001-2011, 2005). Los procedimientos invasivos, incluyendo la amniocentesis, el muestreo de la vellosidad coriónica o la cordocentesis, se utilizan para obtener material genético fetal directo (y realizar el análisis del cariotipo) y conseguir una precisión superior al 99%. Sin embargo, dichos procedimientos invasivos plantean un riesgo potencial de pérdida del feto. Actualmente existe un gran interés en el desarrollo de métodos para el diagnóstico directo pero no invasivo de las enfermedades genéticas fetales sin riesgo de aborto.
- 15 El descubrimiento de los ácidos nucleicos fetales libres de células en el plasma materno ha abierto nuevas posibilidades para el diagnóstico prenatal no invasivo (Lo YMD et al., Lancet 350:485-487, 1997; Lo YMD y Chiu, RWK, Nat. Rev. Genet. 8:71-77, 2007). Durante los últimos pocos años, este enfoque se ha aplicado al diagnóstico prenatal de trastornos ligados al sexo (Costa J.M. et al., N. Engl. J. Med. 346:1502, 2002) y determinados trastornos de genes individuales (Lo Y.M.D. et al., N. Eng. J. Med. 339:1734-1738, 1998). Otras aplicaciones implican el enriquecimiento en ADN fetal mediante un método que suprime el fondo materno, tal como mediante la adición de formaldehído (Dhallan R. et al., JAMA291:1114-1119, 2004) o el reconocimiento de moléculas de ácidos nucleicos específicamente fetales, incluyendo marcadores epigenéticos específicamente fetales y marcadores de ARNm específicos de la placenta (Chan K.C. et al., Clin. Chem. 52:2211-8, 2006).
- 20 Los últimos informes han indicado que el desarrollo de métodos cuantitativos altamente discriminatorios para el análisis de dosis cromosómicas utilizando tecnología digital de reacción en cadena de la polimerasa podrían resultar útiles para la detección no invasiva de aneuploidías fetales mediante el análisis de ADN y ARN libre de células en el plasma o suero materno (Fan H.C. y Quake S.R., Anal. Chem. 79:7576-7579, 2007; Lo Y.M.D. et al., PNAS 104:13116-13121, 2007). Sin embargo, la fracción reducida de ácidos nucleicos fetales que coexisten en el plasma materno con el elevado fondo de ácidos nucleicos de origen materno con frecuencia puede interferir con dichos análisis (Lo Y.M.D. et al., Am. J. Hum. Genet. 62:768-775, 1998).
- 25 Con el rápido desarrollo de la tecnología de secuenciación de alto rendimiento, que permite la secuenciación masiva en paralelo de decenas de millones de etiquetas de secuenciación cortas, se ha explorado recientemente la posibilidad de detectar la presencia de genomas fetales trisómicos en la muestra de ADN de plasma materno obtenida durante el primer trimestre del embarazo (Fan H.C. et al., PNAS 105:16266-16271, 2008; Chiu R.W. et al., PNAS 105:20458-20463, 2008). Dicha tecnología de secuenciación "shotgun" (pistola génica) permite un muestreo más profundo del que se consigue mediante PCR digital. Sin embargo, los costes de la secuenciación del ADN utilizando la secuenciación "shotgun" son excesivos para que este enfoque se convierta en un ensayo médico rutinario como parte de la atención prenatal.
- 30 Además, una desventaja técnica de la secuenciación aleatoria con pistola es el riesgo potencial de introducción de sesgos de secuencia. La secuenciación aleatoria "shotgun" no puede producir una distribución de lectura perfectamente uniforme en todo el genoma de interés, por lo que estadísticamente siempre existirán algunos tramos de ADN que se lean con mayor frecuencia que otros.
- 35 El documento nº WO 2009/013496 se refiere a un método no invasivo de ensayo diagnóstico prenatal de una aneuploidía cromosómica fetal mediante la determinación de los desequilibrios entre diferentes secuencias de ácidos nucleicos en una muestra materna mediante secuenciación genómica masiva en paralelo (SGMP).
- 40 De acuerdo con lo anteriormente expuesto, existe una fuerte necesidad de un método no invasivo menos costoso y menos laborioso para el diagnóstico prenatal de una anomalía cromosómica fetal, por ejemplo una aneuploidía cromosómica fetal, en una muestra biológica obtenida de una mujer gestante, el cual presente una sensibilidad y especificidad elevadas y no introduzca ningún sesgo de secuencia adicional.
- 45 De esta manera, es un objetivo de la presente invención proporcionar un método mejorado alternativo para llevar a cabo el diagnóstico prenatal.

Más concretamente, la invención se refiere a un método para la determinación de una aneuploidía cromosómica fetal en una muestra biológica obtenida de un individuo hembra gestante, en el que la muestra biológica incluye moléculas de ácidos nucleicos, comprendiendo el método:

- 5 a) seleccionar y aislar a partir de una muestra biológica de un individuo hembra gestante una o más secuencias diana de moléculas de ADN contenidas en la muestra biológica, en la que dichas secuencias diana comprenden secuencias de ADN que presentan regiones de unión a nucleosoma de consenso y en la que dichas moléculas de ADN que comprenden las secuencias diana se aíslan y se enriquecen utilizando cromatografía de afinidad o enriquecimiento por lotes utilizando cebos de ácidos nucleicos marcados, los cuales permiten el enriquecimiento y el
- 10 aislamiento de las moléculas de ADN hibridadas mediante purificación por afinidad,
- b) amplificar dichas secuencias diana seleccionadas,
- c) secuenciar dichas secuencias diana seleccionadas y amplificadas y asignarlas a los cromosomas del genoma e identificar las secuencias diana asignadas únicas,
- 15 d) determinar una primera cantidad de cada uno de uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas,
- e) determinar una segunda cantidad de cada uno de uno o más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho segundo o segundos cromosomas, y
- 20 f) determinar basándose en dicha primera y segunda cantidades una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.

El método anteriormente indicado puede comprender o no etapas adicionales antes, después o entre las etapas esenciales mencionadas explícitamente del método. Preferentemente, el método comprende las etapas de obtener la muestra biológica de un individuo hembra gestante. Además, el método puede comprender preferentemente etapas de pretratamiento de las muestras que, por ejemplo, están destinadas a conservar los ácidos nucleicos presentes en la muestra o que están destinadas a liberar o a aislar los ácidos nucleicos respecto del material celular presente en la muestra. El método puede, preferentemente, estar asistido por automatización. Las etapas a) a c) pueden estar asistidas por un dispositivo robótico; las etapas restantes que evalúan los datos de secuencias pueden estar asistidas por un ordenador o dispositivo de procesamiento de datos que presenta tangiblemente incluido un algoritmo para determinar las cantidades de las secuencias y para llevar a cabo la comparación.

La expresión "aneuploidía cromosómica fetal" tal como se utiliza en la presente memoria se refiere a una aberración cromosómica que se caracteriza por varios cromosomas que difieren del número fisiológico de cromosomas que se espera que se encuentren presentes en un feto, es decir, un número anormal de cromosomas. Dicha diferencia en el número también se denomina aberración numérica. La expresión comprende un número incrementado de un cromosoma, así como un número reducido de un cromosoma presente en el genoma fetal. Una aneuploidía cromosómica tal como se denomina en el contexto de la presente invención puede ser: (i) una monosomía, es decir, la falta de un cromosoma de una pareja de cromosomas, (ii) una disomía en los casos en los que el individuo que debe examinarse normalmente es triploide, tetraploide o incluso más, o una disomía uniparental, (ii) una trisomía, es decir, la presencia de tres en lugar de dos cromosomas que se encuentran presentes fisiológicamente para cada cromosoma en un organismo diploide, (iv) una tetrasomía, es decir, la presencia de cuatro cromosomas en lugar de dos cromosomas que son los presentes fisiológicamente para cada cromosoma en un organismo diploide, o incluso (v) una pentasomía, es decir, la presencia de cinco cromosomas en lugar de los dos cromosomas que se encuentran presentes fisiológicamente para cada cromosoma en un organismo diploide. La aneuploidía puede afectar tanto a los autosomas como a los gonosomas. En algunos casos, una pareja de cromosomas puede faltar por completo en el genoma. Además, una aneuploidía tal como se utiliza el término en la presente memoria comprende además una aberración numérica con respecto a una parte de un cromosoma, es decir, un caso en el que una parte de uno de los cromosomas de una pareja de cromosomas ha sido delecionada o se encuentra presente en una o más copias adicionales. Una parte de un cromosoma puede encontrarse presente en un número inapropiado de copias como resultado de una aberración estructural, por ejemplo traslocaciones desequilibradas, síndromes de genes contiguos o síndromes de deleciones cromosómicas. La aneuploidía cromosómica fetal habitualmente resulta en trastornos o enfermedades graves que afectan al feto y a su desarrollo. Entre las aneuploidías cromosómicas fetales y enfermedades o trastornos acompañantes que resultan preferentes que pueden determinarse mediante el método de la presente invención se incluyen el síndrome de Turner (monosomía gonosómica), el síndrome de Klinefelter (gonosomas XXY), el síndrome de la triple X (gonosomas XXX), el síndrome de Down (trisomía del par 21), el síndrome de Edwards (trisomía del par 18) o el síndrome de Patau (trisomía del par 13). La disomía uniparental es conocida para el cromosoma 15 y se conoce como síndrome de Prader-Willi. En el caso de que una disomía uniparental de este tipo deba detectarse, el ADN analizado también distinguirse también en ADN heredado paternamente o ADN heredado maternamente. Las traslocaciones desequilibradas tal como se utiliza la expresión en la presente memoria comprende, preferentemente, las trisomías de Robertson desequilibradas, rob(13q;14q). Entre otras aberraciones estructurales que pueden determinarse preferentemente mediante el método de la invención se incluyen la deleción de 4q (síndrome de Wolf-Hirschhorn), la deleción de 5q (síndrome de cri du chat) o los síndromes de microdeleción, en particular la deleción 17q11.2 (síndrome de Smith-Magenis) o la deleción 22q11.2 (síndrome de DiGeorge).

