

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 745 556**

51 Int. Cl.:

C12Q 1/6811 (2008.01)

C12Q 1/6827 (2008.01)

C12Q 1/6883 (2008.01)

C12Q 1/6816 (2008.01)

G16H 50/20 (2008.01)

G16B 30/00 (2009.01)

G16B 25/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.07.2016 PCT/US2016/044914**

87 Fecha y número de publicación internacional: **02.02.2017 WO17020023**

96 Fecha de presentación y número de la solicitud europea: **29.07.2016 E 16748036 (7)**

97 Fecha y número de publicación de la concesión europea: **10.07.2019 EP 3329010**

54 Título: **Ácidos nucleicos y métodos para detectar anomalías cromosómicas**

30 Prioridad:

29.07.2015 US 201562198654 P

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

02.03.2020

73 Titular/es:

**PROGENITY, INC. (100.0%)
4330 La Jolla Village Drive, Suite 200
San Diego, CA 92122, US**

72 Inventor/es:

**MANN, TOBIAS;
WANG, HENG;
KIM, JUNG H. y
SEKEDAT, MATTHEW**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 745 556 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

Ácidos nucleicos y métodos para detectar anomalías cromosómicas

Campo de la invención

La presente invención se refiere a los sistemas y a los métodos para determinar, *inter alia*, aneuploidía y anomalías cromosómicas en un sujeto que lo necesita.

Antecedentes de la invención

Las principales anomalías cromosómicas se detectan en casi 1 de 140 nacimientos vivos y en una fracción mucho más alta de fetos que no llegan a término o nacen muertos. Hsu (1998) Prenatal diagnosis of chromosomal abnormalities through amniocentesis. In: Milunsky A, editor. Genetic Disorders and the Fetus. 4° ed. Baltimore: The Johns Hopkins University Press. 179-180; Staebler et al. (2005) ¿Debe la determinación del cariotipo ser sistemática para todas las malformaciones detectadas por ecografía obstétrica? Prenat Diagn 25: 567-573. La aneuploidía más común es la trisomía 21 (síndrome de Down), que se produce, normalmente, en 1 de 730 nacimientos. Hsu; Staebler et al. Si bien menos común que la trisomía 21, la trisomía 18 (síndrome de Edwards) y la trisomía 13 (síndrome de Patau) se producen en 1 de 5.500 y en 1 de 17.200 nacimientos vivos, respectivamente. Hsu. Una gran variedad de defectos congénitos, deficiencias en el crecimiento y discapacidades intelectuales se encuentran en niños con aneuploidía cromosómica, y estos presentan desafíos para toda la vida para las familias y para las sociedades. Jones (2006) Smith's recognizable patterns of human malformation. Philadelphia: Elsevier Saunders. Hay una variedad de pruebas prenatales que pueden indicar un aumento en el riesgo de aneuploidía fetal, que incluyen pruebas de diagnóstico invasivas como por ej., la amniocentesis o la muestra de vellosidades coriónicas, las cuales son el estándar de oro actual, pero están vinculadas con un riesgo nada despreciable de pérdida del feto. American College of Obstetricians and Gynecologists (2007) ACOG Practice Bulletin No. 88, Diciembre 2007. Invasive prenatal testing for aneuploidy. Obstet Gynecol 110: 1459-1467. Por lo tanto se han buscado pruebas más confiables, no invasivas para la aneuploidía fetal. Las más prometedoras de estas se basan en la detección del ADN fetal en el plasma materno. Se ha demostrado que la secuenciación masiva paralela de bibliotecas que se generan en el plasma materno pueden detectar, de manera confiable, las anomalías del cromosoma 21. Chiu et al., (2008) Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci USA 105:20458-20463; Fan et al., (2008) Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A 105: 16266-16271.

Los métodos actuales para la detección sistemática de aneuploidía, como por ej., la secuenciación por bombardeo con microproyectiles, requieren mucho tiempo o son muy costosos, o requieren un análisis bioinformático extenso.

Por lo tanto, existe la necesidad de desarrollar pruebas rentables y eficientes que tengan altas sensibilidades y especificidades. Los métodos para la detección de aneuploidía en un feto a partir de la sangre materna son conocidos en la técnica anterior, se incluyen en los documentos WO 2007/147076 y WO/2011/066476 A1. Dichos documentos, no obstante, no proporcionan ninguna solución que entre dentro de las características distinguidas de la invención que se reivindica, entre estas están la MIP (sonda de inversión molecular) que transporta dos etiquetas identificatorias moleculares únicas, la detección de amplicones MIP tiene lugar a través de la secuenciación, y el recuento de amplicones tiene lugar mediante el uso de etiquetas moleculares únicas. Existe, además, técnica anterior que se refiere a las variantes en el número de copias en general, o que están siendo detectadas para otro tejido, como por ej., en tumores (Wang et al., Nucleic Acid Research, vol. 33, no. 21, 1 Noviembre 2005, páginas e183/1-14; Wang et al., Cancer Genetics, vol. 205, no. 7, 11 junio 2012, páginas 341-355; Wang et al., Genome Biology, vol. 8, no. 11, 20, Noviembre 2007, página R246; Johnson et al., Breast Cancer Research and Treatment, vol. 133, no. 3, 4 Noviembre 2011, páginas 889-898). Cada uno de estos documentos, no obstante, describe una variante del método que se divulga en la Patente Estadounidense 2013/072390 y Wang et al., Nucleic Acid Research, vol. 33, no. 21, 1 Noviembre 2005, páginas e183/1-14, mediante la cual la MIP transporta solo una única etiqueta identificatoria, y mediante la cual una única etiqueta indica un sitio de unión cromosómico específico, mientras que en los métodos que se reivindican en la presente invención el mismo sitio de unión cromosómico está vinculado con una pluralidad de diferentes primeras y segundas etiquetas.

Sumario de la invención

La invención se presenta en las reivindicaciones adjuntas. Las realizaciones de la descripción que no están dentro del alcance de dichas reivindicaciones se proporcionan a los fines ilustrativos únicamente, y no forman parte de la presente invención. La invención comprende un método de detección sistemática para la aneuploidía en un feto, que comprende:

- a) obtener una muestra de ácido nucleico que se aísla a partir de una muestra de sangre materna;
- b) capturar una pluralidad de secuencias diana de interés en la muestra de ácido nucleico que se obtiene en la etapa a) mediante el uso de una o más poblaciones de sondas de inversión molecular (MIP) para producir una pluralidad de replicones, en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes: primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización; en donde los pares del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico

que, respectivamente, flanquean cada secuencia en la pluralidad de secuencias diana de interés; en donde la primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;

- 5 c) secuenciar una pluralidad de amplicones de MIP que se amplían a partir de los replicones que se obtienen en la etapa b);
- d) determinar el número de eventos de captura de cada una de una primera población de amplicones de la pluralidad de amplicones que se proporciona en la etapa c), basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la primera población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- 10 e) determinar el número de eventos de captura de cada una de una segunda población de amplicones de la pluralidad de amplicones que se proporciona en la etapa c), basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la segunda población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- 15 f) determinar, para cada secuencia diana de interés a partir de la cual se produjo la primera población de amplicones, una métrica de captura del sitio basado, al menos, en parte en el número de eventos de captura que se determinaron en la etapa d);
- g) identificar un primer subconjunto de la métrica de captura del sitio que se determinó en la etapa f) que satisface, al menos, un criterio;
- 20 h) determinar, para cada secuencia diana de interés a partir de la cual se produjo la segunda población de amplicones, una métrica de captura del sitio basado, al menos, en parte en el número de eventos de captura que se determinaron en la etapa e);
- i) identificar un segundo subconjunto de la métrica de captura del sitio que se determinó en la etapa h) que satisface, al menos, un criterio;
- 25 j) normalizar una primera medida que se determinó a partir del primer subconjunto de métrica de captura del sitio que se identificó en la etapa g) mediante una segunda medida que se determinó a partir del segundo subconjunto de la métrica de captura del sitio que se identificó en la etapa i) para obtener una relación de prueba;
- 30 k) comparar la relación de prueba con una pluralidad de relaciones de referencia que se computan basado en las muestras de ácido nucleico de referencia que se aislaron de los sujetos de referencia, que se sabe exhiben euploidía o aneuploidía, para determinar la presencia o la ausencia de una relación de prueba que indica la aneuploidía en el feto.

En un método preferido la muestra de ácido nucleico es ADN o ARN.

Por ejemplo, la muestra de ácido nucleico es ADN genómico.

La muestra de sangre puede ser una muestra de sangre entera, una muestra de plasma o una muestra de suero.

35 La extensión del primer brazo de polinucleótidos de focalización puede estar entre 14 y 30 pares base.

La extensión del segundo brazo de polinucleótidos de focalización puede estar entre 14 y 30 pares base.

Cada uno de los brazos de polinucleótidos de focalización puede tener una temperatura de fusión de entre 45 °C y 80 °C.

40 Cada uno de los brazos de polinucleótidos de focalización puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %. La extensión de la primera etiqueta molecular única puede estar entre 4 y 15 pares base.

La extensión de la segunda etiqueta molecular única puede estar entre 4 y 15 pares base.

Cada una de las etiquetas moleculares únicas puede tener una temperatura de fusión de entre 45 °C y 80 °C.

Cada una de las etiquetas moleculares únicas puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %

45 Preferiblemente, el ligador polinucleotídico no es sustancialmente complementario con alguna región genómica del sujeto. El ligador polinucleotídico puede tener una extensión de entre 14 y 30 pares base.

El ligador polinucleotídico puede tener una temperatura de fusión de entre 45 °C y 80 °C. El ligador polinucleotídico puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %. El ligador polinucleotídico puede comprender, al menos, un cebador de amplificación.

50 El ligador polinucleotídico puede comprender un cebador de amplificación hacia adelante y un cebador de amplificación inverso, en donde la secuencia del cebador de amplificación hacia adelante puede comprender la secuencia de nucleótidos de 5'-CTTCAGCT- TCCCGATTACGG-3' (SEC ID NO: 1) y en donde la secuencia del cebador de amplificación inverso comprende la secuencia de nucleótidos de 5'-GCACGATCCGACGGTAGTGT-3' (SEC ID NO: 2).

ES 2 745 556 T3

El ligador polinucleotídico puede comprender la secuencia de nucleótidos de 5'-CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT-3' (SEC ID NO: 3).

El primer brazo de polinucleótidos de focalización puede comprender la secuencia de nucleótidos de 5'-CACTGCACTCCAGCCT-GG-3' (SEC ID NO: 4).

5 El segundo brazo de polinucleótidos de focalización puede comprender la secuencia de nucleótidos de 5'-GAGGCTGAGGCAG-GAGAA - 3' (SEC ID NO: 5).

La MIP puede comprender la secuencia de nucleótidos de 5'-CACTGCACTCCA GCCTGG(N1-6)CTTCAGCTTC-CCGATTACGGGCACGATCCGACGGTAGTGT(N7-12) GAGGCTGAGGCAGGAGAA-3' (SEC ID NO: 6), en donde (N1-6) representa la primera etiqueta molecular única y (N7-12) representa la segunda etiqueta molecular única.

10 La MIP puede comprender la secuencia de nucleótidos de cualquiera de MIP 001-008 (SEC ID Nos.: 7-14).

La población de MIP puede tener una concentración de entre 10 fM y 100 nM.

Cada uno de los replicones de las MIP puede ser una molécula de ácido nucleico circular monocatenaria.

La métrica de captura del sitio es un índice de eficiencia de captura del sitio (SCE).

La métrica de captura del sitio es una medida de consistencia de la captura del sitio (SCE).

15 Cada uno de los replicones de las MIP que se proporcionó en la etapa b) se pueden producir mediante:

i) el primero y el segundo brazo de polinucleótidos de focalización, respectivamente, que hibridan a la primera y a la segunda región en la muestra de ácido nucleico, respectivamente, en donde la primera y la segunda región flanquean una secuencia diana de interés; y

20 ii) después de la hibridación, utilizan una mezcla de ligación/extensión para extender y ligar la región con huecos entre los dos brazos de polinucleótidos de focalización para formar moléculas de ácido nucleico circulares monocatenarias.

Cada uno de los replicones de las MIP puede ser una molécula de ácido nucleico circular monocatenaria.

25 La etapa de secuenciación de c) puede comprender un método de secuenciación de próxima generación, en donde el siguiente método de secuenciación de próxima generación puede comprender un método de secuenciación paralelo masivo, o un método de secuenciación masivo, paralelo, de corta lectura.

30 El método puede comprender, antes de la etapa de secuenciación de c), una reacción de PCR para amplificar los replicones de las MIP para secuenciación, en donde la reacción de PCR puede ser una reacción de PCR de indexado, en donde dicha reacción de PCR de indexado puede introducir en cada uno los amplicones de las MIP los siguientes componentes: un par de cebadores de indexado, un único código de barras de muestra y un par de adaptadores de secuenciación, y en donde los amplicones de las MIP con código de barra pueden comprender en secuencia los siguientes componentes:

35 un primer adaptador de secuenciación, un primer cebador de secuenciación, la primera etiqueta molecular de focalización única, el primer brazo de polinucleótidos de focalización, el ácido nucleico capturado, el segundo brazo de polinucleótidos de focalización, la segunda etiqueta única de focalización, un único código de barras de muestra, un segundo cebador de secuenciación, un segundo adaptador de secuenciación.

La primera pluralidad de secuencias diana de interés puede estar sobre un único cromosoma.

40 La segunda pluralidad de secuencias diana de interés puede estar sobre múltiples cromosomas. La métrica de captura del sitio que se determina en la etapa f) puede ser el número de eventos de captura que se determinó en la etapa d), y la métrica de captura del sitio que se determinó en la etapa h) puede ser el número de eventos de captura que se determinó en la etapa e).

45 El método puede comprender, además, computar un coeficiente de variabilidad para una pluralidad de métricas de captura del sitio para un sitio particular, en donde cada métrica de captura del sitio en la pluralidad de métricas de captura del sitio se evalúa a partir de una muestra de ácido nucleico de un sujeto diferente, y en donde el, al menos, un criterio que se utiliza en las etapas g) y h) incluye un requisito de que el coeficiente de variabilidad para el sitio particular esté por debajo de un valor umbral.

La primera medida que se determinó en la etapa j) puede ser una suma del primer subconjunto de métrica de captura del sitio y corresponde a un cromosoma de interés, y la segunda medida que se determina en la etapa j) puede ser una suma del segundo subconjunto de métrica de captura del sitio, y corresponde a los cromosomas además del cromosoma de interés.

50 La que se determina en la etapa k) puede comprender realizar una prueba estadística para evaluar si la relación de prueba que se obtiene en la etapa j) es estadísticamente diferente de la pluralidad de las relaciones de referencia.

La primera población de amplicones puede corresponder a un cromosoma de interés.

La segunda población de amplicones puede corresponder a cromosomas además del cromosoma de interés.

La relación de prueba y las relaciones de referencia pueden ser fracciones cromosómicas, en donde las fracciones cromosómicas se pueden definir mediante una relación entre una suma de todos los eventos de captura únicos a partir de un cromosoma de interés (S1) y una suma de todos los eventos de captura únicos a partir de todos los cromosomas (S1 + S2).

5 El tamaño del replicón de la MIP puede estar entre 80-90 pares base.

La etapa de secuenciación puede tener una profundidad de lectura de entre 6-8 millones de lecturas.

La secuencia diana de interés puede estar ubicada en un elemento Alu, en donde la secuencia diana de interés puede estar ubicada en el brazo derecho de un elemento Alu.

10 La aneuploidía puede ser una aneuploidía autosómica, y los números de eventos de captura que se determinan en las etapas d) y e) pueden excluir cualquier evento de captura a partir de los cromosomas sexuales.

La aneuploidía puede ser una aneuploidía cromosómica sexual, y los números de eventos de captura que se determinan en las etapas d) y e) pueden incluir eventos de captura de al menos un cromosoma sexual.

La presente invención comprende, además, un método de detección sistemática para la aneuploidía en un feto, que comprende:

- 15 a) obtener una muestra de ADN genómico a partir de una muestra de sangre materna;
- b) agregar la muestra de ADN genómico en cada pocillo de una placa de múltiples pocillos, en donde cada pocillo de la placa de múltiples pocillos comprende una mezcla de sonda, en donde la mezcla de sonda comprende una población de sondas de inversión molecular (MIP) y un amortiguador;
 en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes:
 - 20 primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización;
 en donde el par del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en una pluralidad de secuencias diana de interés; en donde la
 - 25 primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;
 - c) incubar la muestra de ADN genómico con la mezcla de sonda para que las MIP capturen la pluralidad de secuencias diana de interés;
 - d) agregar una mezcla de extensión/ligación a la muestra de c) para que las MIP y la pluralidad de secuencias de interés formen una pluralidad de amplicones de las MIP, en donde la mezcla de extensión/ligación comprende una polimerasa, una pluralidad de dNTP, una ligasa y un amortiguador;
 - e) agregar una mezcla de exonucleasa a los amplicones de las MIP de focalización y control para eliminar las sondas en exceso o el ADN genómico en exceso;
 - f) agregar una mezcla de PCR de indexado a la muestra de e) para agregar un par de cebadores de indexado, un
 - 35 único código de barras de muestra y un par de adaptadores de secuenciación a la pluralidad de amplicones;
 - g) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una primera población de amplicones con código de barras que se proporcionan en la etapa f), basado en el número de las etiquetas moleculares de focalización únicas, en donde la primera población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés;
 - h) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una segunda población de amplicones con código de barras que se proporcionan en la etapa f), basado en el número de las etiquetas moleculares de focalización únicas, en donde la segunda población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés;
 - i) computar la métrica de captura del sitio basado, al menos, en parte en el número de las primeras lecturas de secuenciación que se determinaron en la etapa g) y una pluralidad de métricas de captura de sonda de control basado, al menos, en parte en los números de las segundas lecturas de secuenciación que se determinaron en la etapa h);
 - 45 j) identificar un subconjunto de métricas de captura del sitio de la población de amplicones de las MIP que tienen métricas de captura de la sonda de control que satisfacen, al menos, un criterio;
 - k) normalizar la métrica de captura del sitio mediante un factor que se computa a partir del subconjunto de métricas de captura de sonda de control que satisfacen el, al menos, un criterio, para obtener una métrica de captura del sitio normalizada de prueba;
 - 50 l) comparar la métrica de captura del sitio normalizada de prueba con una pluralidad de métricas de captura del sitio normalizadas de referencia, que se computan basado en las muestras de ADN genómico de referencia de

los sujetos de referencia que exhiben genotipos conocidos, mediante el uso de los mismos sitios específicos y de control, la población específica, el subconjunto de poblaciones de control en las etapas b)-h), para determinar la presencia o la ausencia de una métrica de captura del sitio normalizada de prueba que indica la aneuploidía en el feto.

- 5 La muestra de sangre puede ser una muestra de sangre entera, una muestra de plasma o una muestra de suero.
 La extensión del primer brazo de polinucleótidos de focalización puede estar entre 14 y 30 pares base.
 La extensión del segundo brazo de polinucleótidos de focalización puede estar entre 14 y 30 pares base.
 Cada uno de los brazos de polinucleótidos de focalización puede tener una temperatura de fusión de entre 45 °C y 80 °C.
- 10 Cada uno de los brazos de polinucleótidos de focalización puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %.
 La extensión de la primera etiqueta molecular única puede estar entre 4 y 15 pares base.
 La extensión de la segunda etiqueta molecular única puede estar entre 4 y 15 pares base.
 Cada una de las etiquetas moleculares únicas puede tener una temperatura de fusión de entre 45 °C y 80 °C.
- 15 Cada una de las etiquetas moleculares únicas puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %.
 Preferiblemente, el ligador polinucleotídico no es sustancialmente complementario con alguna región genómica del sujeto. El ligador polinucleotídico puede tener una extensión de entre 20 y 1.000 pares base.
 El ligador polinucleotídico puede tener una temperatura de fusión de entre 45 °C y 80 °C.
 El ligador polinucleotídico puede tener un contenido de GC entre 30 % y 80 %, o entre 30 % y 70 %.
- 20 El ligador polinucleotídico puede comprender, al menos, un cebador de amplificación.
 Preferiblemente, el ligador polinucleotídico comprende un cebador de amplificación hacia adelante y un cebador de amplificación inverso.
 La secuencia del cebador de amplificación hacia adelante puede comprender la secuencia de nucleótidos de 5' - CTTAGCT-TCCCATTACGG - 3' (SEC ID NO: 1).
- 25 Preferiblemente, la secuencia del cebador de amplificación inverso comprende la secuencia de nucleótidos de 5' - GCAC-GATCCGACGGTAGTGT - 3' (SEC ID NO: 2).
 El ligador polinucleotídico puede comprender la secuencia de nucleótidos de 5' - CTTAGCTTCCCATTACGGGCAC-GATCCGACGGTAGTGT - 3' (SEC ID NO: 3).
- 30 El primer brazo de polinucleótidos de focalización puede comprender la secuencia de nucleótidos de 5'- CACTGCACTCCAGCCTGG-3' (SEC ID NO: 4).
 El segundo brazo de polinucleótidos de focalización puede comprender la secuencia de nucleótidos de 5'- GAGGCTGAGGACG - GAGAA - 3' (SEC ID NO: 5).
 La MIP puede comprender la secuencia de nucleótidos de 5'-CACTGCACTCCA GCCTGG(N1-6)CTTAGCTTC-CCGATTACGGGCACGATCCGA CGGTAGTGT(N7-12)GAGGCTGAGGCAGGAGAA-3' (SEC ID NO: 6), en donde (N1-6) representa la primera etiqueta molecular única y (N7-12) representa la segunda etiqueta molecular única.
- 35 La población de MIP puede tener una concentración de entre 10 fM y 100 nM.
 El tamaño del replicón de la MIP puede estar entre 80-90 pares base.
 La etapa de secuenciación puede tener una profundidad de lectura de entre 6-8 millones de lecturas.
- Breve descripción de los dibujos**
- 40 **La fig. 1** es una realización ilustrativa de un dispositivo de computación para realizar cualquiera de los procedimientos que se describen de acuerdo con los métodos de la divulgación.
La fig. 2 es un diagrama de flujo del procedimiento representativo para diseñar y seleccionar una sonda, de acuerdo con algunos métodos de la divulgación.
- 45 **La fig. 3** es un diagrama de flujo del procedimiento representativo para pronosticar el estado de aneuploidía en un sujeto de prueba, de acuerdo con algunos métodos de la divulgación.
La fig. 4 es otro diagrama de flujo del procedimiento representativo y más detallado para pronosticar el estado de aneuploidía de un sujeto de prueba, de acuerdo con algunos métodos de la divulgación.
La fig. 5 muestra la secuencia de una sonda de inversión molecular (MIP) ilustrativa que se utiliza en algunos métodos de esta divulgación. La MIP comprende en secuencia los siguientes componentes: un primer brazo de polinucleótidos

de focalización (etiquetado «brazo de ligación»), un ligador polinucleotídico (etiquetado «esqueleto») y comprende una primera etiqueta molecular de focalización única (etiquetada «6N»), un cebador de PCR hacia adelante, un cebador de PCR inverso, una segunda etiqueta molecular de focalización única (también etiquetada «6N») y un segundo brazo de polinucleótidos de focalización (etiquetado «brazo de extensión»). El primero y el segundo brazo de polinucleótidos de focalización en cada una de las MIP son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean un sitio de interés. Las etiquetas moleculares únicas son secuencias de polinucleótidos aleatorias. En algunas realizaciones, la frase «sustancialmente complementaria» se refiere a 0 apareamiento erróneo en ambos brazos, o como máximo 1 apareamiento erróneo en solo un brazo (por ej., cuando los brazos de polinucleótidos de focalización hibridan a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean un sitio de interés). En algunas realizaciones, la frase «sustancialmente complementaria» se refiere, como máximo, a un pequeño número de apareamientos erróneos en ambos brazos, como por ej., 1, 2, 3, 3, 5, 6, 7 u 8.

La fig. 6 describe la hibridación de la sonda y la extensión/ligación en un método de la divulgación. La MIP se agrega al ADN en condiciones adecuadas para la hibridación del primer brazo de polinucleótidos de focalización (etiquetado «brazo de ligación») y el segundo brazo de polinucleótidos de focalización (etiquetado «brazo de extensión») al molde de ADN. Después de la hibridación, se agregan una polimerasa y una ligasa en condiciones de extensión/ligación, y un oligonucleótido circular (la «sonda capturada») es producido mediante la síntesis de ADN a lo largo de la secuencia diana de interés que contiene la secuencia con huecos única entre los brazos de ligación y extensión. Al cabo de la fusión del amplicón y del ADNcs, la sonda capturada ya está lista para la amplificación.

La fig. 7 describe la amplificación y la secuenciación de las sondas capturadas. Las moléculas de ácido nucleico que comprenden un adaptador de secuenciación y un cebador de PCR hacia adelante o inverso se enlazan al esqueleto del amplicón circular, y todos los oligonucleótidos circulares que han sido creados por las MIP se amplifican mediante el uso de PCR. Posteriormente, los amplicones se secuencian mediante el uso de, por ejemplo, la secuenciación de próxima generación (NGS), y el recuento de lectura para los amplicones resultantes se determina mediante el recuento del número de ocurrencias de las etiquetas moleculares únicas en cada amplicón.

La fig. 8 describe los resultados de una prueba para la trisomía del cromosoma 21 (síndrome de Down). De las pruebas de 48 muestras, 46 fueron negativas para el síndrome de Down, mientras que 2 muestras fueron positivas para el síndrome de Down. Las dos muestras positivas se muestran en la parte derecha superior, con una puntuación Z superior a 6.

La fig. 9 describe la hibridación de la sonda a un elemento Alu en una realización de la divulgación.

Las figuras 10-12 describen el rendimiento de la prueba según se evaluó mediante una MIP ilustrativa en la detección de la trisomía 13, 18 y 21, respectivamente.

Descripción detallada de la invención

La invención se presenta en las reivindicaciones adjuntas. Las realizaciones de la descripción que no están dentro del alcance de dichas reivindicaciones se proporcionan a los fines ilustrativos únicamente, y no forman parte de la presente invención. La divulgación proporciona un sistema y un método para detectar la aneuploidía.

A fin de que la divulgación que se describe en la presente memoria se pueda entender completamente, se establece la siguiente descripción detallada.

A menos que se defina lo contrario en la presente memoria, los términos técnicos y científicos que se utilizan en esta solicitud tienen los significados que normalmente entienden los expertos en la técnica, a los cuales pertenece esta divulgación. En general, la nomenclatura y las técnicas que se utilizan con relación al cultivo de células y tejidos, a la biología molecular, la biología celular, la biología del cáncer, la neurobiología, la neuroquímica, la virología, la inmunología, la microbiología, la genética, la química de las proteínas y el ácido nucleico, la química y la farmacología que se describen en la presente memoria, son las que se conocen y se utilizan, normalmente, en la técnica. Cada realización de la divulgación que se describe en la presente memoria se puede tomar sola o en combinación con una o más de otras realizaciones de la divulgación.

Los métodos y las técnicas de la presente divulgación se llevan a cabo, en general, a menos que se indique lo contrario, de acuerdo con los métodos bien conocidas en la técnica, de la biología molecular, la biología celular, la bioquímica, la tecnología de micromatrices y secuenciación, y según se describen en varias referencias generales y más específicas que se mencionan y se discuten a lo largo de esta memoria descriptiva. Ver por ej., Motulsky, «Intuitive Biostatistics», Oxford University Press, Inc. (1995); Lodish et al., «Molecular Cell Biology, 4^o ed.», W. H. Freeman & Co., New York (2000); Griffiths et al., «Introduction to Genetic Analysis, 7^o ed.», W. H. Freeman & Co., N.Y. (1999); Gilbert et al., «Developmental Biology, 6^o ed.», Sinauer Associates, Inc., Sunderland, MA (2000).