La expresión "determinación de una aneuploidía cromosómica fetal" tal como se utiliza en la presente memoria se refiere a evaluar la probabilidad según la cual un sujeto sufre o sufrirá de una enfermedad o condición a la que se

hace referencia en la presente memoria. Tal como entenderá el experto en la materia, dicha evaluación habitualmente no se pretende que sea correcta para 100% de los sujetos que deben ser diagnosticados. Sin embargo, la expresión requiere que una parte estadísticamente significativa de sujetos pueda diagnosticarse correctamente que sufre de la enfermedad o condición. Si una parte es estadísticamente significativa o no puede ser determinado sin hacer ninguna otra referencia por el experto en la materia utilizando diversas herramientas de evaluación estadística bien conocidos, por ejemplo la determinación de intervalos de confianza, la determinación del valor de p, la prueba t de Student, la prueba de Mann-Whitney, etc. Se encontrará más información en Dowdy y Wearden, *Statistics for Research*, John Wiley & Sons, New York 1983. Los intervalos de confianza preferentes son por lo menos 90%, por lo menos 95%, por lo menos 97%, por lo menos 98% o por lo menos 99%. Los valores de p preferentemente son 0,1, 0,05, 0,01, 0,005 o 0,0001. Preferentemente, la probabilidad contemplada por la presente invención permite que el diagnóstico sea correcto para por lo menos 60%, por lo menos 70%, por lo menos 80% o por lo menos 90% de los sujetos de una cohorte o población dada.

El término "individuo" tal como se utiliza en la presente memoria se refiere a un organismo multicelular que es capaz de portar su progenie en el interior de un útero. Más preferentemente, dicho organismo se refiere a organismos eutéricos y, más preferentemente, a mamíferos. Todavía más preferentemente, el individuo es un ser humano. Se entenderá que el individuo del que se utiliza una muestra en el método de la presente invención será una hembra gestante que porta un embrión o feto en el interior de su útero. Durante la gestación y en particular en las etapas fetales del desarrollo embrionario, se libera ADN genómico fetal al cuerpo del individuo hembra. En particular, se observa dicho ADN genómico fetal en la sangre del individuo hembra.

De esta manera, la expresión "muestra biológica" tal como se utiliza en la presente memoria se refiere a una muestra de un individuo, preferentemente a una muestra de un líquido corporal obtenido de una mujer gestante. Más preferentemente, dicho líquido corporal se selecciona de entre el grupo que consiste de una muestra de sangre, una muestra de orina y una muestra de saliva. La muestra de sangre a la que se hace referencia en la presente memoria comprende una muestra de sangre completa, una muestra de plasma o una muestra de suero. Se entenderá que la muestra que debe utilizarse en el método de la presente invención comprenderá ácidos nucleicos y, en particular, moléculas de ADN que son representativas del genoma del feto en desarrollo. Preferentemente, dichas moléculas de ADN derivadas del genoma fetal representan cada cromosoma presente en dicho genoma fetal de una manera estadística. De esta manera, cuantas más copias de un cromosoma dado se encuentren presentes en el genoma fetal, más moléculas de ADN reflejarán estos cromosomas en la muestra, y viceversa.

Las moléculas de ADN presentes en la muestra, incluyendo las moléculas de ADN derivadas del genoma fetal, se seleccionan específicamente para las moléculas de ADN que comprenden secuencias diana que permite un análisis de las secuencias significativo y eficiente en el contexto de la presente invención. Las secuencias diana adecuadas comprenden secuencias de consenso para regiones de unión a nucleosoma tal como se indica en otros sitios de la presente memoria o regiones de unión a la matriz del ADN cromosómico. La expresión "región de unión a nucleosoma de consenso" tal como se utiliza en la presente memoria se refiere a patrones de secuencias de ADN en o en torno a sitios de inicio de transcripción (SIT) de genes transcritos que están asociados a nucleosomas bien situados que se encuentran localizados en o próximos a dichos sitios de inicio de transcripción. Las secuencias de consenso adecuadas son conocidas de la técnica y se describen en Schones et al. o en Fan et al. (Schones Cell 132(5): 887-898; Fan, Proc. Natl. Acad. Sci. USA 105: 16266-16271, 2008).

La selección y aislamiento de dichas moléculas de ADN que comprenden secuencias diana para el análisis de secuencias tal como se utilizan en la presente memoria incluyen la provisión de secuencias adecuadas para cebadores utilizados para la amplificación del ADN. Preferentemente se proporcionan secuencias adecuadas mediante la ligación de oligonucleótidos adaptadores que comprenden secuencias adecuadas para la amplificación del ADN a las moléculas de ADN. Preferentemente, los oligonucleótidos adaptadores pueden comprender además secuencias adicionales, preferentemente secuencias de etiqueta de código de barras o de índice que permiten la identificación de un grupo de especies idénticas de moléculas de ADN (ver Parameswaran, Nucl. Acids Res. 35(19): e130, 2007). Mediante la utilización de dichos oligonucleótidos adaptadores que comprenden secuencias de etiqueta de código de barras o de índice, resulta posible procesar un grupo de moléculas de ADN de diferentes muestras en el método de la presente invención en paralelo. Además, el término comprende además procedimientos de aislamiento y enriquecimiento de moléculas de ADN que comprenden dichas secuencias diana utilizando, por ejemplo el enriquecimiento mediante cromatografía de afinidad o por lotes utilizando cebadores de ácidos nucleicos marcados, que permite el enriquecimiento y aislamiento de moléculas de ADN hibridadas mediante purificación por afinidad, por ejemplo mediante perlas magnéticas. En una realización preferente del método de la presente invención, las moléculas de ARN cebo biotiniladas se ponen en contacto con las moléculas de ADN comprendidas en la muestra. Las moléculas de ARN cebo están diseñadas específicamente para presentar una secuencia que es capaz de hibridarse con la secuencia diana de las moléculas de ADN deseadas, en particular presentan secuencias capaces de hibridarse con las secuencias de consenso para regiones de unión a nucleosoma tal como se indica en otros sitios de la presente memoria o regiones de unión a matriz del ADN cromosómico. Tras la hibridación, las moléculas híbridas quiméricas de ADN/ARN se aíslan de la muestra restante mediante captura del marcaje de biotina mediante estreptavidina. Preferentemente pueden aplicarse perlas magnéticas con estreptavidina inmovilizada para el aislamiento. Las moléculas de ADN aisladas pueden liberarse de las perlas magnéticas. Dependiendo de cómo se lleva a cabo la liberación, también puede llevarse a cabo el enriquecimiento de las

moléculas de ADN. Una técnica adecuada particular para el aislamiento y el enriquecimiento de las moléculas de ADN deseadas es el sistema disponible comercialmente SureSelect Target Enrichment System, de Agilent Technologies, Inc., US. También se proporciona más información en Gnirke, Nat. Biotechnology 27(2): 182-189, 2009. A continuación, dichas moléculas de ADN aisladas que comprenden la secuencia diana pueden procesarse adicionalmente según el método de la invención. Otra técnica particular es la tecnología de secuenciación mediante síntesis del ADN en cuatro colores, de Illumina, Inc., US (ver Johnson, Science 316: 1497-1502, 2007; Robertson, Nature Methods 4: 651-657, 2007; Barski, Cell 129: 823-837, 2007; Mikkelsen, Nature 448: 553-560, 2007; Fields, Science 316: 1441-1442, 2007).

La amplificación de la secuencia diana que comprende moléculas de ADN puede conseguirse mediante métodos bien conocidos de amplificación de los ácidos nucleicos. Preferentemente la amplificación se lleva a cabo mediante PCR. La secuenciación de los productos de amplificación puede llevarse a cabo mediante técnicas bien conocidas de secuenciación de ácidos nucleicos. La calidad de las muestras analizadas puede garantizarse después de la amplificación. Además, las moléculas de ADN en la muestra pueden clasificarse en moléculas de ADN con secuencias de origen paterno o materno.

Preferentemente, las secuencias paternas se analizan al analizar los genomas fetales mediante el método de la presente invención. Mediante el análisis de dichas moléculas de ADN de herencia paterna, la fiabilidad del método puede mejorarse adicionalmente, ya que estas secuencias se originan inequívocamente del ADN genómico fetal.