Los términos químicos que se utilizan en la presente memoria, se utilizan de acuerdo con el uso convencional en la técnica, como se ejemplifica en «The McGraw-Hill Dictionary of Chemical Terms», Parker S., Ed., McGraw-Hill, San Francisco, C.A. (1985).

En caso de conflicto, la presente memoria descriptiva, incluso sus definiciones específicas, prevalecerá.

A lo largo de la presente memoria descriptiva, la palabra «comprender» o sus variantes como «comprende» o «que

comprende» se entenderá que implica la inclusión de un entero establecido (o componentes) o un grupo de enteros (o componentes), pero no la exclusión de cualquier otro entero (o componentes) o grupo de enteros (o componentes).

Las formas singular «un», «una» y «el/la» incluyen la forma plural a menos que el contexto indique claramente lo contrario.

- 5 El término «que incluye» se utiliza para indicar «que incluye pero sin estar limitado». Las frases «que incluye» y «que incluye pero sin estar limitado» se utilizan intercambiamente.

A fin de definir mejor la divulgación, los siguientes términos y definiciones se proporcionan en la presente memoria.

Definiciones

- 10 El término «aneuploidía», como se utiliza en la presente memoria, se refiere a una anomalía cromosómica que se caracteriza por una variante anómala en el número de cromosomas, por ej., un número de cromosomas que no es un múltiple exacto del número haploide de cromosomas. Por ejemplo, un individuo euploide tendrá un número de cromosomas que es igual a $2n$, donde n es el número de cromosomas en el individuo haploide. En los humanos, el número haploide es 23. De este modo, un individuo diploide tendrá 46 cromosomas. Un individuo aneuploide puede contener una copia extra de un cromosoma (trisomía de ese cromosoma) o carecer de una copia del cromosoma (monosomía de ese cromosoma). La variante anómala es con respecto a cada cromosoma individual. De este modo, un individuo con una trisomía y una monosomía es aneuploide, a pesar de tener 46 cromosomas. Ejemplos de enfermedades o afecciones de aneuploidía incluyen, pero sin estar limitadas, síndrome de Down (trisomía del cromosoma 21), síndrome de Edwards (trisomía del cromosoma 18), síndrome de Patau (trisomía del cromosoma 13), síndrome de Turner (monosomía del cromosoma X en una niña), y el síndrome de Klinefelter (una copia extra del cromosoma X en un varón). Otras anomalías cromosómicas no aneuploide incluye traslocación (en donde un segmento de un cromosoma ha sido transferido a otro cromosoma), supresión (en donde una parte de un cromosoma se ha perdido), y otros tipos de daño cromosómico (por ej., síndrome de X frágil, el cual se origina por un cromosoma X que es anormalmente susceptible al daño).

- 25 En otras realizaciones de la divulgación, los métodos se pueden utilizar para detectar las variantes en el número de copias. Como se utiliza en la presente memoria, una «variante en el número de copias» es, en general, una clase o tipo de variante genética o aberración cromosómica. Una variante en el número de copias puede ser una supresión (por ej., micro-supresión), una duplicación (por ej., una micro-duplicación), o una inserción (por ej., una micro-inserción). En ciertas realizaciones, el prefijo «micro» como se utiliza en la presente memoria, puede referirse a un segmento de un ácido nucleico que tiene menos de 5 pares base en extensión. Una variante en el número de copias puede incluir una o más supresiones (por ej., micro-supresión), duplicaciones y/o inserciones (por ej., una micro-duplicación, una micro-inserción) de un segmento de un cromosoma. En ciertas realizaciones una duplicación comprende una inserción. En ciertas realizaciones, una inserción es una duplicación. En ciertas realizaciones una inserción no es una duplicación. Por ejemplo, una duplicación de una secuencia en una parte aumenta los conteos para una parte en la cual se encuentra la duplicación. Con frecuencia una duplicación de una secuencia en una parte aumenta la elevación o el nivel. En ciertas realizaciones, una duplicación presente en las partes que crean una primera elevación o nivel aumenta la elevación o el nivel con relación a una segunda elevación o nivel donde una duplicación está ausente. En ciertas realizaciones, una inserción aumenta los conteos de una parte y está presente una secuencia que representa la inserción (es decir, duplicada) en otra ubicación dentro de la misma parte. En ciertas realizaciones, una inserción no aumenta, significativamente, los conteos de una parte o elevación o nivel, y la secuencia que se inserta no es una duplicación de una secuencia dentro de la misma parte. En ciertas realizaciones, una inserción no se detecta o no está representada como una duplicación y una secuencia duplicada que representa la inserción no está presente en la misma parte. En algunas realizaciones, una variante en el número de copias es una variante en el número de copias fetales. Con frecuencia, una variante en el número de copias fetales es una variante en el número de copias en el genoma de un feto. En algunas realizaciones, una variante en el número de copias es una variante en el número de copias maternas y/o fetales. En ciertas realizaciones, una variante en el número de copias maternas y/o fetales es una variante en el número de copias dentro del genoma de una mujer embarazada (por ej., un sujeto femenino que porta un feto), un sujeto femenino que dio a luz o es capaz de portar un feto. Una variante en el número de copias puede ser una variante en el número de copias heterocigotas donde la variante (por ej., una duplicación o supresión) está presente sobre un alelo de un genoma. Una variante en el número de copias puede ser una variante en el número de copias homocigotas donde la variante está presente sobre ambos alelos de un genoma. En algunas realizaciones, una variante en el número de copias es una variante en el número de copias fetales heterocigotas u homocigotas. En algunas realizaciones, una variante en el número de copias es una variante en el número de copias maternal y/o fetal, heterocigota u homocigota. Una variante en el número de copias, algunas veces, está presente en un genoma materno y en un genoma fetal, en un genoma materno y no en un genoma fetal, o en un genoma fetal y no en un genoma materno.

- 55 Los términos «sujeto» y «paciente», como se utilizan en la presente memoria, se refieren a cualquier animal, como por ej., un perro, un gato, un pájaro, ganado, y en particular, un mamífero, y preferiblemente, un ser humano. Las frases «sujeto de referencia» y «pacientes de referencia» se refieren a cualquier sujeto o paciente que exhibe genotipos conocidos (por ej., euploidía o aneuploidía conocida).

- 60 Los términos «polinucleótido», «ácido nucleico» y «moléculas de ácido nucleico», como se utilizan en la presente memoria, se utilizan intercambiamente, y se refieren a moléculas de ADN (por ej., ADNc o ADN genómico), moléculas de ARN (por ej., ARNm), híbridos de ADN-ARN, y análogos del ADN o ARN generados mediante el uso de

análogos de nucleótidos. La molécula de ácido nucleico puede ser un nucleótido, un oligonucleótido, ADN bicatenario, ADN monocatenario, ADN de múltiples cadenas, ADN complementario, ADN genómico, ADN no codificante, ARN mensajero (ARNm), ARN micro (ARNm), ARN pequeño nucleolar (ARNsno), ARN ribosómico (ARNr), ARN de transferencia (ARNt), ARN interferente pequeño (ARNsi), ARN nuclear heterogéneo (ARNhr), o pequeño ARN de horquilla (ARNsh).

El término «muestra», como se utiliza en la presente memoria, se refiere a una muestra que deriva, típicamente, de un fluido biológico, una célula, un tejido, un órgano o un organismo, que comprende un ácido nucleico o una mezcla de ácidos nucleicos que comprende, al menos, una secuencia de ácidos nucleicos que debe ser detectada para, por ej., aneuploidía u otras anomalías cromosómicas. En algunas realizaciones, una muestra es una muestra de sangre como por ej., una muestra de sangre entera, una muestra de suero o una muestra de plasma. En algunas realizaciones, la muestra comprende, al menos, una secuencia de ácido nucleico cuyo genoma se sospecha que tiene una variante sometida. Estos ejemplos incluyen, pero sin estar limitados, esputo/fluido bucal, líquido amniótico, sangre, una fracción de sangre, o muestras de punción exploradora (por ej., biopsia quirúrgica, punción exploradora, etc.), orina, líquido peritoneal, líquido pleural, y similares. Si bien la muestra, con frecuencia, se toma de un sujeto humano (por ej., paciente), los ensayos se pueden utilizar para detectar aneuploidía en muestras de cualquier mamífero, que incluye, pero sin estar limitados, perros, gatos, caballos, cabras, ovejas, ganado vacuno, cerdos, etc. La muestra se puede utilizar directamente según se obtiene de la fuente biológica o después de un pretratamiento para modificar el carácter de la muestra. Por ejemplo, dicho pretratamiento puede incluir la preparación del plasma a partir de la sangre, la dilución de los líquidos viscosos, etc. Los métodos de pretratamiento pueden involucrar, además, pero sin estar limitados, filtración, precipitación, dilución, destilación, mezclado, centrifugación, congelación, liofilización, concentración, amplificación, fragmentación del ácido nucleico, inactivación de los componentes interferentes, la adición de reactivos, el lisado, etc. Si dichos métodos de pretratamiento se emplean con respecto a la muestra, dichos métodos de pretratamiento son típicamente tales, que los ácidos nucleicos de interés permanecen en la muestra de prueba, preferiblemente, a una concentración proporcional a la de una muestra de prueba no tratada (por ej., concretamente, una muestra que no se somete a ningún método de pretratamiento de este tipo). Según el tipo de muestra que se utilice, se pueden realizar etapas de procesamiento y/o purificación adicionales para obtener fragmentos de ácido nucleico de una pureza o tamaño deseado, mediante el uso de métodos de procesamiento que incluyen, pero sin estar limitados, sonicación, nebulización, purificación en gel, sistemas de purificación por PCR, escisión de nucleasa, captura o exclusión específica del tamaño, captura focalizada o una combinación de estos métodos. Opcionalmente, el ADN libre de células se puede aislar de la muestra antes de un análisis posterior. En algunas realizaciones, la muestra es del sujeto cuya euploidía o aneuploidía se debe determinar mediante los sistemas y métodos de la divulgación, a la que se hace referencia también como «una muestra de prueba».

El término «MIP», como se utiliza en la presente memoria, se refiere a una sonda de inversión molecular (también conocida como sonda de captura circular). Tal como se utiliza en la presente memoria, el término «cebador» o «sonda» también hace referencia a una MIP. Las sondas de inversión molecular son moléculas de ácido nucleico que contienen dos brazos de polinucleótidos de focalización, una o más etiquetas moleculares únicas (también conocidas como identificadoras moleculares únicas) y un ligador polinucleotídico (por ej., un ligador de esqueleto universal). Ver, por ejemplo la figura 5. En algunas realizaciones, una MIP puede comprender más de una etiqueta molecular única, como por ej., dos etiquetas moleculares únicas, tres etiquetas moleculares únicas, o más. En algunas realizaciones, los brazos de polinucleótidos únicos en cada MIP se ubican en los extremos 5' y 3' de la MIP, mientras que las etiquetas moleculares únicas y el ligador de polinucleótidos se ubican en el medio. Por ejemplo, las MIP que se utilizan en la divulgación comprenden en secuencia los siguientes componentes: primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización. En algunas realizaciones, el ligador polinucleotídico (o el ligador del esqueleto) en las MIP es universal en todas las MIP que se utilizan en un método de la divulgación.

En las MIP, los brazos de polinucleótidos únicos están diseñados para hibridar, de inmediato, corriente arriba y corriente abajo de una secuencia diana específica (o sitio) de interés en una muestra de ácido nucleico genómico. Como se utiliza en la presente memoria, las frases «secuencia diana de interés» y «sitio blanco de interés» se utilizan intercambiabilmente para referirse a una parte de la molécula de ácido nucleico genómico para cuya captura está diseñada la MIP. En algunas realizaciones, los brazos de polinucleótidos únicos son complementarios a la corriente arriba y corriente abajo inmediata de una o más secuencias de interés (o sitios de interés) en una muestra de ácido nucleico genómico. En algunas realizaciones, estos brazos de polinucleótidos únicos son complementarios a una o más secuencias de interés (o sitios de interés) en una muestra de ácido nucleico genómico. En algunas realizaciones, los brazos de polinucleótidos únicos comprenden una secuencia de ligación y una secuencia de extensión. Una MIP que comprende brazos de polinucleótidos de focalización que son complementarios a una pluralidad de secuencias de interés en una muestra de ADN se puede hacer referencia como una «MIP causante de repetición» o «RO-MIP»: Por ejemplo, una RO-MIP puede focalizar cientos, miles, cientos de miles, o millones de secuencias de interés de una muestra de ADN (por ej., una muestra que comprende un genoma humano). En algunas realizaciones, una RO-MIP focaliza, por ejemplo más que 1.000, más que 10.000, más que 20.000, más que 30.000, más que 40.000, más que 50.000, más que 60.000, más que 70.000, más que 80.000, más que 90.000, más que 100.000, más que 200.000, más que 300.000, más que 400.000, más que 500.000, más que 600.000, más que 700.000, más que 800.000, más que 900.000, y/o más que 1.000.000 secuencias de interés. En algunas realizaciones, una RO-MIP focaliza, por ejemplo más que 100.000, más que 110.000, más que 120.000, más que 130.000, más que 140.000, más que 150.000, más que 160.000, más que 170.000, más que 180.000, más que 190.000, y/o más que 200.000 secuencias de interés, o

cualquier intervalo entre 100.000 y 200.000 secuencias de interés. En algunas realizaciones, una RO-MIP focaliza 140.000-160.000 secuencias de interés.

Estas secuencias de interés pueden ser flanqueadas por secuencias de repetición a las cuales hibridan los brazos de polinucleótidos de focalización. En ciertas realizaciones, las secuencias de repetición tienen 0, 1, 2, 3, 4 o más apareamientos erróneos al hibridar con los brazos de polinucleótidos de focalización. En realizaciones específicas, las secuencias de repetición tienen 0 o 1 apareamiento erróneo al hibridar con los brazos de polinucleótidos de focalización. En algunas realizaciones, una RO-MIP no se une a los elementos de nucleótidos intercalados largos (LÍNEA) en el genoma.

En algunas realizaciones, las etiquetas moleculares únicas son secuencias cortas de nucleótidos que se generan aleatoriamente. En ciertas realizaciones, las etiquetas moleculares únicas no hibridan a cualquier secuencia o sitio ubicado sobre un fragmento de ácido nucleico genómico o en una muestra de ácido nucleico genómico. En ciertas realizaciones, la etiqueta molecular única es cualquier etiqueta con una marca detectable adecuada que se puede incorporar o unirse a un ácido nucleico (por ej., un polinucleótido) que permite la detección y/o identificación de los ácidos nucleicos que comprenden o se unen a la etiqueta. En algunas realizaciones, la etiqueta se incorpora o se une a un ácido nucleico durante un método de secuenciación (por ej., mediante una polimerasa). Ejemplos no limitantes de etiquetas incluyen etiquetas de ácido nucleico, índices de ácido nucleico o códigos de barra, una radiomarca (por ej., un isótopo), una etiqueta metálica, una etiqueta fluorescente, una etiqueta quimioluminiscente, una etiqueta fosforescente, un desactivador de fluoróforo, un tinte, una proteína (por ej., una enzima, un anticuerpo o una de sus partes, un ligador, un miembro de un par de unión), similares, o sus combinaciones. En algunas realizaciones, la etiqueta (por ej., un índice de ácido nucleico o código de barras) es una secuencia única, conocida y/o identificable de nucleótidos o análogos nucleotídicos. En algunas realizaciones, las etiquetas son seis o más nucleótidos contiguos. Una multitud de fluoróforos están disponibles con una variedad de diferentes espectros de excitación y emisión. Cualquier tipo y/o número adecuado de fluoróforos se puede utilizar como etiqueta. En algunas realizaciones, 1 o más, 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 20 o más, 30 o más, 50 o más, 100 o más, 500 o más, 1000 o más, 10.000 o más, 100.000 o más etiquetas diferentes se utilizan en un método que se describe en la presente memoria (por ej., un método de detección y/o secuenciación del ácido nucleico). En algunas realizaciones, uno o dos tipos de etiquetas (por ej., etiquetas fluorescentes) se enlazan a cada ácido nucleico en una biblioteca. En algunas realizaciones, las etiquetas específicas del cromosoma se utilizan para realizar el conteo cromosómico con más rapidez y facilidad. La detección y/o cuantificación de una etiqueta se puede realizar mediante un método, una máquina o un aparato adecuado; ejemplos no limitantes de los cuales incluyen citometría de flujo, reacción en cadena de polimerasa cuantitativa (PCRq), electroforesis, un luminómetro, un fluorómetro, un espectrofotómetro, un chip adecuado o análisis de micromatrices, inmunoelectrotransferencia, espectrometría de masa, cromatografía, análisis citofluorimétrico, microscopía de fluorescencia, un método de formación de imágenes por fluorescencia o digital, microscopía confocal de barrido con láser, citometría de barrido con láser, cromatografía por afinidad, separación discontinua manual, suspensión del campo eléctrico, un método de secuenciación de ácido nucleico adecuado y/o un aparato de secuenciación del ácido nucleico, similares, y sus combinaciones. En realizaciones particulares, la etiqueta es adecuada para uso con en análisis de micromatrices.

Las MIP se introducen a los ácidos nucleicos (por ej., fragmentos de ácido nucleico) para realizar la captura de secuencias o sitios específicos ubicados sobre una muestra de ácido nucleico (por ej., un ADN genómico). En algunas realizaciones, por ejemplo, si el ADN genómico está presente en una muestra, la fragmentación puede ayudar en la captura del ácido nucleico específico mediante sondas de inversión molecular. Como se describe en mayor detalle en la presente memoria, después de la captura de la secuencia específica (por ej., locus) de interés, el blanco capturado puede ser, además, sometido a una etapa de llenado de huecos y ligación enzimática, de modo que una copia de la secuencia específica se incorpora a un círculo. La eficiencia de captura de la MIP a la secuencia específica sobre el fragmento de ácido nucleico puede ser mejorada mediante el alargamiento de los periodos de hibridación e incubación por llenado de huecos. (Ver, por ej., Turner E. H.; et al., *Nat Methods* 2009 Abr. 6:1-2.)

La tecnología de MIP puede ser utilizada para detectar o amplificar las secuencias de ácido nucleico particulares en mezclas complejas. Una de las ventajas de utilizar la tecnología de MIP está en su capacidad de un alto grado de multiplexación, lo cual permite que miles de secuencias específicas se capturen en una única reacción que contiene miles de MIP. Varios aspectos de la tecnología MIP se describen, por ejemplo, en Hardenbol et al., «Multiplexed genotyping with sequence-tagged molecular inversion probes», *Nature Biotechnology*, 21(6): 673-678 (2003); Hardenbol et al., «Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay», *Genome Research*, 15: 269-275 (2005); Burmester et al., «DMET microarray technology for pharmacogenomics-based personalized medicine», *Methods in Molecular Biology*, 632: 99-124 (2010); Sissung et al., «Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform», *Pharmacogenomics*, 11(1): 89-103 (2010); Deeken, «The Affymetrix DMET platform and pharma- cogenetics in drug development», *Current Opinion in Molecular Therapeutics*, 11(3): 260-268 (2009); Wang et al., «High quality copy number and genotype data from FFPE samples using Molecular Inversion Probe (MIP) microarrays», *BMC Medical Genomics*, 2:8 (2009); Wang et al., «Analysis of molecular inversion probe performance for allele copy number determination», *Genome Biology*, 8(11): R246 (2007); Ji et al., «Molecular inversion probe analysis of gene copy alternations reveals distinct categories of colorectal carcinoma», *Cancer Research*, 66(16): 7910-7919 (2006); y Wang et al., «Allele quantification using molecular inversion probes (MIP)», *Nucleic Acids Research*, 33(21): e183 (2005). Ver, además, las Patentes estadounidenses Nos. 6.858.412, 5.817.921, 6.558.928, 7.320.860, 7.351.528, 5.866.337, 6.027.889 y 6.852.487.

La tecnología MIP ha sido previamente aplicada con éxito a otras áreas de investigación, que incluyen la identificación novedosa y la subclasificación de biomarcadores en cánceres. Ver, por ej., Brewster et al., «Copy number imbalances between screen- and symptom-detected breast cancers and impact on disease-free survival», *Cancer Prevention Research*, 4(10): 1609-1616 (2011); Geiersbach et al., «Unknown partner for USP6 and unusual SS18 rearrangement detected by fluorescence in situ hybridization in a solid aneurysmal bone cyst», *Cancer Genetics*, 204(4): 195-202 (2011); Schiffman et al., «Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas», *Cancer Research*, 70(2): 512-519 (2010); Schiffman et al., «Molecular inversion probes reveal patterns of 9p21 deletion and copy number aberrations in childhood leukemia», *Cancer Genetics and Cytogenetics*, 193(1): 9-18 (2009); Press et al., «Ovarian carcinomas with genetic and epigenetic BRCA1 loss have distinct molecular abnormalities», *BMC Cancer*, 8:17 (2008); and Deeken et al., «A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform», *Pharmacogenomics*, 10(3): 191-199 (2009).

La tecnología MIP ha sido aplicada a la identificación de nuevos biomarcadores relacionados con los medicamentos. Ver, por ej., Caldwell et al., «CYP4F2 genetic variant alters required warfarin dose», *Blood*, 111(8): 4106-4112 (2008); and McDonald et al., «CYP4F2 Is a Vitamin K1 Oxidase: An Explanation for Altered Warfarin Dose in Carriers of the V433M Variant», *Molecular Pharmacology*, 75: 1337-1346 (2009). Otras aplicaciones de la MIP incluyen el desarrollo de medicamentos y la investigación de seguridad. Ver, por ej., Mega et al., «Cytochrome P-450 Polymorphisms and Response to Clopidogrel», *New England Journal of Medicine*, 360(4): 354-362 (2009); Dumaul et al., «Comprehensive assessment of metabolic enzyme and transporter genes using the Affymetrix Targeted Genotyping System», *Pharmacogenomics*, 8(3): 293-305 (2007); and Daly et al., «Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport», *Clinical Chemistry*, 53(7): 1222-1230 (2007). Otras aplicaciones de la tecnología MIP incluyen la base de datos del genotipo y el fenotipo. Ver, por ej., Man et al., «Genetic Variation in Metabolizing Enzyme and Transporter Genes: Comprehensive Assessment in 3 Major East Asian Subpopulations With Comparison to Caucasians and Africans», *Journal of Clinical Pharmacology*, 50(8): 929-940 (2010).

El término «captura» o «que captura», como se utiliza en la presente memoria, se refiere a la reacción de unión o hibridación entre una sonda de inversión molecular y el correspondiente sitio de focalización.

El término «sensibilidad», como se utiliza en la presente memoria, se refiere a una medida estadística de rendimiento de un ensayo (por ej., método, prueba), que se calcula dividiendo el número de positivos ciertos por la suma de los positivos ciertos y los negativos falsos.

El término «especificidad», como se utiliza en la presente memoria, se refiere a una medida estadística de rendimiento de un ensayo (por ej., método, prueba), que se calcula dividiendo el número de negativos ciertos por la suma de los negativos ciertos y los positivos falsos.

El término «replicón de MIP» o «replicón circular», como se utiliza en la presente memoria, se refiere a una molécula de ácido nucleico circular que se generó mediante una reacción de captura (por ej., una reacción de unión o hibridación entre una MIP y su secuencia de focalización). En algunas realizaciones, el replicón de MIP puede ser una molécula de ácido nucleico circular monocatenaria. En algunas realizaciones, una MIP de focalización captura o hibrida a una secuencia o sitio específico. Después de la reacción de captura o hibridación, se introduce una mezcla de ligación/extensión para extender y ligar la región con huecos entre los dos brazos de polinucleótidos de focalización para formar moléculas de nucleótidos circulares monocatenarias, es decir, un replicón de MIP de focalización. Los replicones de MIP se pueden amplificar a través de la reacción en cadena de polimerasa (PCR) para producir una pluralidad de amplicones de MIP de focalización, los cuales son moléculas de nucleótidos bicatenarias.

El término «amplicón», como se utiliza en la presente memoria, se refiere a un ácido nucleico que se genera mediante la reacción de amplificación. En algunas realizaciones, el amplicón es una molécula de ácido nucleico monocatenaria. En algunas realizaciones, el amplicón es una molécula de ácido nucleico circular monocatenaria. En algunas realizaciones, el amplicón es una molécula de ácido nucleico bicatenaria. Por ejemplo, una MIP (por ej., una RO-MIP) captura o hibrida a una secuencia o sitio específico. Después de la reacción de captura o hibridación, se introduce una mezcla de ligación/extensión para extender y ligar la región con huecos entre los dos brazos de polinucleótidos de focalización para formar una molécula de nucleótido circular monocatenario, es decir, un replicón de MIP. El replicón de MIP se puede amplificar a través de la reacción en cadena de polimerasa (PCR) para producir una pluralidad de amplicones de MIP, los cuales son moléculas de nucleótidos bicatenarios. Los replicones y amplicones de MIP se pueden introducir a partir de una primera pluralidad de secuencias diana de interés (por ej., un cromosoma que se ensaya para aneuploidía) y una segunda pluralidad de secuencias diana de interés (por ej., secuencias específicas que se distribuyen a través del genoma).

El término «secuenciación», como se utiliza en la presente memoria, se utiliza en un sentido amplio y puede referirse a cualquier técnica conocida en la técnica que permite el orden de, al menos, algunos nucleótidos consecutivos en, al menos, parte de un ácido nucleico a ser identificado, que incluye sin limitación, al menos, parte de un producto de extensión o un inserto de vector. La secuenciación puede referirse, además, a una técnica que permite la detección de diferencias entre las bases nucleotídicas en una secuencia de ácido nucleico. Las técnicas de secuenciación ilustrativas incluyen la secuenciación focalizada, la secuenciación en tiempo real de única molécula, la secuenciación basada en la microscopía de electrones, la secuenciación mediada por transistor, la secuenciación directa, la secuenciación por bombardeo con microproyectiles aleatoria, la secuenciación Sanger por terminación didesoxi, la secuenciación focalizada, la secuenciación exón, la secuenciación de genoma entero, la secuenciación por hibridación (por ej., en una

matriz como por ej., una micromatriz), la pirosecuenciación, la electroforesis capilar, la electroforesis en gel, la secuenciación dúplex, la secuenciación del ciclo, la secuenciación de la extensión de base única, la secuenciación de fase sólida, la secuenciación de alto rendimiento, la secuenciación paralela masiva de la firma, la PCR de emulsión, la coamplificación a baja temperatura de enaturación de PCR (COLD-PCR), PCR múltiple, la secuenciación por terminador de tinte reversible, la secuenciación de extremo apareado, la secuencia a término cercano, la secuenciación de exonucleasa, la secuenciación por ligación, la secuenciación de lectura corta, la secuencia de única molécula, la secuenciación por síntesis, la secuenciación en tiempo real, la secuenciación por terminador inverso, la secuenciación por semiconductor de iones, la secuenciación nanoball, la secuenciación de nanoporos, la secuenciación 454, la secuenciación por analizador del genoma de Solexa, miSeq (Illumina), HiSeq 2000 (Illumina), HiSeq 2500 (Illumina), analizador del genoma de Illumina (Illumina), por Ion Torrent PGM™ (Life Technologies), MinION™ (Oxford Nanopore Technologies), tecnología en tiempo real SMRT™ (Pacific Biosciences), ligación de sonda-ancla (cPAL™) (Complete Genomics/BGI), secuenciación SOLiD®, secuenciación MS-PET, espectrometría de masa, y sus combinaciones. En algunas realizaciones, la secuenciación comprende la detección del producto de secuenciación mediante el uso de un instrumento, por ejemplo pero sin limitación, un secuenciador de ADN ABI PRISM® 377, un analizador genético ABI PRISM® 310, 3100, 3100-Avant, 3730 o 3730xl, un analizador de ADN ABI PRISM® 3700, o un sistema de Applied Biosystems SOLiD™ (todo de Applied Biosystems), una secuenciador del genoma 20 System (Roche Applied Science) o un espectrómetro de masa. En ciertas realizaciones, la secuenciación comprende la PCR en emulsión. En ciertas realizaciones, la secuenciación comprende una técnica de secuenciación de alto rendimiento, por ejemplo pero sin estar limitada, la secuenciación paralela masiva de la firma (MPSS).