Las secuencias identificadas mediante la secuenciación de las moléculas de ADN amplificadas que comprenden las secuencias diana deseadas pueden asignarse a cromosomas o partes de los mismos por la presencia de secuencias características de dichos cromosomas. Las secuencias adecuadas que permiten la asignación son secuencias genéticas marcadoras, por ejemplo marcadores de polimorfismo de secuencia única, secuencias de marcador de ADN repetitivo tales como repeticiones de secuencias únicas, repeticiones en tándem cortas, repeticiones en tándem de número variable o secuencias de microsatélite en general. Alternativamente, la asignación puede llevarse a cabo mediante la comparación de las secuencias con secuencias de consenso para cromosomas, tales como las secuencias de consenso hg18 o hg19 para el genoma humano. Tras asignar las secuencias diana a un cromosoma, es decir, tras localizar las secuencias diana en los cromosomas, se identifican las secuencias diana asignadas únicas, es decir, las secuencias diana que han sido asignadas únicamente una vez a un cromosoma, de entre las secuencias diana localizadas. Entre los algoritmos adecuados para la comparación de secuencias y las asignaciones a cromosomas se incluyen la alineación local eficiente de datos de nucleótidos (ELAND, por sus siglas en inglés, una herramienta de alineación integrada en el paquete de procesamiento de datos Illumina-Solexa, que permite la alineación sin huecos para lecturas con tamaños de hasta 32 pb; Cox, no publicado) o el alineador de Burrows-Wheeler (BWA, por sus siglas en inglés; Li y Durbin, Bioinformatics 25:1754-1760, 2009).

En una etapa posterior, se determina una primera cantidad de cada uno de uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas. Dicha cantidad puede ser cualquier número o valor que refleje o se correlacione con el número de secuencias diana únicas que se asignan a un cromosoma que se han obtenido de las moléculas de ADN. La cantidad puede ser una cantidad absoluta o relativa, es decir, una cantidad normalizada, por ejemplo con respecto al número total de secuencias diana amplificadas y/o la cantidad total de ácidos nucleicos observada en la muestra. Simultánea o separadamente, se determina una segunda cantidad de cada uno de uno o más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho segundo o segundos cromosomas.

Basándose en dichas primera y segunda cantidades, puede determinarse una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas. Lo anterior puede llevarse a cabo para, por ejemplo, un individuo diploide mediante la comparación de una primera cantidad relativa de un primer cromosoma con una segunda cantidad relativa de un segundo cromosoma. En el caso de un individuo diploide, la primera y segunda cantidades relativas son esencialmente idénticas. Una primera cantidad relativa incrementada con respecto a la segunda cantidad será indicativa de un número incrementado de primer cromosoma y, de esta manera, de una aneuploidía del primer cromosoma en el genoma fetal. Lo anterior es de aplicación, *mutatis mutandis*, a una primera cantidad reducida. Preferentemente, en el ser humano, pueden utilizarse los cromosomas 1, 2, 3, 5 a 15, 18 y 21 para derivar una segunda cantidad tal como se ha indicado anteriormente.

También preferentemente, puede determinarse un parámetro basándose en la primera y segunda cantidades. Dicho parámetro será estadísticamente más robusto y permitirá la comparación entre diferentes individuos o colectivos de individuos.

De esta manera, en una realización preferente del método de la presente invención, dicho método comprende además:

- i) determinar un parámetro a partir de dicha primera cantidad respecto a dicha segunda cantidad,

ii) comparar el parámetro con un valor de control de corte correspondiente y, basándose en la comparación, determinar si existe una diferencia o no que permita la predicción de una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.

5 El término "parámetro" tal como se utiliza en la presente memoria se refiere a un valor numérico que caracteriza un conjunto de datos cuantitativos y/o una relación numérica entre conjuntos de datos cuantitativos. Más concretamente, dicho parámetro puede calcularse como una proporción entre la primera cantidad y la segunda cantidad o puede ser un valor derivado a partir de dicha proporción, tal como un valor de porcentaje. Preferentemente, el parámetro refleja la cantidad de un primer cromosoma que debe investigarse con respecto a la cantidad o cantidades de por lo menos otro segundo cromosoma y, preferentemente, con respecto a los cromosomas restantes según se determinen por las secuencias diana. Además, el parámetro puede incluso hacerse más robusto con respecto a las variaciones individuales y, de esta manera, adecuado para la comparación con parámetros de individuos de referencia o grupos de los mismos. Con este fin, puede, preferentemente, restarse un parámetro que ha sido calculado tal como se ha indicado anteriormente (parámetro de referencia), respecto del parámetro determinado y dividirse por la desviación estándar para el parámetro de referencia. El parámetro transformado resultante, también denominado "puntuación z" (ver Chiu, Proc. Natl. Acad. Sci. USA 105(51):20458-63, 2008), seguidamente puede compararse con un parámetro transformado de un individuo de referencia o grupo de individuos que es conocido que muestran la aneuploidía cromosómica o que es conocido que no muestran dicha aneuploidía. De esta manera, el parámetro de referencia constituye un valor de corte y, basándose en la comparación, resulta posible determinar si existe una diferencia o no en el número de cromosomas del genoma fetal, permitiendo la predicción de una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas. La expresión "valor de corte" o "valor de control de corte" tal como se utiliza en la presente memoria, preferentemente se refiere a un valor numérico calculado como la proporción entre dos cantidades de cromosomas uno respecto a otro, es decir, la proporción entre la cantidad de uno o más primeros cromosomas y la cantidad de uno o más segundos cromosomas de una muestra promedio o agrupada de sangre, orina o saliva procedente de uno o más donantes hembra sanos, y en la que el número de dichos segundos cromosomas es idéntico al número de dichos primeros cromosomas con lo que debe compararse. El valor de corte permite distinguir entre un estado de enfermedad y de no enfermedad de una muestra biológica experimental. Son técnicas adicionales para determinar parámetros y compararlos entre diferentes individuos o grupos de los mismos que pueden aplicarse preferentemente en el método de la presente invención aquellas que han sido descritas en los documentos nº EP 2 183 692 A1, nº EP 2 183 693 A1 o nº WO 2010/033578.

Recientemente ha sido encontrado por Fan et al. (Fan, loc. cit., 2008) que el ADN en plasma libre de células es de origen principalmente apoptótico y que comparte características del ADN nucleosómico. De esta manera, según la presente invención, el método no invasivo de diagnóstico prenatal dado a conocer en el documento nº WO 2009/013496 se desarrolló y mejoró adicionalmente al sugerir la realización de una etapa adicional de enriquecimiento antes de la etapa de secuenciación para capturar y enriquecer específicamente regiones de ADN seleccionadas que muestran posiciones nucleosómicas de consenso en los cromosomas de interés. Mediante dicha modificación del método del estado de la técnica, puede mejorarse sustancialmente el coste y eficiencia del procedimiento, y particularmente en comparación con la secuenciación genómica masiva en paralelo utilizando la secuenciación aleatoria "shotgun". El presente procedimiento de enriquecimiento de la diana ofrece un incremento significativo de la velocidad y ventajas adicionales de escalabilidad respecto a las técnicas actuales de reacción en cadena de la polimerasa (PCR) en la mayoría de los flujos de trabajo de secuenciación de la siguiente generación. Se basa en una técnica eficiente de selección de híbridos que mejora el coste y eficiencia del procedimiento del flujo de trabajo denominado de "secuenciación genómica masivamente en paralelo" (Gnirke A. et al., Nat. Biotechnol. 27:182-189, 2009). El análisis de múltiples grupos de loci diana específicos se utiliza para incrementar la cantidad de datos obtenibles de una muestra sin incrementar el número de muestreos de PCR (digitales) realizados para la secuenciación genómica masivamente en paralelo posterior, tal como la alcanzable con la tecnología 454 (Roche).

El método según la presente invención resulta adecuado para determinar en una muestra de sangre procedente de un individuo gestante la aneuploidía cromosómica de la progenie en desarrollo. En particular, puede determinarse una aneuploidía cromosómica de un feto en desarrollo, particularmente desde la sexta semana de gestación en adelante.

Se da a conocer además un método para la determinación de una aberración cromosómica en una muestra biológica obtenida de un individuo, en el que la muestra biológica incluye moléculas de ácidos nucleicos, comprendiendo el método:

- a) seleccionar y aislar a partir de dicha muestra biológica de un individuo una o más secuencias diana de moléculas de ADN contenidas en la muestra biológica, en la que dichas secuencias diana comprenden secuencias de ADN que presentan regiones de unión a nucleosoma de consenso,
- b) amplificar dichas secuencias diana seleccionadas,
- c) secuenciar dichas secuencias diana seleccionadas y amplificadas y asignarlas a los cromosomas del genoma e identificar las secuencias diana asignadas únicas,
- d) determinar una primera cantidad de cada uno de uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas,

- e) determinar una segunda cantidad de cada uno de uno más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho segundo o segundos cromosomas, y
- f) determinar, basándose en dicha primera y segunda cantidades, una aberración cromosómica.

5 La expresión "aberración cromosómica" tal como se utiliza en la presente memoria preferentemente se refiere a aberraciones numéricas tal como se definen en otros sitios de la presente memoria.