Los métodos y aparatos que se describen en la presente memoria pueden emplear, de forma alternativa, la tecnología de micromatrices para cuantificar los productos de RO-MIP. «Micromatriz» o «matriz» se refiere a un soporte de fase sólida que tiene una superficie, preferiblemente, pero no exclusivamente una superficie plana o sustancialmente plana, que transporta una matriz de sitios que contiene ácidos nucleicos de modo que cada sitio de la matriz comprende copias sustancialmente idénticas o idénticas de oligonucleótidos o polinucleótidos, y se define espacialmente y no se superpone con otros sitios de miembros de la matriz; es decir los sitios son espacialmente discretos. La matriz o la micromatriz también pueden comprender una estructura no plana no comparable con una superficie como por ej., una microesfera o un pocillo. Los oligonucleótidos o polinucleótidos de la matriz se pueden unir covalentemente al soporte sólido, o se pueden unir de manera no covalente. La tecnología de micromatrices convencional es revisada en, por ej., Schena, Ed., Microarrays: A Practical Approach, IRL Press, Oxford (2000). El «análisis de matriz», «análisis por matriz» o «análisis por micromatriz» se refiere al análisis, como por ej., análisis de secuencias, de una o más moléculas biológicas mediante el uso de una micromatriz. En algunas realizaciones, cada muestra se hibrida individualmente a una micromatriz única. En otras realizaciones, el rendimiento del procesamiento puede ser mejorado mediante las micromatrices múltiples que se conectan físicamente sobre una única placa de micromatriz múltiple para el manejo de alto rendimiento conveniente. En ciertas realizaciones, las micromatrices de ADN habituales, por ejemplo de Affymetri Inc. (Santa Clara, Calif., EE. UU.), se pueden producir para cuantificar específicamente los productos del ensayo de RO-MIP.

Un experto en la técnica entenderá que las composiciones y los métodos que se describen en la presente memoria se pueden adaptar y modificar.

Esta divulgación se entenderá mejor a partir de los detalles experimentales que siguen a continuación. No obstante, un experto en la técnica apreciará que los métodos y los resultados específicos que se discuten son meramente ilustrativos de la divulgación ya que esta se describe por completo en las realizaciones que siguen de aquí en adelante.

Métodos para detectar enfermedades o afecciones

Los métodos de secuenciación existentes emplean etapas de preparación de la biblioteca de secuenciación detalladas, requieren decenas de millones de lecturas para obtener coeficientes útiles de variación, y pueden perder la validez en la fracción fetal en la muestra que cae por debajo del 4 %. Los métodos de «bombardeo con proyectiles» no focalizados requieren grandes números de lecturas para obtener la cobertura de las regiones deseadas en los cromosomas pertinentes a la aneuploidía humana. Los métodos específicos requieren la manipulación de un gran número de cebadores de PCR y multiplexación. Los métodos que utilizan un par único de cebadores en la amplificación por PCR de las regiones de repetición en la preparación de la biblioteca pueden tener artefactos de PCR que producen ambigüedades (interferencia) en las secuencias del producto, lo que disminuye la proporción de las lecturas cartográficas exclusivas y la eficiencia general.

Las realizaciones de la presente divulgación proporcionan una solución a los problemas de los métodos de secuenciación existentes para detectar la aneuploidía. Estas realizaciones reemplazan las preparaciones de la biblioteca anteriores con un método de captura que utiliza un pequeño número de MIP de oligonucleótidos que comprenden los brazos de polinucleótidos de focalización que hibridan a las secuencias de repetición, dichos brazos son los brazos que se unen a las estructuras del esqueleto universal de alto rendimiento. Estas MIP están diseñadas para flanquear e incorporar las secuencias que se alinean únicamente sobre el genoma humano entero, pero están enriquecidas por los blancos pertinentes para la detección de las aneuploidías comunes (por ej., trisomía del cromosoma 21, 18 o 13). Los métodos que se contemplan para la selección de moléculas de captura tratan la necesidad de seleccionar secuencias únicas en un área conveniente para la cuantificación, y no dependen de la presencia de algunas secuencias únicas en la amplificación de las secuencias de repetición convenientes.

El uso de secuencias de repetición (es decir, «causantes de repetición») en el método de captura optimizado permite

matrices de mosaico densas en un área específica con poca o ninguna interferencia de secuencias similares en la creación de blancos con código de barras para la cinética de la molécula única durante la preparación de la biblioteca. El análisis molecular único permite la cuantificación superior y el conteo de cromosomas. De manera alternativa se puede contar el número de lecturas. No obstante, el análisis molecular único es insesgado, y por lo tanto es menos probable que afecte la cuantificación. Al contar las etiquetas moleculares, los métodos que se describen en la presente memoria, obtienen una imagen más precisa de la abundancia relativa de cada secuencia en la muestra de ADN original. La presente divulgación proporciona, además, un método que tiene beneficios económicos sobre los métodos anteriores. En particular, los métodos proporcionan ahorros a través del uso de un pequeño número de reactivos de captura (cebadores) que todavía tienen la capacidad de medir los índices del genoma completo. Los métodos proporcionan, además, un rápido análisis con un conteo de lectura bajo en un ensayo que se multiplexa con facilidad. Por ejemplo, las capas múltiples de etiquetas moleculares únicas y/o los códigos de barra se pueden utilizar, en los métodos, para identificar las especies específicas del cebador, como así también, los datos de multiplexación desenmarañados a fin de trazar las señales de nuevo en las muestras individuales. Además, los métodos se pueden utilizar en aplicaciones de cobertura ultra baja como por ej., en la detección de trisomías en una muestra fetal al 100 %, como por ej., un producto de la concepción, o una muestra de diagnóstico no fetal. Una muestra se puede mezclar (por ej., fetal vs. maternal, o enfermo vs. no enfermo) o no mezclar (por ej., un niño que se sospecha que tiene una aneuploidía), en cuyo caso la «cobertura» o profundidad de lectura puede ser bastante baja (por ej., una profundidad de lectura inferior a 20.000) porque la señal será fuerte. Los métodos son, además, rápidos en comparación con la secuenciación del genoma entero, la secuenciación del exoma entero, y la secuenciación por bombardeo con microproyectiles masiva paralela.

Los métodos de la divulgación se refieren al campo del análisis genético. En general, estos métodos se pueden utilizar como un medio rápido y económico para detectar y cuantificar supresiones y duplicaciones de características genéticas en un intervalo que se extiende desde cromosomas completos y brazos de cromosomas hasta supresiones y duplicaciones microscópicas, supresiones sub-microscópicas y supresiones, e incluso características nucleotídicas únicas que incluyen polimorfismos de nucleótidos únicos, supresiones e inserciones. En ciertas realizaciones, los métodos de la divulgación se pueden utilizar para detectar lesiones genéticas sub-cromosómicas, por ej., micro-supresiones. Las aplicaciones ilustrativas de los métodos incluyen el diagnóstico pediátrico de aneuploidía, la prueba para el producto de la concepción o el riesgo de aborto prematuro, la prueba prenatal no invasiva (prueba genética tanto cualitativa como cuantitativa, como por ej., la detección de los trastornos de Mendelian, las inserciones/supresiones y los desequilibrios cromosómicos), la prueba genética preimplantacional, la caracterización del tumor, la prueba posnatal que incluye la citogenética y el monitoreo del efecto mutagénico.

Las moléculas de ácido nucleico (por ej., las MIP) que se proporcionan mediante la divulgación también tienen el beneficio de incrementar la estabilidad de unión en comparación con los pares de cebadores de PCR que no son parte de la misma molécula. En ciertas realizaciones, las secuencias exactas del brazo de focalización son, de algún modo, cortas para los cebadores de PCR, y por lo tanto, tendrán muy bajas temperaturas de fusión en un contexto de PCR. Sin embargo, en una configuración de MIP, los cebadores mejorarán la especificidad de unión mediante su colaboración para estabilizar la interacción. Si un brazo tiene una alta eficiencia de unión, la captura es mejorada incluso si el brazo opuesto tiene una eficiencia inferior. La extensión aditiva del par mejora el equilibrio de «activar/desactivar» para la captura porque el brazo de eficiencia inferior está con mayor frecuencia en proximidad con su blanco en una MIP que lo que estaría como cebador de PCR libre.

Los métodos que se proporcionan mediante la divulgación tienen varias ventajas en comparación con la secuenciación focalizada. En ciertas realizaciones, los métodos que se describen en la presente memoria utilizan un reconocimiento simultáneo de dos elementos de secuencia en el punto de captura, y los dos brazos están limitados por la proximidad. Por el contrario, un método de secuenciación focalizado típico permitirá que una polimerasa se inicie en un sitio único. El producto mal formado que se creó mediante la secuenciación típica produce no solo ineficacia, sino además un cebado interno o un «cebado de desviación» con el segundo cebador. El «reconocimiento dual» inherente de los ácidos nucleicos de la divulgación (por ej., RO-MIP) aumenta la restricción, un efecto que acarrea la cuantificación por parte del elemento identificador molecular en la estructura de la MIP. Una etiqueta molecular única se puede colocar en un sitio en el esqueleto de la MIP, pero en la secuenciación focalizada estándar mediante el uso de un identificador molecular, se utiliza una secuencia aleatoria en ambos cebadores. Además, los métodos que proporciona la divulgación permiten costos de reactivos más bajos, dado que la cobertura del genoma completo puede obtenerse con pocas RO-MIP en comparación con los cientos o miles de cebadores de PCR multiplexados necesarios para la secuenciación focalizada.

No obstante, los métodos de la divulgación cuentan, en su mayoría, si no todos, con ventajas económicas y de rendimiento que la secuencia focalizada muestra sobre los métodos de bombardeo con proyectiles.

Los métodos y los ácidos nucleicos de la presente divulgación ofrecen claras ventajas sobre los métodos genéticos previamente descritos. Por ejemplo, la secuenciación del genoma entero y la secuenciación por bombardeo con proyectiles masiva paralela requieren, en general, análisis costosos de grandes partes no informativas del genoma, mientras que los métodos actuales pueden producir respuestas similares mediante el uso de una fracción del genoma, reduciendo, de este modo, los costos y el tiempo de ensayo. Otros enfoques confían en las partes informativas de ensayo selectivas del genoma. Mientras que ciertos aspectos de la presente divulgación comparten alguna similitud, los métodos actuales utilizan un enfoque novedoso y exhaustivo para identificar los sitios de unión al cebador, de repetición, que permiten parámetros de diseño del ensayo más grandes (secuencia agnóstica, por ejemplo, no limitada a elementos de la línea de repetición), más cebadores candidatos (por ej., porque todos los posibles cebadores se enumeran),

ensayos simples de menor costo que son específicos y lo suficientemente sensibles para utilidad clínica, y una mayor capacidad de multiplexación.

Los métodos y los ácidos nucleicos que se describen en la presente memoria, tienen claras ventajas sobre los métodos alternativos para identificar los sitios específicos de interés a lo largo del genoma que comprenden las regiones de repetición, por ejemplo, los métodos que utilizan cebadores para capturar sitios blanco de interés (o secuencias diana de interés) para detectar las aneuploidías cromosómicas. En ciertas realizaciones, los métodos de la divulgación utilizan las MIP para capturar los sitios blanco de interés (o las secuencias diana de interés).

En ciertas realizaciones, los replicones de la MIP (o los amplicones) que se generan en los métodos que se describen en la presente memoria tienen un tamaño de entre 50 y 120 pb (por ej., 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110 o 115 pb, o cualquier tamaño entre 50 y 120 pb, o cualquier intervalo de tamaño entre 50 y 120 pb). En algunas realizaciones, los replicones de la MIP (o los amplicones) tienen un tamaño de entre 80 y 90 pb, o entre 80 y 100 pb, o entre 80 y 110 pb, o entre 80 y 120 pb, o entre 70 y 90 pb, o entre 70 y 100 pb, o entre 70 y 110 pb o entre 70 y 120 pb. En algunas realizaciones, los replicones de la MIP (o los amplicones) tienen un tamaño de entre 80 y 90 pb. Los métodos de captura del cebador generan replicones (o amplicones) que son más largos que los replicones (o los amplicones) de la MIP que se generan en los métodos que se describen en la presente memoria. Con frecuencia el ADN circulante de las muestras de plasma está fragmentado. Cuando se utiliza dicho ADN como moldes, los replicones (o los amplicones) más cortos ofrecen claras ventajas sobre los más largos porque los replicones (o los amplicones) más cortos aumentan la probabilidad de capturar fragmentos cortos. Si un amplicón es largo, es menos probable que los fragmentos cortos tengan ambos sitios de unión de dicho amplicón largo. Además, la profundidad de lectura por muestra en los métodos de captura del cebador conocidos es más alta que la de los métodos que se describen en la presente. Esta es una desventaja de los métodos de captura del cebador conocidos. En ciertas realizaciones, los métodos que se describen en la presente memoria proporcionan una profundidad de lectura inferior a 20 lecturas por muestra, o inferior a 19 millones de lecturas por muestra, o inferior a 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4 o 3 millones de lecturas por muestra, pero no inferior a 2 millones de lecturas por muestra, o cualquier intervalo entre 2 y 20 millones de lecturas por muestra, o cualquier intervalo entre 3 y 20 millones de lecturas por muestra. En algunas realizaciones, los métodos que se describen en la presente memoria proporcionan una profundidad de lectura de entre 6 y 8 millones de lecturas por muestra, por ej., 6, 7 u 8 millones de lecturas por muestra. Más aún, cuando se compara con los métodos de captura del cebador, los métodos que se describen en la presente memoria apuntan a más sitios de interés (o secuencias de interés) del genoma completo y/o sobre el cromosoma de interés (por ej., cromosoma 221) que los métodos de captura del cebador. En ciertas realizaciones, los métodos que se describen en la presente memoria, tienen un número total de sitios de unión a lo largo del genoma en un intervalo de 50k a 250k (o cualquier número o intervalo entre 50k y 250k). En algunas realizaciones, el número total de sitios de unión a lo largo del genoma es mayor que 50k, 60k, 70k, 80k, 90k, 100k, 110k, 120k, 130k, 140k, 150k, 160k, 170k, 180k, 190k, 200k, 210k, 220k, 230k o 240k. En algunas realizaciones, el número total de sitios de unión a lo largo del genoma es entre 125k-175k. En ciertas realizaciones, los métodos que se describen en la presente memoria, tienen un número total de sitios de unión sobre un cromosoma de interés (por ej., cromosoma 21) en un intervalo de 500 a 3000 sitios (o cualquier número o intervalo entre 500 y 3000 sitios). En algunas realizaciones, el número total de sitios de unión sobre un cromosoma de interés es mayor que 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900 o 3000 sitios. En algunas realizaciones, los índices de alineación únicos son mayores que 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 % o 50 % o más. Como se utiliza en la presente, la frase «índice de alineación único» se refiere al porcentaje de las lecturas de secuenciación total que se alinean únicamente a una ubicación cromosómica sobre el genoma de un sujeto (por ej., un genoma humano).

En ciertas realizaciones, los métodos que se describen en la presente memoria, utilizan pares de cebadores que no son MIP para capturar, o unirse a, los sitios blanco de interés (o la secuencias diana de interés). En algunas realizaciones, los pares de cebadores no-MIP están dispuestos de manera lineal o circular. Como se utiliza en la presente memoria, las frases «secuencia diana de interés» y «sitio blanco de interés» se utilizan intercambiabilmente para referirse a una parte de la molécula de ácido nucleico genómico para cuya captura o unión están diseñados los pares del cebador. En algunas realizaciones, uno o más pares de cebadores están diseñados para hibridar de inmediato corriente arriba y corriente abajo de una secuencia diana específica (o sitio) de interés en una muestra de ácido nucleico genómico. En algunas realizaciones, uno o más pares de cebadores comprenden secuencias que son complementarias a una o más secuencias de interés (o sitios de interés) en una muestra de ácido nucleico genómico.

En algunas realizaciones, la divulgación proporciona un método para detectar aneuploidía, o la ausencia de aneuploidía, en un individuo o feto que lo necesita. En algunas realizaciones, la divulgación proporciona un método para detectar aneuploidía, o la ausencia de aneuploidía, en un individuo o feto que lo necesita. En algunas realizaciones, la divulgación proporciona un método para detectar aneuploidía en un feto que comprende:

- a) obtener una muestra de ácido nucleico que se aísla a partir de una muestra de sangre materna;
- b) capturar una pluralidad de secuencias diana de interés en la muestra de ácido nucleico que se obtiene en la etapa a) mediante el uso de una o más poblaciones de sondas de inversión molecular (MIP) para producir una pluralidad de replicones, en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes:

un primer brazo de polinucleótidos de focalización y un segundo brazo de polinucleótidos de focalización; en

donde el par del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en la primera pluralidad de secuencias diana de interés;

- 5 c) amplificar la secuencias diana de interés;
- d) secuenciar la secuencias diana de interés;
- e) aparear las secuencias diana de interés *in silico* a las secuencias genómicas en el locus genómico; y
- f) contar el número de amplicones de apareamiento en el locus genómico individual; comparar el número de amplicones apareados al locus genómico sobre un cromosoma de prueba con el número de amplicones apareados al locus genómico sobre los cromosomas de referencia.
- 10 En algunas realizaciones, la divulgación proporciona un método para detectar aneuploidía en un feto que comprende:
- a) obtener una muestra de ácido nucleico que se aísla a partir de una muestra de sangre materna;
- b) capturar una pluralidad de secuencias diana de interés en la muestra de ácido nucleico que se obtiene en la etapa a) mediante el uso de una o más poblaciones de sondas de inversión molecular (MIP) para producir una pluralidad de replicones, en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes: primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización; en donde el par del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en la pluralidad de secuencias diana de interés; en donde la primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;
- 15 c) secuenciar una pluralidad de amplicones de MIP que se amplían a partir de los replicones que se obtienen en la etapa b);
- d) determinar el número de eventos de captura de cada una de una primera población de amplicones de la pluralidad de amplicones que se proporciona en la etapa c), basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la primera población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- 20 e) determinar el número de eventos de captura de cada una de una segunda población de amplicones de la pluralidad de amplicones que se proporciona en la etapa c), basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la segunda población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- 25 f) determinar, para cada secuencia diana de interés a partir de la cual se produjo la primera población de amplicones, una métrica de captura del sitio basado, al menos, en parte en el número de las lecturas de secuenciación que se determinaron en la etapa d);
- 30 g) identificar un primer subconjunto de la métrica de captura del sitio que se determinó en la etapa f) que satisface, al menos, un criterio;
- h) determinar, para cada secuencia diana de interés a partir de la cual se produjo la segunda población de amplicones, una métrica de captura del sitio basado, al menos, en parte en el número de eventos de captura que se determinaron en la etapa e);
- 35 i) identificar un segundo subconjunto de la métrica de captura del sitio que se determinó en la etapa h) que satisface, al menos, un criterio;
- j) normalizar una primera medida que se determina a partir del primer subconjunto de métricas de captura del sitio que se identifica en la etapa g) mediante una segunda medida que se determina a partir del segundo subconjunto de métricas de captura del sitio que se identifica en la etapa i) para obtener una relación de prueba;
- 40 k) comparar la relación de prueba con una pluralidad de relaciones de referencia que se computan basado en las muestras de ácido nucleico de referencia que se aíslan de los sujetos de referencia que se sabe exhiben euploidía o aneuploidía, y
- 45 l) determinar, basado en la comparación de la etapa k), si se detecta aneuploidía en el feto. De manera alternativa, este método se puede utilizar para detectar aneuploidía en un sujeto no fetal. En ciertas realizaciones, como una alternativa para detectar aneuploidía, los métodos de la divulgación se pueden utilizar para detectar y cuantificar supresiones y duplicaciones de características genéticas en los brazos de cromosomas, como así también supresiones y duplicaciones microscópicas, supresiones sub-microscópicas y supresiones, y características nucleotídicas únicas que incluyen polimorfismos nucleotídicos únicos, supresiones e inserciones.
- 50
- 55

- En ciertas realizaciones, los métodos de la divulgación se pueden realizar sobre una muestra de ácido nucleico como por ej., ADN o ARN, por ej., ADN genómico. Una muestra de ácido nucleico se puede aislar de cualquier manera conocida por parte de un experto en la técnica (por ej., por centrifugación). El experto apreciará que el sujeto puede ser cualquier ser humano. Cuando se detecta euploidía, aneuploidía o una enfermedad o afección en un feto, el sujeto es una mujer embarazada.
- En algunas realizaciones, los métodos de la divulgación utilizan una única especie de MIP. En realizaciones alternativas, los métodos son útiles con 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 o más especies de MIP. Por ejemplo, se pueden utilizar múltiples especies de MIP para detectar diferentes enfermedades o afecciones (por ej., anomalías cromosómicas como por ej., aneuploidía) en una muestra simple. En ciertas realizaciones, una única MIP se puede utilizar para detectar diferentes enfermedades o afecciones (por ej., anomalías cromosómicas como por ej., aneuploidía) en una muestra simple.
- El experto en la técnica apreciará que las extensiones del primero y del segundo brazo de polinucleótidos de focalización pueden variar, según sea apropiado, para proporcionar hibridación eficiente entre el polinucleótido de focalización y la muestra de ácido nucleico. Por ejemplo, el primero y/o el segundo brazo de polinucleótidos de focalización pueden estar entre 14 y 30 pares base, por ej., 18-21 pares base. En ciertas realizaciones, la extensión del primero y/o del segundo brazo de polinucleótidos de focalización es de 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 pares base, o cualquier intervalo entre 14 y 30 pares base. En ciertas realizaciones, los brazos de polinucleótidos de focalización tienen una temperatura de fusión (TM) de entre 45 °C y 80 °C (por ej., 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C, 51 °C, 52 °C, 53 °C, 54 °C, 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, 60 °C, 61 °C, 62 °C, 63 °C, 64 °C, 65 °C, 66 °C, 67 °C, 68 °C, 69 °C, 70 °C, 71 °C, 72 °C, 73 °C, 74 °C, 75 °C, 76 °C, 77 °C, 78 °C, 79 °C o 80 °C, o cualquier intervalo de entre 45 °C y 80 °C) y/o un contenido de GC entre 30 % y 80 % (por ej., aproximadamente 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 % o 80 %, o cualquier intervalo entre 30 % y 80 %). En ciertas realizaciones, los brazos de polinucleótidos de focalización tienen una temperatura de fusión (TM) de entre 45 °C y 80 °C (por ej., 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C, 51 °C, 52 °C, 53 °C, 54 °C, 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, 60 °C, 61 °C, 62 °C, 63 °C, 64 °C, 65 °C, 66 °C, 67 °C, 68 °C, 69 °C, 70 °C, 71 °C, 72 °C, 73 °C, 74 °C, 75 °C, 76 °C, 77 °C, 78 °C, 79 °C o 80 °C, o cualquier intervalo entre 45 °C y 80 °C) y/o un contenido de GC entre 30 % y 70 % (por ej., aproximadamente 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 % o 70 %, o cualquier intervalo de entre 30 % y 70 %). En ciertas realizaciones, los brazos de polinucleótidos de focalización tienen una TM de entre 60 °C y 70 °C y/o un contenido de GC de entre 30 % y 70 %. En ciertas realizaciones, los brazos de polinucleótidos de focalización tienen, al menos, uno o más de lo siguiente: 1) una extensión de 14-30 nucleótidos; 2) una TM de entre 45 °C y 80 °C y 3) un contenido de GC de entre 30 % y 70 %. En ciertas realizaciones, los brazos de polinucleótidos de focalización tienen la misma secuencia esqueleto (es decir, el mismo ligador polinucleotídico) para la amplificación pos-captura. En algunas realizaciones, la secuencia del primer brazo de polinucleótidos de focalización es CACTGCACTCCAGCCTGG. En algunas realizaciones, la secuencia del segundo brazo de polinucleótidos de focalización es GAGGCTGAGGCAGGAGAA. En algunas realizaciones, los brazos de polinucleótidos de focalización, por ejemplo, tienen más que 1.000, más que 10.000, más que 20.000, más que 30.000, más que 40.000, más que 50.000, más que 60.000, más que 70.000, más que 80.000, más que 90.000, más que 100.000, más que 200.000, más que 300.000, más que 400.000, más que 500.000, más que 600.000, más que 700.000, más que 800.000, más que 900.000, y/o más que 1.000.000 secuencias de interés (o sitios de interés). En algunas realizaciones, las secuencias diana de interés (o sitios de interés) tienen un tamaño de 150 pb (como por ej., 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150 pb, o cualquier intervalo de entre 50-150 pb). En algunas realizaciones, una RO-MIP no se une a los elementos de nucleótidos intercalados largos (LÍNEA) en el genoma.
- En ciertas realizaciones, las MIP que se describen en la presente memoria capturan o se unen a una pluralidad de elementos Alu en el genoma. Los elementos Alu son los elementos transponibles más abundantes en un sujeto humano, que tiene más de un millón de copias dispersas a lo largo del genoma. Los elementos Alu son secuencias repetitivas y tienen una extensión de aproximadamente 300 pares base. Ver la figura 9. En algunas realizaciones, las MIP capturan o se unen al brazo derecho de los elementos Alu. En algunas realizaciones, las MIP capturan o se unen al brazo izquierdo de los elementos Alu. En algunas realizaciones, las MIP capturan o se unen a la región de inserción 31-nt sobre el brazo derecho de los elementos Alu (ver la figura 9).
- Las etiquetas moleculares únicas proporcionan una forma de determinar el número de eventos de captura para un amplicón dado. Una MIP puede comprender una o más etiquetas moleculares únicas, por ej., 1, 2, 3, 4 o 5 etiquetas moleculares únicas. En ciertas realizaciones, la extensión de la primera y/o de la segunda etiqueta molecular única está entre 4 y 15 pares base, por ej., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 pares base. En ciertas realizaciones, cada una de las etiquetas moleculares únicas tienen una temperatura de fusión de entre 45 °C y 80 °C (por ej., 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C, 51 °C, 52 °C, 53 °C, 54 °C, 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, 60 °C, 61 °C, 62 °C, 63 °C, 64 °C o 65 °C) y/o un contenido de GC de entre 30 % y 80% (por ej., aproximadamente 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 % u 80 %, o cualquier intervalo entre 30 % y 80 %, como por ej., 30 % a 70 %).
- Un ligador polinucleotídico puentea el hueco entre los dos brazos de polinucleótidos de focalización. En algunas realizaciones, el ligador polinucleotídico se ubica directamente entre la primera y la segunda etiqueta molecular única. En ciertas realizaciones, el ligador polinucleotídico no es sustancialmente complementario a alguna región genómica del sujeto. En ciertas realizaciones, el ligador polinucleotídico tiene una extensión de entre 20 y 1.000 pares base (por ej., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40 pares base.) y/o una temperatura de

fusión de entre 45 ° y 80 °C (por ej., 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C, 51 °C, 52 °C, 53 °C, 54 °C, 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, 60 °C, 61 °C, 62 °C, 63 °C, 64 °C o 65 °C) y/o un contenido de GC de entre 30 % y 80 % (por ej., aproximadamente 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 % u 80 %, o cualquier intervalo entre 30 % y 80 %, como por ej., 30-70 %). En ciertas realizaciones, el ligador polinucleotídico comprende, al menos, un cebador de amplificación, por ej., un cebador de amplificación hacia adelante y un cebador de amplificación inverso. Por ejemplo, la secuencia del cebador de amplificación hacia adelante puede comprender la secuencia de nucleótidos de 5'-CTTCAGCTTCCCGATTACGG-3' (SEC ID NO: 1) y/o la secuencia del cebador de amplificación inverso puede comprender la secuencia de nucleótidos de GCACGATCCGACGGTAGTGT-3' (SEC ID NO: 2). De este modo, la secuencia del ligador polinucleotídico puede comprender la secuencia de nucleótidos de 5'-CTTCAGCTTCCCGATTACGGGCAC-GATCCGACGGTAGTGT-3' (SEC ID NO: 3).