Preferentemente la etapa f) de dicho método comprende además:

- 10 i) determinar un parámetro a partir de dicha primera cantidad respecto a dicha segunda cantidad,
- ii) comparar el parámetro con un valor de control de corte correspondiente y, basándose en la comparación, determinar si existe o no una diferencia en la que la presencia de una diferencia es indicativa de una aberración cromosómica en uno o más de dichos primeros cromosomas.

15 Más preferentemente, la aberración cromosómica es indicativa de una predisposición al cáncer.

El término "predisposición" tal como se utiliza en la presente memoria se refiere a un riesgo incrementado de que un individuo desarrolle cáncer dentro de una ventana de predicción en el futuro. Preferentemente, dicha ventana es de hasta 1 año, de hasta 5 años, de hasta 10 años o el periodo de vida entero del individuo.

20 El término "cáncer" tal como se utiliza en la presente memoria comprende cualquier neoplasma maligno. Son bien conocidos de la técnica diversos tipos de cáncer. Preferentemente el cáncer al que se hace referencia según la presente exposición se asocia a aberraciones cromosómicas. En particular, los cánceres siguientes se ha informado que se asocian frecuentemente a aberraciones cromosómicas, seleccionados de entre el grupo que consiste de:
25 leucemias humanas, tales como la leucemia mieloide aguda o la leucemia mieloide crónica (ver, por ejemplo, Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer, Mitelman F, Johansson B and Mertens F (eds.), 2010, <http://cgap.nci.nih.gov/Chromosomes/Mitelman>).

30 Preferentemente, dicho individuo en el contexto del método anteriormente indicado es un individuo tal como se define en otros sitios de la presente memoria y, más preferentemente, es un ser humano.

Preferentemente, dicha muestra biológica comprende una célula que se sospecha que es una célula de cáncer. Pueden obtenerse muestras adecuadas de líquidos corporales en el caso de tumores no sólidos, por ejemplo sangre en el caso de los cánceres hematopoyéticos, o de muestra de biopsia de tejidos. Según el tipo de cáncer, el experto en la materia conocerá cómo obtener una muestra que comprende células que se sospecha que son células de cáncer.

40 Se da a conocer además un producto programa informático que al ser ejecutado en un ordenador o dispositivo de procesamiento de datos es capaz de llevar a cabo las etapas siguientes:

- a) asignar secuencias diana seleccionadas y aisladas que se han amplificado, que presentan regiones de unión a nucleosoma de consenso de una muestra biológica de un individuo, a cromosomas del genoma, e identificar las secuencias diana asignadas únicas,
- 45 b) determinar una primera cantidad de cada uno de uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas, y
- c) determinar una segunda cantidad de cada uno de uno o más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de uno o más segundos cromosomas, y d) determinar basándose en dicha primera y segunda cantidades de una aneuploidía cromosómica fetal de uno o más de entre dichos primeros cromosomas o una aberración cromosómica.

50 Una realización preferente de dicho producto programa informático, de dicha determinación en la etapa d), comprende además:

- i) determinar un parámetro a partir de dicha primera cantidad respecto a dicha segunda cantidad,
- 55 ii) comparar el parámetro con un valor de control de corte correspondiente y, basándose en la comparación, determinar si existe o no una diferencia que permita la predicción de una aberración cromosómica en uno o más de dichos primeros cromosomas o determinar si existe o no una diferencia que permite la predicción de una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.

60 Se da a conocer además un dispositivo para llevar a cabo un método tal como se da a conocer en la presente memoria, que comprende:

- a) una unidad de análisis que comprende una subunidad capaz de seleccionar y aislar a partir de dicha muestra biológica de un individuo, una o más secuencias diana de moléculas de ADN contenidas en la muestra biológica, en la que dichas secuencias diana comprenden secuencias de ADN que presentan regiones de unión a nucleosoma de

65

consenso, una subunidad para la amplificación de dichas moléculas de ADN diana y una subunidad para secuenciar secuencias diana amplificadas, y

5 b) una unidad de evaluación que comprende una subunidad receptora para los datos de secuencias y un ordenador o subunidad de procesamiento de datos que ejecuta el producto programa informático dado a conocer en la presente memoria.

10 El término "dispositivo" tal como se utiliza en la presente memoria se refiere a un sistema que comprende por lo menos la unidad de análisis anteriormente indicada y la unidad de evaluación ligada operativamente con ella. El cómo asociar las unidades del dispositivo de una manera operativa dependerá del tipo unidades incluidas en el dispositivo. Por ejemplo, en el caso de que se utilicen unidades para el análisis automático de una muestra, los datos obtenidos por dicha unidad de análisis de operación automática pueden ser procesados por el producto programa informático de la unidad de evaluación con el fin de obtener los resultados diagnósticos deseados. Preferentemente, las unidades están comprendidas en un único dispositivo en dicho caso. Sin embargo, el dispositivo puede comprender también unidades físicamente separadas que se encuentran conectadas mediante cables, inalámbricamente o a través de Internet.

FIGURAS

20 La figura 1 muestra un diagrama de flujo que delinea el principio de flujo de trabajo de un método para llevar a cabo el diagnóstico prenatal para la determinación de una aneuploidía cromosómica fetal en una muestra biológica obtenida de una mujer gestante según la presente invención.

25 La figura 2 (A) muestra un gráfico de la representación en porcentaje de lecturas de localización única sin ningún desapareamiento con el cromosoma 13 en muestras seleccionadas de ADN libres de células, según una realización de la presente invención, (B) muestra un gráfico de la representación en porcentaje de lecturas de localización única sin ningún desapareamiento con el cromosoma 21 en muestras seleccionadas de ADN libres de células, según una realización de la presente invención.

30 La figura 3 (A) muestra un gráfico de la representación en porcentaje de lecturas de localización única sin ningún desapareamiento con regiones cebo en el cromosoma 13 (150 pb cadena arriba-150 pb cadena abajo) sin ningún desapareamiento en muestras seleccionadas de ADN libre de células según una realización de la presente invención, (B) muestra un gráfico de la representación en porcentaje de lecturas de localización única en regiones cebo en el cromosoma 21 (150 pb cadena arriba-150 pb cadena abajo) sin ningún desapareamiento en muestras seleccionadas de ADN libres de células según una realización de la presente invención.

35 La figura 4 (A) muestra un gráfico de la representación en porcentaje de lecturas de secuencia de localización única en SIT (500 pb cadena arriba-1.500 pb cadena abajo) en el cromosoma 13 sin ningún desapareamiento en muestras seleccionadas de ADN libres de células según una realización de la presente invención, (B) muestra un gráfico de la representación en porcentaje de lecturas de secuencia de localización única en SIT (500 pb cadena arriba-1.500 pb cadena abajo) en el cromosoma 21 sin ningún desapareamiento en muestras seleccionadas de ADN libres de células según una realización de la presente invención.

45 La figura 5 muestra un gráfico de la representación en porcentaje de lecturas de secuencia en el cromosoma 13 en muestras seleccionadas de ADN libres de células según resultados anteriores utilizando la estrategia de secuenciación "shotgun" sin procedimiento de enriquecimiento.

50 Con el fin de entender más completamente la invención descrita en la presente memoria, se proporciona el ejemplo siguiente. Se proporciona meramente con fines ilustrativos y no debe interpretarse como limitativo de la presente invención en modo alguno.

Se entiende además que la presente invención comprenderá además variaciones de las realizaciones dadas a conocer expresamente en la medida que serían contempladas por el experto ordinario en la materia. La invención se define a partir del alcance de las reivindicaciones.

55 Ejemplo 1: realización del diagnóstico prenatal para la detección de trastornos cromosómicos fetales mediante tecnología de enriquecimiento de la diana

60 En primer lugar, se extraen hasta 15 ml de sangre periférica de una mujer gestante y se recogieron en tubos que contenían EDTA. Se obtiene plasma libre de células mediante centrifugación de la muestra de sangre. Se extrajo el ADN del plasma libre de células a partir del plasma mediante la utilización del minikit QIAamp DSP DNA Blood Mini Kit (Qiagen) o el microkit QIAamp DNA (Qiagen).

65 Tras la extracción del ADN se llevó a cabo una etapa de enriquecimiento de la diana. Más concretamente, una o más secuencias de ADN específicas que comprendían regiones de unión a nucleosoma de consenso en torno a los sitios de inicio de transcripción de los genes codificantes de proteína del cromosoma o cromosomas de interés se enriquecieron selectivamente mediante la técnica de selección de híbridos en solución. Para dicho enriquecimiento

específico, puede utilizarse el sistema de enriquecimiento de dianas SureSelect (Agilent Technologies) siguiendo el manual del usuario. Para iniciar la preparación de muestra, se utilizan hasta 10 ng de ADN del plasma libre de células para la producción de bibliotecas preparadas específica de la plataforma de secuenciación utilizada posteriormente, tal como el instrumento de secuenciación de Illumina. La posterior preparación de la biblioteca se llevó a cabo siguiendo el protocolo del fabricante correspondiente con la modificación de que no se realizó fragmentación mediante nebulización o sonicación de la muestra de ADN del plasma libre de células.