En ciertas realizaciones, la MIP comprende la secuencia de nucleótidos de 5' CACTGCACTCCAGCCT-GG(N1-6)CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT(N7-12)GAGGCTGAGGCAGGAGAA-3' (SEC ID NO: 6) en donde (N1-6) representa la primera etiqueta molecular única y (N7-12) representa la segunda etiqueta molecular única.

En ciertas realizaciones, la divulgación en la presente memoria proporciona moléculas de ácido nucleico que comprenden una secuencia de nucleótidos de 5'- A - (N)x - B - (N)y - C -3', en donde (N)x representa una primera etiqueta molecular única y (N)y representa una segunda etiqueta molecular única, y en donde X e Y están entre 4 y 5 pares base, en donde A i) comprende la secuencia de 5'-TGCACTCCAGCCTG-3' (SEC ID NO: 15), o una secuencia que es al menos 85 % similar a la secuencia de 5'-TGCACTCCAGCCTG-3' (SEC ID NO: 15); y ii) tiene una extensión de no más de 30 pares base, en donde C i) comprende la secuencia de 5'-GAGGCTGAGGCAGGA-3' (SEC ID NO: 16), o una secuencia que es al menos 85 % similar a la secuencia de 5'-GAGGCTGAGGCAGGA-3' (SEC ID NO: 16); y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, A i) comprende una secuencia que es al menos 90 % o 95 % similar a la secuencia de 5'-TGCACTCCAGCCTG-3' (SEC ID NO: 15); y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, C i) comprende una secuencia que es al menos 90 % o 95 % similar a la secuencia de 5'-GAGGCTGAGGCAGGA-3' (SEC ID NO: 16); y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, B i) comprende la secuencia de 5'-CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT-3' (SEC ID NO: 3); o una secuencia que es al menos 85 % (o 90 % o 95 %) similar a la secuencia de nucleótidos de 5'-CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT-3' (SEC ID NO: 3). En algunas realizaciones, A o C tiene una temperatura de fusión de entre 45 °C y 80 °C. En algunas realizaciones, A o C tienen un contenido de GC de entre 30 % y 80 %, o de entre 30 % y 70 %.

En ciertas realizaciones, la divulgación en la presente memoria proporciona moléculas de ácido nucleico que comprenden una secuencia de nucleótidos de 5'- A - (N)x - B - (N)y - C -3', en donde (N)x representa una primera marca molecular única y (N)y representa una segunda marca molecular única, y en donde X e Y están entre 4 y 15 pares base, en donde A i) comprende la secuencia de 5'-TCCTGCCTCAGCCTC-3' (SEC ID NO: 17), o una secuencia que es al menos 85 % similar a la secuencia de 5'-TCCTGCCTCAGCCTC-3' (SEC ID NO: 17); y ii) tiene una extensión de no más de 30 pares base, y en donde C i) comprende la secuencia de 5'-AGGCTGGAGTGC-3' (SEC ID NO: 18), o una secuencia que es al menos 85 % similar a la secuencia de 5'-AGGCTGGAGTGC-3'(SEC ID NO: 18); y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, A i) comprende una secuencia que es al menos 90 % o 95 % similar a la secuencia de 5'-TCCTGCCTCAGCCTC- 3' (SEC ID NO: 17), y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, C i) comprende una secuencia que es al menos 90 % o 95 % similar a la secuencia de 5'-AGGCTGGAGTGC-3' (SEC ID NO: 18), y ii) tiene una extensión de no más de 30 pares base. En algunas realizaciones, B comprende la secuencia de 5'-CTTCAGCTTCCCGATTACG-GGCACGATCCGACGGTAGTGT-3' (SEC ID NO. 3), o una secuencia que es al menos 85 % (o 90 % o 95 %) similar a la secuencia de 5'-CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT-3' (SEC ID NO: 3). En algunas realizaciones, A o C tiene una temperatura de fusión de entre 45 °C y 80 °C. En algunas realizaciones, A o C tienen un contenido de GC de entre 30 % y 80 %, o de entre 30 % y 70 %.

En algunas realizaciones, las MIP que se utilizan en los métodos que se describen en la presente memoria son las siguientes, donde los correspondientes valores para A, B, C, D, E, F, G y H son como se describen con relación a las tablas 1 y 2, y los valores correspondientes para la puntuación son como se describen con relación a la EQ. 9:

ES 2 745 556 T3

ID	SECUENCIA								PUNTUACIÓN
	A	B	C	D	E	F	G	H	
MIP001	/5Phos/CCACTGCACTCCAGCCTGNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAGAA (SEQ ID NO: 7)								8,151
	757	439	1329	576	63973	38697	120953	50631	
MIP002	/5Phos/TCTCCTGCCTCAGCCTCCNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGGAGTGCAGTGGC (SEQ ID NO: 8)								5,868
	539	341	1092	559	44384	30852	100409	48075	
MIP003	/5Phos/CACTGCACTCCAGCCTGGNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAGAA (SEQ ID NO: 9)								8,557
	784	424	1326	556	65221	37461	122258	48707	
MIP004	/5Phos/CCACTGCACTCCAGCCTGNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAGA (SEQ ID NO: 10)								8,167
	755	420	1277	546	62323	36938	119507	48684	
MIP005	/5Phos/CACTGCACTCCAGCCTGGNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNCAGGAGGCTGAGGCAGGA (SEQ ID NO: 11)								4,301
	272	71	781	282	20949	6673	69142	23425	
MIP006	/5Phos/ACTGCACTCCAGCCTGGNNNNNNCTTCAGCTTCCCGATTACGGGCACG ATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAG (SEQ ID NO: 12)								9,761
	865	407	1393	513	70197	36142	131805	46075	
MIP007	/5Phos/TGCACTCCAGCCTGGGCANNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAGAA (SEQ ID NO: 13)								5,283
	370	80	902	315	32498	8732	76997	26567	
MIP008	/5Phos/CTGCACTCCAGCCTGGGCNNNNNNCTTCAGCTTCCCGATTACGGGCAC GATCCGACGGTAGTGTNNNNNNAGGCTGAGGCAGGAGAA (SEQ ID NO: 14)								5,965
	475	261	1078	386	37864	22133	96980	34718	

En algunas realizaciones, la población de las MIP que se utiliza en un método de la divulgación tiene una concentración entre 10 fM y 100 nM, por ejemplo 0,5 nM. En ciertas realizaciones, la concentración de las MIP que se utiliza en un método de la divulgación variará con el número de secuencias que se focaliza, por ej., según se calcula multiplicando el número de secuencias diana de interés por el número de equivalentes genómicos en una reacción (el «número específico total»). En realizaciones particulares, la relación aproximada del número de moléculas MIP y el número específico total es 1:50, 1:100, 1:150, 1:200, 1:250, 1:300, 1:350, 1:400, 1:450, 1:500, 1:550, 1:600, 1:650, 1:700, 1:750, 1:800, 1:850, 1:900, 1:950, o 1:1,000. En ciertas realizaciones, cada uno de los replicones y/o amplicones de las MIP es una molécula de ácido nucleico circular monocatenaria.

En algunas realizaciones, los replicones de las MIP son producidos por: i) el primero y el segundo brazo de polinucleótidos de focalización, respectivamente, que hibridan a la primera y a la segunda región en la muestra de ácido nucleico, respectivamente, en donde la primera y la segunda región flanquean una secuencia diana de interés; y ii) después de la hibridación, el uso de una mezcla de ligación/extensión para extender y ligar la región con huecos entre los dos brazos de polinucleótidos de focalización para formar moléculas de ácido nucleico circulares monocatenarias. En ciertas realizaciones, un amplicón de MIP es producido por la amplificación de un replicón de MIP, por ej., a través de PCR.

En algunas realizaciones, la etapa de secuenciación comprende un método de secuenciación de próxima generación, por ejemplo, un método de secuenciación paralelo masivo, o un método de secuenciación de corta lectura. En algunas realizaciones, la secuenciación puede ser mediante cualquier método conocido en la técnica, por ejemplo, la secuenciación focalizada, la secuenciación en tiempo real de única molécula, la secuenciación basada en la microscopía de electrones, la secuenciación mediada por transistor, la secuenciación directa, la secuenciación por bombardeo con microproyectiles aleatoria, la secuenciación Sanger por terminación didesoxi, la secuenciación focalizada, la secuenciación exón, la secuenciación de genoma entero, la secuenciación por hibridación, la pirosecuenciación, la electroforesis capilar, la electroforesis en gel, la secuenciación dúplex, la secuencia del ciclo, la secuenciación de la extensión de base única, la secuenciación de fase sólida, la secuenciación de alto rendimiento, la secuenciación paralela masiva de la firma, la PCR en emulsión, la coamplificación a baja temperatura de desnaturalización de PCR (COLD-PCR), la PCR múltiplex, la secuenciación por terminador de tinte reversible, la secuenciación de extremo apareado, la secuencia a término cercano, la secuenciación de exonucleasa, la secuenciación por ligación, la secuenciación de lectura corta, la secuencia de única molécula, la secuenciación por síntesis, la secuenciación en tiempo real, la secuenciación por terminador inverso, la secuenciación de nanoporos, la secuenciación 454, la secuenciación por analizador del genoma de Solexa, la secuenciación SOLID®, la secuenciación MS-PET, la espectrometría de masa, y sus combinaciones. En algunas realizaciones, la secuenciación comprende la detección del producto de secuenciación mediante el uso de un instrumento, por ejemplo pero sin limitación, un secuenciador de ADN ABI PRISM® 377, un analizador genético ABI PRISM® 310, 3100, 3100-Avant, 3730 o 3730xl, un analizador de ADN ABI PRISM® 3700, o un sistema de Applied Biosystems SOLiD™ (todo de Applied Biosystems), una secuenciador del genoma 20 System (Roche Applied Science) o un espectrómetro de masa. En ciertas realizaciones, la secuenciación comprende la PCR en emulsión. En ciertas realizaciones, la secuenciación comprende una técnica de secuenciación de alto rendimiento, por ejemplo, pero sin estar limitada, la secuenciación paralela masiva de la firma (MPSS).

Una técnica de secuenciación que se puede utilizar en los métodos de la divulgación incluye, por ejemplo, la secuenciación Illumina. La secuenciación Illumina se basa en la amplificación del ADN sobre una superficie sólida mediante el uso de PCR plegada y cebadores anclados. El ADN genómico es fragmentado, y los adaptadores son adicionados a los extremos 5' y 3' de los fragmentos. Los fragmentos de ADN que se unen a la superficie de los canales de la célula de flujo se extienden y se produce la amplificación de puente. Los fragmentos se convierten en bicatenarios y las moléculas bicatenarias se desnaturalizan. Los múltiples ciclos de la amplificación de fase sólida seguido de la desnaturalización pueden crear varios millones de clústeres de aproximadamente 1.000 copias de moléculas de ADN monocatenarias del mismo molde en cada canal de la célula de flujo. Los cebadores, la ADN polimerasa y cuatro nucleótidos de terminación reversible, marcados con fluoróforo se utilizan para llevar a cabo la secuenciación secuencial. Después de la incorporación de los nucleótidos, se utiliza un láser para excitar los fluoróforos, y se captura una imagen y se registra la identidad de la primera base. Los terminadores de 3' y los fluoróforos de cada base incorporada se eliminan y se repiten las etapas de incorporación, detección e identificación. La secuenciación de acuerdo con esta tecnología se describe en la Patente estadounidense No. 7.960.120; la Patente estadounidense No. 7.835.871; la Patente estadounidense No. 7.232.656; la Patente estadounidense No. 7.598.035; la Patente estadounidense No. 6.911.345; la Patente estadounidense No. 6.833.246; la Patente estadounidense No. 6.828.100; la Patente estadounidense No. 6.306.597; la Patente estadounidense No. 6.210.891; la Publicación estadounidense 2011/0009278; la Publicación estadounidense 2007/0114362; la Publicación estadounidense 2006/0292611; y la Publicación estadounidense 2006/0024681.

En algunas realizaciones, un método de la divulgación comprende, antes de la secuenciación (por ej., la etapa de secuenciación de d) como se describió anteriormente), una reacción de PCR para amplificar los amplicones de las MIP para la secuenciación. Esta reacción de PCR puede ser una reacción de PCR de indexado. En ciertas realizaciones, la reacción de PCR de indexado introduce en cada uno de los amplicones de las MIP los siguientes componentes: un par de cebadores de indexado, un único código de barras de muestra y un par de adaptadores de secuenciación. En realizaciones particulares, los amplicones de las MIP de focalización con código de barra comprenden en secuencia los siguientes componentes: un primer adaptador de secuenciación, un primer cebador de secuenciación, la primera

etiqueta molecular de focalización única, el primer brazo de polinucleótidos de focalización, el ácido nucleico capturado, el segundo brazo de polinucleótidos de focalización, la segunda etiqueta única de focalización, un único código de barra de muestra, un segundo cebador de secuenciación, un segundo adaptador de secuenciación.

5 En algunas realizaciones, la primera pluralidad de secuencias diana de interés está sobre un único cromosoma. En algunas realizaciones, la segunda pluralidad de secuencias diana de interés está sobre múltiples cromosomas. Dado que la única secuencia de la MIP se puede utilizar para las secuencias diana de interés a lo largo de un genoma entero, en ciertas realizaciones los métodos de la divulgación proporcionan el beneficio de poder detectar la aneuploidía de más de un cromosoma a la vez. Por ejemplo, la primera pluralidad de secuencias diana se puede definir como las secuencias sobre el cromosoma 21, y la segunda pluralidad de secuencias diana se puede definir como las secuencias sobre los cromosomas restantes. No obstante, mediante el uso de la misma reacción, la primera pluralidad de secuencias diana se puede definir como las secuencias sobre el cromosoma 13, y la segunda pluralidad de secuencias diana se puede definir como las secuencias sobre los cromosomas restantes. De este modo, los datos de secuenciación de la misma reacción se pueden utilizar para detectar tanto el síndrome de Down (trisomía 21) como el síndrome de Patau (trisomía 13). De igual modo, las MIP se pueden diseñar, y se pueden analizar los datos, para detectar 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más afecciones vinculadas con la aneuploidía, u otros tipos de anomalías cromosómicas o sub-cromosómicas.