En paralelo con la producción de biblioteca, se creó un kit SureSelect específico que contenía una mezcla de oligonucleótidos de ARN SureSelect diseñados en la herramienta web de diseño de Agilent. La Tabla 1 muestra un ejemplo de un diseño personalizado de kit adecuado para la detección de aneuploidías cromosómicas y enfermedades genéticas ligadas al sexo causadas por una mutación en el cromosoma X. En detalle, para cada cromosoma de interés incluyendo el cromosoma o cromosomas de interés, se generan las sondas de hibridación para las regiones en torno a los sitios de inicio de transcripción (aproximadamente 1,5 kb) de todos los genes codificantes de proteína conocidos en el cromosoma respectivo, resultando en una cantidad de ADN enriquecido de aproximadamente 4 Mb.

Tabla 1:

Ejemplo de un diseño personalizado de kit para la detección de aneuploidías cromosómicas y enfermedades genéticas ligadas al X			
Cromosoma	Genes codificantes de proteína (correspondientes al número de sondas de hibridación) (Ensembl versión 55 - julio de 2009; http://www.ensembl.org/index.html)	Longitud aproximada en kb enriquecida por cada 1,5 kb en torno al sitio de inicio de transcripción de genes codificantes de proteína	Síndrome o trastorno cromosómico
13	359	520	Síndrome de Patau (trisomía 13)
16	1038	1250	Trisomía 16
18	315	470	Síndrome de Edwards (trisomía 18)
21	265	390	Síndrome de Down (trisomía 21)
X	883	1250	Síndrome de Turner, síndrome de la triple X, enfermedades genéticas ligadas al X
Y	86	120	Síndrome XYY, enfermedades genéticas ligadas al sexo
		4 Mb	

Para llevar a cabo la captura del ADN, se incubaba la biblioteca seleccionada según tamaño, con los oligonucleótidos de ARN SureSelect diseñados y los híbridos de ARN-ADN generados de esta manera se incuban con perlas magnéticas marcadas con estreptavidina para permitir la captura de los híbridos de ARN-ADN mediante la unión de los mismos a las perlas. Tras recolectar las perlas cargadas mediante su atracción hacia un imán, las perlas se lavan y los oligonucleótidos de ARN se digieren, de manera que sólo se recolecta el ADN enriquecido de interés restante. Tras la amplificación final del ADN y la estimación de la calidad de los productos de PCR utilizando, por ejemplo el instrumento Agilent 2100 Bioanalyzer (Agilent Technologies), el grupo enriquecido de secuencias de ADN diana seguidamente se somete a secuenciación mediante secuenciación en paralelo masiva utilizando, por ejemplo, la plataforma 454 (Roche) (Margulies M. et al., Nature 437:376-380, 2005), el analizador genómico de Illumina o el sistema SOLiD (Applied Biosystems), que permite la secuenciación de muchas moléculas de ácidos nucleicos aislados a partir de una muestra de ADN de plasma humano en paralelo.

El procedimiento bioinformático posterior seguidamente se utiliza para localizar cada una de dichas secuencias de ADN en el genoma humano. Más concretamente, se recogen las lecturas cortas del instrumento de secuenciación y se alinean con el genoma de referencia humano (hg18, NCBI revisión 36 (GenBank números de acceso: NC_000001 a NC_000024) utilizando varias herramientas bioinformáticas, tales como ELAND (por sus siglas en inglés, alineación eficiente a gran escala de bases de nucleótidos). Con el fin de garantizar una calidad elevada de los resultados, resulta preferente que sólo se consideren aquellas lecturas para el análisis posterior que se encuentran localizadas en regiones genómicas preseleccionadas que comprenden regiones de unión a nucleosoma de consenso del cromosoma o cromosomas de interés y que se localizan de manera única en el genoma humano con sólo uno o dos desapareamientos respecto al genoma humano.

La lectura digital resultante de las moléculas de ácidos nucleicos se utiliza a continuación para la detección de aneuploidías cromosómicas fetales, por ejemplo la trisomía 13, 18 o 21, y de manera similar puede utilizarse para la determinación del sexo del feto.

Un desequilibrio, tal como una aneuploidía cromosómica, en una muestra experimental dada queda revelada por diferencias en el número o porcentaje de secuencias alineadas con una región cromosómica de interés en comparación con el número o porcentaje correspondiente de dichas secuencias esperado o predeterminado para una muestra de genoma humano euploide.

5 El presente método resulta adecuado para la detección de una o más aneuploidías cromosómicas en una operación, en la que el cromosoma afectado típicamente se selecciona de entre el grupo que consiste de cromosoma 21 (trisomía 21), cromosoma 18 (trisomía 18), cromosoma 13 (trisomía 13) y cromosoma X (síndrome de Turner).

10 El método de secuenciación mediante enriquecimiento selectivo de la diana según la presente invención también puede resultar aplicable a otras aplicaciones diagnósticas que implican la evaluación cualitativa y/o cuantitativa del contenido de ácidos nucleicos del suero o del plasma, por ejemplo en oncología y medicina de los trasplantes. Por ejemplo, la técnica de secuenciación mediante enriquecimiento selectivo de la diana indicado anteriormente aplicado sobre ADN libre de células también puede utilizarse para detectar alteraciones cromosómicas específicas de tumores asociadas a un cáncer específico.

15 El principio de la invención se describe adicionalmente en la reivindicación independiente posteriormente en la presente memoria, siendo las diversas realizaciones de la invención la materia objeto de las reivindicaciones dependientes.

20 Ejemplo 2: realización del diagnóstico prenatal para la detección de la trisomía 21 y la trisomía 13 mediante tecnología de enriquecimiento de la diana y secuenciación "barcode" (de código de barras) multiplexada

25 Para el estudio, se seleccionaron muestras de sangre materna procedentes de 4 embarazos simples. Una de las mujeres gestantes portaba un feto masculino euploide y los otros tres portaban un feto con trisomía 21, trisomía 13 y trisomía 18, respectivamente (Tabla 2, posteriormente). Se extrajeron hasta 15 ml de sangre venosa periférica de dichas mujeres gestantes y se recolectaron en tubos con EDTA. Se obtuvo el plasma de las muestras de sangre mediante centrifugación a 1.600g a 4°C durante 10 minutos. Para separar las células residuales, el plasma se centrifugó adicionalmente a 16.000 g a 4°C durante 10 minutos. De las muestras de plasma se extrajo el ADN libre de células a partir de 0,8 a 1 ml de plasma mediante la utilización del kit de ácidos nucleicos circulantes QIAamp (Qiagen) siguiendo el protocolo del fabricante.

30 A continuación, se utilizaron hasta 10 ng de ADN del plasma libre de células para construir bibliotecas de secuenciación de Illumina. La preparación de la biblioteca de ADN siguió el protocolo de preparación estándar de muestras de Illumina para la secuenciación de extremos apareados con unas cuantas modificaciones. Brevemente, no se llevó a cabo ninguna fragmentación mediante nebulización o sonicación de las muestras de ADN del plasma libre de células. La preparación de biblioteca se llevó a cabo siguiendo el protocolo de preparación de muestras beta Chromatin Immunoprecipitation Sequencing (ChIP-Seq) (Illumina, componente nº 11257047 Rev. A) utilizando enzimas de Fermentas (ADN polimerasa de T4, ADN polimerasa Klenow, polinucleótido quinasa de T4, ADN ligasa), así como de Finnzymes Oy (polimerasa Phusion*). Se repararon los extremos de los productos y se añadieron A en el extremo 3' no del molde. Para que se encontrase disponible la secuenciación "barcode" (de código de barras) multiplexada, se etiquetaron bibliotecas de ADN con diferentes identificadores (códigos de barra) durante la ligación de adaptadores de extremos apareados. El primer adaptador contenía los sitios de cebador de secuenciación para la aplicación lectura 1 y un identificador de 4 pb, así como una 'T' que resulta necesaria para la ligación de adaptadores. El segundo adaptador contenía los sitios de cebador de secuenciación para la aplicación lectura 2 y también una 'T'. Las bibliotecas de ADN con muestras con código de barras seguidamente se amplificaron adicionalmente utilizando una PCR de 12 ciclos y cebadores que contenían los sitios de unión para las celdas de flujo. Los fragmentos de ADN ligados a adaptador se seleccionaron según tamaño en el intervalo de 150 a 300 pb utilizando una electroforesis en gel de agarosa al 2%. Se llevó a cabo un control de calidad y la cuantificación de las bibliotecas utilizando un kit de ADN de alta sensibilidad en el instrumento Agilent 2100 Bioanalyzer siguiendo las instrucciones del fabricante.

35 Para que la diana fuesen todos los exones humanos y sus regiones genómicas humanas asociadas correspondientes a los sitios de inicio de transcripción (SIT), se incubaron 500 ng de bibliotecas con el kit SureSelect Human All Exon (Agilent Technologies) y se enriquecieron siguiendo el protocolo del fabricante. Tras la elución de los fragmentos de ADN capturados, las bibliotecas se amplificaron nuevamente durante 12 a 14 ciclos de PCR con cebadores específicos de Sure Select Illumina. La amplificación permite la cuantificación exacta utilizando el chip de alta sensibilidad del instrumento Bioanalyzer antes de la secuenciación.

40 A continuación, los cuatro muestras diferentes con código de barras seguidamente se agruparon en un único tubo y se generaron agrupaciones clonales utilizando el sistema de amplificación clonal cBOT con el kit de generación de agrupaciones con extremos apareados cBOT. Siguiendo el flujo de trabajo de secuenciación de Illumina, los moldes de ADN de molécula única amplificados se secuenciaron utilizando la síntesis en paralelo masiva en el analizador genómico Ilx de Illumina.

65

El procedimiento bioinformático posterior incluía análisis de imágenes, identificación de bases y alineación mediante la utilización del software "en pipeline" de Illumina. Para el análisis individual posterior, una estrategia semiautomática de clasificación de etiquetas identifica cada muestra que presenta un código de barras único. Las primeras 32 pb de cada lectura de cada muestra se alinean con la secuencia de referencia genómica humana con repeticiones enmascaradas de NCBI revisión 36 (también conocido como hg18; GenBank números de acceso: NC_000001 to NC_000024) descargado del explorador UCSC Genome Browser utilizando el software de alineación ELAND (software GAPipeline-1.4.0) proporcionados por Illumina.

Las lecturas digitales resultantes de las moléculas de ácidos nucleicos se utiliza a continuación para la detección de aneuploidías cromosómicas fetales, por ejemplo la trisomía 13 o 21, y de manera similar puede utilizarse para la determinación del sexo del feto.

Inicialmente se realizó un recuento del número de total de lecturas secuenciadas de cada muestra. A continuación, sólo se utilizaron para el análisis posterior las lecturas separadas que habían sido localizadas en un único sitio en la secuencia de referencia genómica humana con enmascarado de repeticiones y sin ningún desapareamiento de nucleótidos (ver la Tabla 2, posteriormente).

En primer lugar, se identificó un desequilibrio, tal como la trisomía 21 o la trisomía 13, en las muestras experimentales dadas, a partir de las diferencias en el número o el porcentaje de lecturas de localización única con enmascarado de las repeticiones sin ningún desapareamiento de interés (originadas en los cromosomas 13 y 21, respectivamente) en comparación con el número o porcentaje correspondiente de dichas secuencias determinado para la muestra del genoma humano euploide. El porcentaje esperado de representación de cada cromosoma se obtuvo dividiendo el número de lecturas de localización única con enmascarado de repeticiones sin ningún desapareamiento por cromosoma, por el número total de lecturas de localización única con enmascarado de repeticiones sin ningún desapareamiento de todos los cromosomas. Tal como se muestra en la figura 2A, el porcentaje de lecturas de localización única en el cromosoma 13 procedentes de la muestra S_T13 era más elevado que el de la muestra S_euploide con un feto euploide, así como de la muestra S_T21 que portaba un feto con trisomía 21. El porcentaje de lecturas de localización única en el cromosoma 21 procedentes de la muestra S_T21 también era más elevado que el de la muestra S_euploide de un feto euploide, así como de la muestra S_T13 que portaba un feto con trisomía 13 (figura 2B).

En segundo lugar, se identificó un desequilibrio, tal como la trisomía 21 o la trisomía 13, en las muestras experimentales dadas, a partir de las diferencias en el número o el porcentaje de lecturas de localización única con enmascarado de las repeticiones sin ningún desapareamiento alineadas con una región cromosómica dada de interés en comparación con el número o porcentaje correspondiente de dichas secuencias determinado para la muestra del genoma humano euploide. La región cromosómica de interés se caracterizó por las regiones cebo de 120 pb predeterminadas (disponibles de la plataforma eArray de Agilent Technologies) del kit SureSelect Human All Exon Kit (Agilent Technologies) más regiones de 150 pb flanqueantes cadena arriba y cadena abajo de las regiones de cebo. A continuación, se obtuvo el porcentaje esperado de representación de cada cromosoma tal como se ha indicado anteriormente. Se observó una sobrerrepresentación de lecturas de localización única para los cromosomas 13 y 21 en los casos T13 y T21, respectivamente (figuras 3A y 3B).

En tercer lugar, se identificó un desequilibrio, tal como la trisomía 21 o la trisomía 13, en las muestras experimentales dadas, a partir de las diferencias en el número o el porcentaje de lecturas de localización única con enmascarado de las repeticiones sin ningún desapareamiento alineadas con una región de unión a nucleosoma de consenso dada en comparación con el número o porcentaje correspondiente de dichas secuencias esperado o predeterminado para la muestra del genoma humano euploide. La región de unión a nucleosoma de consenso utilizada en la presente memoria incluía regiones de secuencia entre 500 pb cadena arriba del SIT y 1.500 pb cadena abajo del SIT. El porcentaje de lecturas de localización única en el cromosoma 13 de la muestra S_T13 era más elevado que el de la muestra S_euploide con un feto euploide, así como el de la muestra S_T13 (figura 4A). El porcentaje de lecturas de localización única en el cromosoma 21 de la muestra S_T21 sólo era más elevado que el de la muestra S_T13 y no era más elevado que el de la muestra S_euploide con un feto euploide (figura 4B).

El presente método resulta adecuado para la detección de una o más aneuploidías cromosómicas en una operación, en la que el cromosoma afectado típicamente se selecciona de entre el grupo que consiste de cromosoma 21 (trisomía 21), cromosoma 18 (trisomía 18), cromosoma 13 (trisomía 13) y cromosoma X (síndrome de Turner).

En comparación con experimentos anteriores (figura 5 y Tabla 3, posteriormente) en los que se utilizaba únicamente la secuenciación "shotgun", el presente método resultó apropiado incluso para detectar una trisomía 13.

Además, los métodos indicados tienden a resultar en la reducción de las capacidades de almacenamiento necesarias para los datos en bruto, así como para el mapeado y la alineación de las lecturas de secuencia generadas.

Tabla 2:

(i) Resumen de datos clínicos y número de lecturas de secuencia del Ejemplo 2 "Realización del diagnóstico prenatal para la detección de la trisomía 21 y la trisomía 13 mediante tecnología de enriquecimiento de la diana y secuenciación "barcode" (de código de barras) multiplexada

Muestra	Cariotipo	Gestación (semanas+días)	Edad total nº de lecturas de secuencia lectura1	Nº total de lecturas de secuencia lectura2
S_T13	47XY+13	13+5	5005094	5005094
S_T21	47XY+21	13+0	6739231	6739231
S_euploide	46XY	16+0	3415786	3415786

(ii) Resumen de número de lecturas de secuencia del Ejemplo 2

Muestra	Nº total de lecturas de localización única sin ningún desapareamiento_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento_lectura2	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura2
S_T13	2856635	2745340	8643	83057	32290	31010
S_T21	3913173	3761195	109842	105110	45647	44115
S_euploide	1959738	1895396	56616	54658	2191	21186

(iii) Resumen del número de lecturas de secuencia de una región cromosómica de interés (caracterizada por las regiones cebo de 120 pb predeterminadas del kit SureSelect Human All Exon Kit más regiones de 150 pb flanqueantes cadena arriba y cadena abajo de las regiones cebo) del Ejemplo 2

Muestra	Nº total de lecturas de localización única sin ningún desapareamiento_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento_lectura2	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura2
S_T13	1599317	1530800	42125	40162	17060	16256
S_T21	2215860	2120634	54244	51598	24433	23345
S_euploide	1091265	1051800	27601	26577	11679	11303

(iv) Resumen del número de lecturas de secuencia alineadas con una región de unión a nucleosoma de consenso dada que incluye regiones de secuencia entre las 500 pb cadena arriba del SIT y 1.500 pb cadena abajo del SIT (Ejemplo 2)

Muestra	Nº total de lecturas de localización única sin ningún desapareamiento_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento_lectura2	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 13_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura1	Nº total de lecturas de localización única sin ningún desapareamiento con el cromosoma 21_lectura2
S_T13	202723	196213	4254	4087	2755	2652
S_T21	282640	273834	5520	5408	3951	3814
S_euploide	135892	131693	2665	2567	1880	1857

Resumen del número de lecturas de secuencia de experimentos anteriores utilizando el método de secuenciación "shotgun"

Muestra	Nº total de lecturas de secuencia	Nº total de lecturas de localización única sin ningún	Nº total de lecturas de localización única sin ningún	Nº total de lecturas de localización única sin ningún

ES 2 564 656 T3

		desapareamiento	desapareamiento con el cromosoma 13	desapareamiento con el cromosoma 21
S_T13	16611762	5992776	215666	73885
S_T21	26137898	10752119	402319	141130
S_euploide	20289419	7928946	285433	99306

REIVINDICACIONES

- 5 1. Método para la determinación de una aneuploidía cromosómica fetal en una muestra biológica obtenida de un individuo femenino gestante, en el que la muestra biológica incluye moléculas de ácidos nucleicos, comprendiendo el método:
- 10 a) seleccionar y aislar a partir de una muestra biológica de un individuo hembra gestante una o más secuencias diana de moléculas de ADN contenidas en la muestra biológica, en la que dichas secuencias diana comprenden secuencias de ADN que presentan regiones de unión a nucleosoma de consenso y en la que dichas moléculas de ADN que comprenden las secuencias diana se aíslan y se enriquecen utilizando cromatografía de afinidad o enriquecimiento por lotes utilizando cebos de ácidos nucleicos marcados, los cuales permiten el enriquecimiento y el aislamiento de las moléculas de ADN hibridadas mediante purificación por afinidad,
- 15 b) amplificar dichas secuencias diana seleccionadas,
- c) secuenciar dichas secuencias diana seleccionadas y amplificadas y asignarlas a los cromosomas del genoma e identificar las secuencias diana asignadas únicas,
- 20 d) determinar una primera cantidad de cada uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas,
- e) determinar una segunda cantidad de cada uno o más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho segundo o segundos cromosomas, y
- f) determinar basándose en dicha primera y segunda cantidades una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.
- 25 2. Método según la reivindicación 1, en el que la etapa f) comprende además:
- i) determinar un parámetro a partir de dicha primera cantidad respecto a dicha segunda cantidad,
- ii) comparar el parámetro con un valor de control de corte correspondiente y, basándose en la comparación, determinar si existe o no una diferencia que permite la predicción de una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.
- 30 3. Método según la reivindicación 1 o 2, en el que dichos ácidos nucleicos cebo son moléculas de ARN cebo biotiniladas.
- 35 4. Método según la reivindicación 3, en el que dichos ácidos nucleicos cebo están diseñados específicamente para presentar una secuencia, que es capaz de hibridarse con las secuencias de consenso para dichas regiones de unión a nucleosoma.
5. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha muestra biológica es una muestra de sangre materna.
- 40 6. Método según la reivindicación 6, en el que dicha muestra de sangre es una muestra de plasma sanguíneo materno o suero sanguíneo materno.
- 45 7. Método según las reivindicaciones 1 a 4, en el que dicha muestra biológica es una muestra de orina o una muestra de saliva.
8. Método según cualquiera de las reivindicaciones 1 a 7, en el que dicho primer o primeros cromosomas se seleccionan de entre el grupo que consiste de cromosoma 21, cromosoma 18, cromosoma 13, cromosoma X y cromosoma Y.
- 50 9. Dispositivo para llevar a cabo el método según cualquiera de las reivindicaciones 1 a 8, que comprende:
- a) una unidad de análisis que comprende una subunidad capaz de seleccionar y aislar a partir de dicha muestra biológica una o más secuencias diana individuales de moléculas de ADN contenidas en la muestra biológica, en la que dichas secuencias diana comprenden secuencias de ADN que presentan regiones de unión a nucleosoma de consenso y en la que dichas moléculas de ADN que comprenden las secuencias diana se aíslan y se enriquecen mediante cromatografía de afinidad o enriquecimiento por lotes utilizando ácidos nucleicos cebo marcados, los cuales permiten el enriquecimiento y el aislamiento de las moléculas de ADN hibridadas mediante purificación por afinidad, una subunidad para la amplificación de dichas moléculas de ADN diana y una subunidad para la secuenciación de las secuencias diana amplificadas, y
- 55 b) una unidad de evaluación que comprende una subunidad receptora para los datos de secuencias y un ordenador o subunidad de procesamiento de los datos que ejecuta un producto programa informático que, al ejecutarlo en un ordenador o dispositivo de procesamiento de datos, es capaz de llevar a cabo las etapas siguientes:
- a) asignar secuencias diana seleccionadas y aisladas que se han amplificado, que presentan regiones de unión a nucleosoma de consenso de una muestra biológica de un individuo, a cromosomas del genoma, e identificar las
- 60 secuencias diana asignadas únicas,
- 65

b) determinar una primera cantidad de cada uno de uno o más primeros cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de dicho primer o primeros cromosomas, y
c) determinar una segunda cantidad de cada uno de uno o más segundos cromosomas identificados basándose en dichas secuencias diana asignadas únicas que se originan a partir de uno o más segundos cromosomas, y d)
5 determinar basándose en dicha primera y segunda cantidades de una aneuploidía cromosómica fetal de uno o más de entre dichos primeros cromosomas.

10. Dispositivo según la reivindicación 9, en el dicha determinación en la etapa d) comprende además:

10 i) determinar un parámetro a partir de dicha primera cantidad respecto a dicha segunda cantidad,
ii) comparar el parámetro con un valor de control de corte correspondiente y, basándose en la comparación, determinar si existe o no una diferencia que permita la predicción de una aberración cromosómica en uno o más de dichos primeros cromosomas o determinar si existe o no una diferencia que permite la predicción de una aneuploidía cromosómica fetal de uno o más de dichos primeros cromosomas.

15

Fig. 1

Recepción de muestra

Muestra materna

*Separación en componentes
y aislamiento de ADN libre de células* ↓

ADN libre de células extraído

Enriquecimiento de la diana ↓

Pool enriquecido de ADN seleccionado

Secuenciación ↓

Datos de ADN secuenciado

Análisis de los datos ↓

Secuencias cromosómicas alineadas y localizadas

Cuantificación del ADN ↓

Cantidad determinada de un primer ADN de un primer cromosoma, cantidad determinada de un segundo, tercer, etc. ADN de cromosomas adicionales

Comparación de las cantidades de ADN ↓

Detección de desequilibrio cromosómico

Fig. 2

A

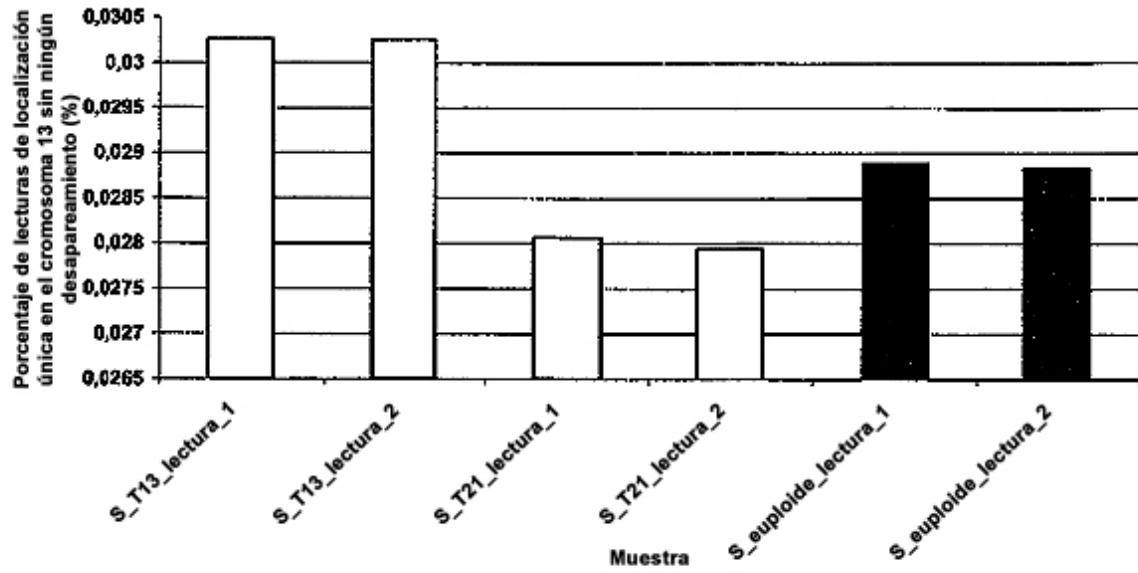


Fig. 2 cont.

B

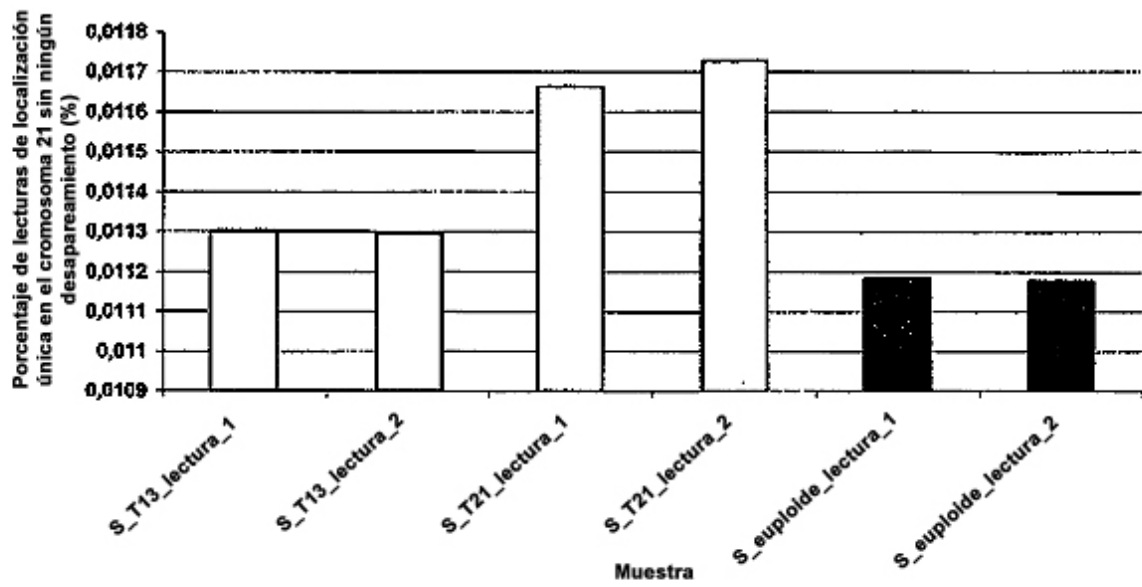


Fig. 3

A

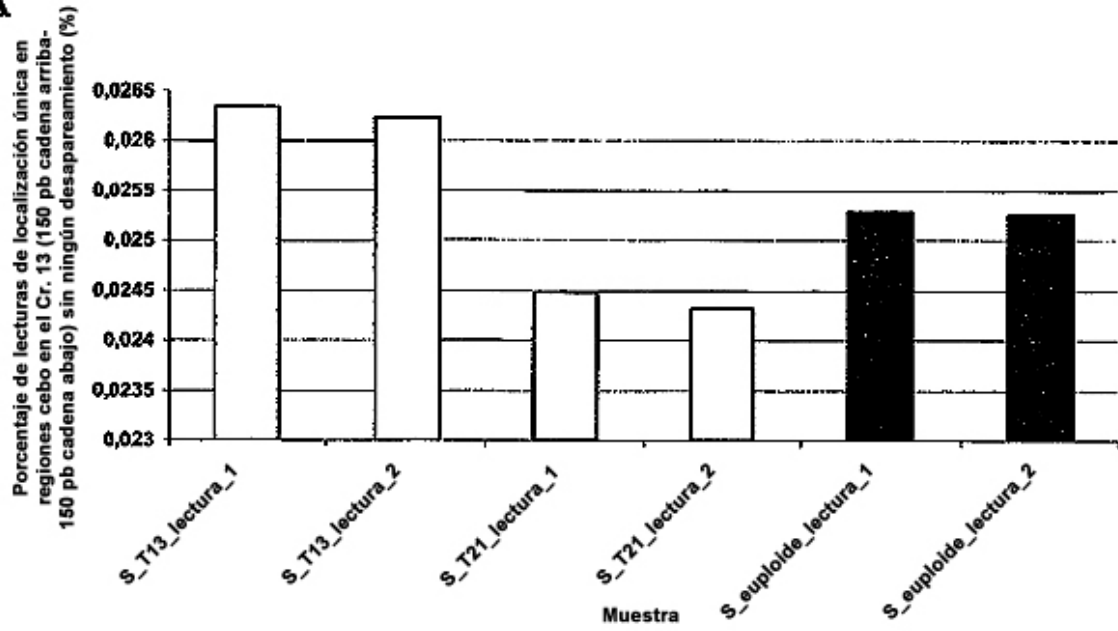


Fig. 3 cont.

B

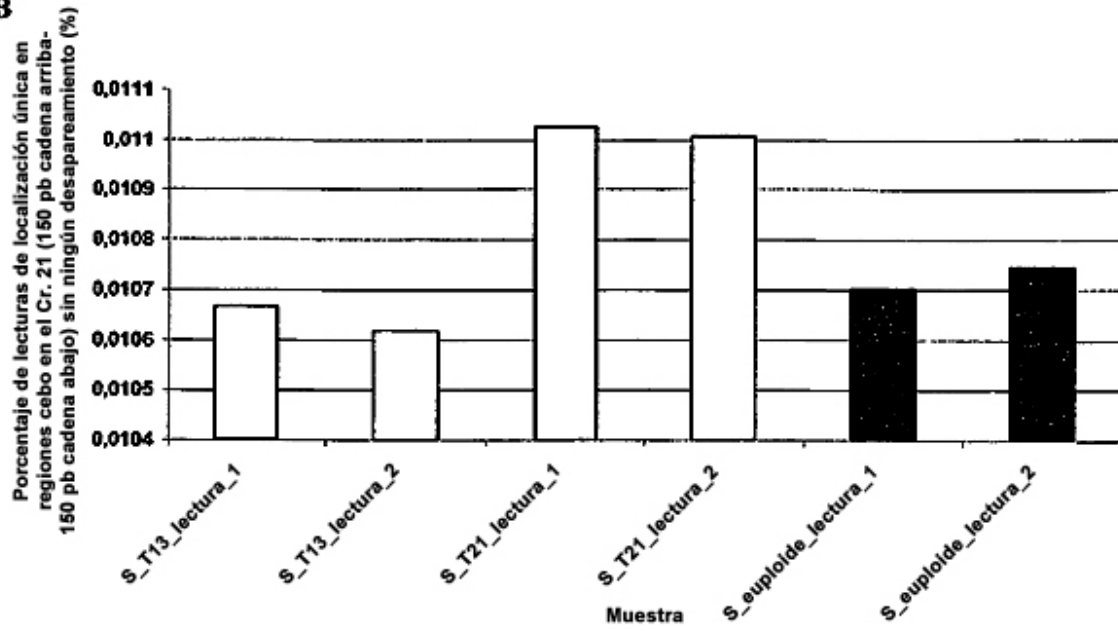


Fig. 4

A

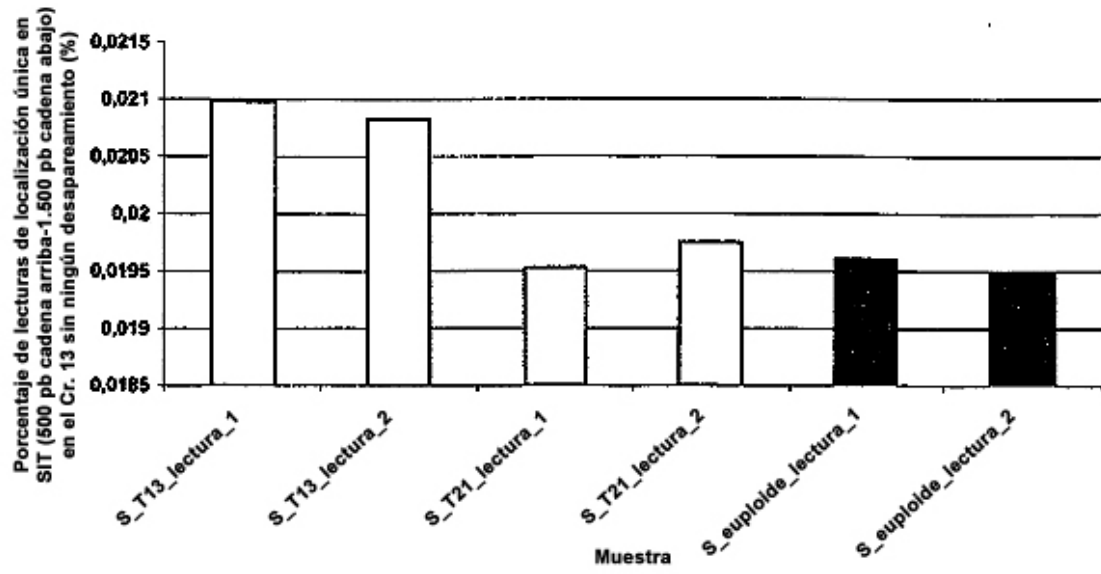


Fig. 4 cont.

B

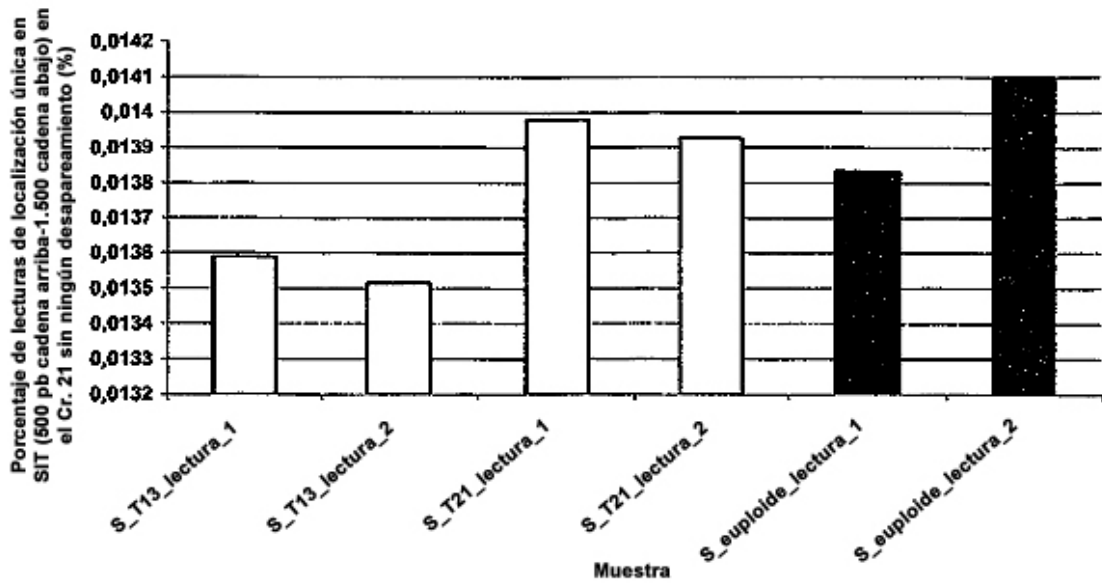


Fig. 5