En algunas realizaciones, la divulgación proporciona un método para detectar aneuploidía en un feto que comprende:

- a) obtener una muestra de ADN genómico a partir de una muestra de sangre materna;
- 20 b) agregar la muestra de ADN genómico a cada pocillo de una placa de múltiples pocillos, en donde cada pocillo de la placa de múltiples pocillos comprende una mezcla de sonda, en donde la mezcla de sonda comprende una población de sondas de inversión moleculares (MIP) y un amortiguador; en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes:
 primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización; en donde los pares del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en una pluralidad de secuencias diana de interés;
 en donde la primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;
- 25 c) incubar la muestra de ADN genómico con la mezcla de sonda para que las MIP capturen la pluralidad de secuencias diana de interés;
- d) agregar una mezcla de extensión/ligación a la muestra de c) para las MIP y la pluralidad de secuencias diana de interés para formar una pluralidad de amplicones de las MIP, en donde la mezcla de extensión/ligación comprende una polimerasa, una pluralidad de dNTP, una ligasa y un amortiguador;
- 35 e) agregar una mezcla de exonucleasa a los amplicones de las MIP de focalización y control para eliminar las sondas en exceso o el ADN genómico en exceso;
- f) agregar una mezcla de PCR de indexado a la muestra de e) para agregar un par de cebadores de indexado, un único código de barras de muestra y un par de adaptadores de secuenciación a la pluralidad de amplicones;
- 40 g) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una primera población de amplicones con código de barras que se proporcionan en la etapa f), basado en el número de las etiquetas moleculares de focalización únicas, en donde la primera población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés;
- 45 h) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una primera población de amplicones con código de barras que se proporcionan en la etapa f) basado en el número de etiquetas moleculares de focalización únicas, en donde la primera población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés.
- i) computar la métrica de captura del sitio basado, al menos, en parte en el número de las primeras lecturas de secuenciación que se determinaron en la etapa g) y una pluralidad de métricas de captura de sonda de control basado, al menos, en parte en los números de las segundas lecturas de secuenciación que se determinaron en la etapa h);
- 50 j) identificar un subconjunto de métricas de captura del sitio de la población de amplicones de las MIP que tienen métricas de captura de la sonda de control que satisfacen, al menos, un criterio;
- k) normalizar la métrica de captura del sitio mediante un factor que se computa a partir del subconjunto de métricas de captura de sonda de control que satisfacen el, al menos, un criterio, para obtener una métrica de captura del sitio normalizada de prueba;
- 55 l) comparar la métrica de captura del sitio normalizada de prueba con una pluralidad de métricas de captura del sitio normalizadas de referencia que se computan basado en las muestras de ADN genómico de referencia de

los sujetos de referencia que exhiben genotipos conocidos, mediante el uso de los mismos sitios específicos y de control, la población específica, el subconjunto de poblaciones de control en las etapas b)-h); y

m) determinar, basado en la comparación de la etapa 1) y los genotipos conocidos de los sujetos de referencia, si se detecta aneuploidía en el feto.

5 En algunas realizaciones, la divulgación proporciona un método de selección de una sonda de inversión molecular (MIP) a partir de una pluralidad de MIP candidatas para su uso en la detección de aneuploidía en un sujeto, el método comprende:

a) recibir secuencias de ácido nucleico de la pluralidad de MIP candidatas;

b) para cada MIP respectiva en la pluralidad de MIP candidatas,

10 i) computar un primer número (A) de sitios únicos previstos, sin apareamiento erróneo, a ser capturado por la MIP respectiva sobre un cromosoma de interés;

ii) computar un segundo número (C) de sitios únicos previstos, con un apareamiento erróneo, a ser capturado por la MIP respectiva sobre el cromosoma de interés;

15 iii) computar un tercer número (E) de sitios únicos previstos, sin apareamiento erróneo, a ser capturado por la MIP respectiva a través de un genoma;

iv) computar un cuarto número (G) de sitios únicos previstos, con un apareamiento erróneo, a ser capturado por la MIP a través del genoma;

v) computar un quinto número (F) de sitios no únicos previstos, sin apareamiento erróneo, a ser capturado por la MIP respectiva a través del genoma;

20 vi) computar un sexto número (H) de sitios no únicos previstos, con un apareamiento erróneo, a ser capturado por la MIP respectiva a través del genoma;

vii) computar una métrica de rendimiento para la MIP respectiva basado, al menos en parte, en el primero, segundo, tercero, cuarto, quinto y sexto número;

25 c) seleccionar una MIP basado, al menos en parte, en la métrica de rendimiento que se computa en la etapa b) vii) para cada MIP en la pluralidad de MIP candidatas.

En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una primera relación entre el primer número (A) y el quinto número (F) es mayor que una relación equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una segunda relación entre el primer número (A) y el tercer número (E) es mayor que una relación equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una tercera relación entre el primer número (A) y el segundo número (C) es mayor que una relación equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una cuarta relación entre una primera suma del primer número (A) y el segundo número (C) y una segunda suma del tercero, cuarto, quinto y sexto número (E, F, G, H) es mayor que una relación equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una quinta relación entre una primera suma ponderada del primer número (A) y el segundo número (C) y una segunda suma ponderada del tercero, cuarto, quinto y sexto número (E, F, G, H) es mayor que una relación equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la quinta relación (P1) entre la primera suma ponderada y la segunda suma ponderada es:

40

$$P1 = \frac{A + K_5 C}{(E + F) + K_6 (G + H)}$$

45 En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que una tercera suma ponderada entre el primer número (A) y el tercer número (C) es mayor que una suma ponderada equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la tercera suma ponderada es: $P2 = A + K_6 C$. En ciertas realizaciones, la MIP en la etapa c) se selecciona de modo que un producto entre la quinta relación (P1) y la tercera suma ponderada (P2) es mayor que un producto equivalente para un conjunto restante de las MIP candidatas. En ciertas realizaciones, la métrica de rendimiento se calcula basado en un número total de lecturas útiles del cromosoma de interés. En ciertas realizaciones, la MIP en la etapa c) se selecciona basado en una relación (Ke) de un coeficiente de captura promedio de un sitio de apareamiento erróneo (K1) y un coeficiente de captura promedio de cero sitios de apareamiento erróneo (K0):

50

$$K^e = \frac{K^0}{K^1}$$

En ciertas realizaciones, la relación (Ke) se estima de manera experimental. En ciertas realizaciones, la MIP en la etapa c) se selecciona basado en un conteo total de etiquetas moleculares(TMTC) que se define como:

$$TMTC = K_0(E + F) + K_1(G + H)$$

5

En algunas realizaciones, la divulgación proporciona, además, una molécula de ácido nucleico que comprende una secuencia de nucleótidos de 5'-CACTGCACTCCAGCCTGG (N1-6)CTTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT(N7-12)GAG-GCTGAGGCAGGAGAA-3' (SEC ID NO: 6) en donde (N1-6) representa la primera etiqueta molecular única y (N7-12) representa una segunda etiqueta molecular única. En ciertas realizaciones, la extensión de la primera etiqueta molecular única es entre 4 y 15 pares base. En ciertas realizaciones, la extensión de la segunda etiqueta molecular única es entre 4 y 15 pares base. En ciertas realizaciones, cada una de las etiquetas moleculares únicas tiene una temperatura de fusión de entre 45 °C y 80 °C. En ciertas realizaciones, cada una de las etiquetas moleculares únicas tiene un contenido de GC de entre 30 % y 80 %, o de entre 30 % y 70 %. La divulgación proporciona, además, una composición que comprende cualquiera de las moléculas de ácido nucleico que se describe en la presente memoria.

15

Métodos para identificar las MIP

La fig. 1 es un diagrama de bloque de un dispositivo de computación 100 para llevar a cabo cualquiera de los procedimientos que se describen en la presente memoria, que incluye los procedimientos 200, 300 y 500. Como se utiliza en la presente memoria, el término «procesador» o «dispositivo de computación» se refiere a una o más computadoras, microprocesadores, dispositivos lógicos, servidores y otros dispositivos que están configurados con hardware, firmware y software para llevar a cabo una o más de las técnicas computarizadas que se describen en la presente memoria. Los procesadores y los dispositivos de procesamiento también pueden incluir uno o más dispositivos de memoria para almacenar entradas, salidas y datos que se procesan actualmente. El dispositivo de computación 100 puede incluir una «interfaz de usuario», la cual puede incluir, sin limitación, cualquier combinación adecuada de uno o más dispositivos de entrada (por ej., teclados, pantallas táctiles, bolas de seguimiento, sistemas de reconocimiento de voz, etc.) y/o uno o más dispositivos de salida (por ej., presentaciones visuales, parlantes, visualizaciones táctiles, dispositivos de impresión, etc.). El dispositivo de computación 100 puede incluir, sin limitación, cualquier combinación adecuada de uno o más dispositivos que están configurados con hardware, firmware y software para llevar a cabo una o más de las técnicas computarizadas que se describen en la presente memoria. Cada uno de los componentes que se describe en la presente memoria se puede implementar sobre uno o más dispositivos de computación 100. En ciertos aspectos, una pluralidad de estos sistemas se puede incluir dentro de un dispositivo de computación 100. En ciertas realizaciones, un componente y un dispositivo de almacenamiento se pueden implementar a través de varios dispositivos de computación 100.

20

25

30

El dispositivo de computación 100 comprende, al menos, una unidad de interfaz de comunicaciones 108, un controlador de entrada/salida 110, una memoria del sistema y uno o más dispositivos de almacenamiento de datos. La memoria del sistema incluye, al menos, una memoria de acceso aleatorio (RAM 102) y al menos una memoria de solo lectura (ROM 104). Todos estos elementos están en comunicación con una unidad de procesamiento central (CPU 106) para facilitar el funcionamiento del dispositivo de computación 100. El dispositivo de computación 100 puede estar configurado en muchas formas diferentes. Por ejemplo, el dispositivo de computación 100 puede ser una computadora autónoma convencional, o de forma alternativa, las funciones del dispositivo de computación 100 se pueden distribuir a lo largo de múltiples sistemas y arquitecturas de computadora. En la figura 1, el dispositivo de computación 100 está vinculado, a través de la red o una red local, a otros servidores o sistemas.

35

40

El dispositivo de computación 100 puede estar configurado en una arquitectura distribuida, en donde las bases de datos y los procesadores se alojan en unidades o ubicaciones separadas. Algunas unidades realizan las funciones de procesamiento primario y contienen, como mínimo, un controlador general o un procesador y una memoria del sistema. En las realizaciones de arquitectura distribuidas, cada una de estas unidades se pueden unir, mediante la unidad de interfaz de comunicaciones 108 a un concentrador o puerto (que no se muestra) que funciona como un vínculo de comunicación primaria con otros servidores, otras computadoras del cliente o del usuario y otros dispositivos relacionados. El concentrador o puerto de comunicaciones puede tener una capacidad de procesamiento mínima, que funciona, esencialmente, como un enrutador de comunicaciones. Una variedad de protocolos de comunicaciones puede ser parte del sistema, que incluyen, pero sin estar limitados: Ethernet, SAP, SAS™, ATP, BLUETOOTH™, GSM y TC/IP.

45

50

La CPU 106 comprende un procesador, como por ej., uno o más microprocesadores convencionales y uno o más co-procesadores complementarios como por ej., co-procesadores de matemáticas para descargar la carga de trabajo de la CPU 106. La CPU 106 está en comunicación con la unidad de interfaz de comunicaciones 108 y el controlador de entrada/salida 110, a través del cual la CPU 106 se comunica con otros dispositivos como por ej., otros servidores, terminales de usuario o dispositivos. La unidad de interfaz de comunicaciones 108 y el controlador de entrada/salida 110 pueden incluir múltiples canales de comunicación para comunicación simultánea, por ejemplo, con otros procesadores, servidores o terminales de cliente.

55

60

La CPU 106 también está en comunicación con el dispositivo de almacenamiento de datos. El dispositivo de almacenamiento de datos puede comprender una combinación apropiada de memoria magnética, óptica o semiconductor, y puede incluir, por ejemplo, RAM 102, ROM 104, unidad flash, un disco óptico como por ej., un disco compacto o un disco duro. La CPU 106 y el dispositivo de almacenamiento de datos pueden estar ubicados, por ejemplo, completamente dentro de una computadora simple u otro dispositivo de computación; o conectado uno con otro mediante un medio de comunicación, como por ej., un puerto USB, un cable de puerto serie, un cable coaxial, un cable Ethernet, una línea telefónica, un transceptor de radiofrecuencia y otro medio inalámbrico o cableado similar, o una combinación de los anteriores. Por ejemplo, la CPU 106 puede ser conectada al dispositivo de almacenamiento de datos mediante la unidad de interfaz de comunicaciones 108. La CPU 106 puede ser configurada para realizar una o más funciones de procesamiento particular.

El dispositivo de almacenamiento de datos puede almacenar, por ejemplo, (i) un sistema operativo 112 para el dispositivo de computación 100; (ii) una o más aplicaciones 114 (por ej., un código de programa de computadora o un producto del programa de computadora) que se adapta para dirigir la CPU 106, de acuerdo con los sistemas y los métodos que se describen en esta memoria, y en particular, de acuerdo con los procedimientos que se describen en detalle con respecto a la CPU 106; o (iii) bases de datos 116 que se adaptan para almacenar información que se puede utilizar para almacenar información que requiere el programa.

El sistema operativo 112 y las aplicaciones 114 se pueden almacenar, por ejemplo, en un formato comprimido, no compilado o cifrado, y puede incluir un código de programa de computadora. Las instrucciones del programa se pueden leer en una memoria principal del procesador desde un medio legible por computadora además del dispositivo de almacenamiento de datos, como por ej., desde el ROM 104 o desde el RAM 102. Mientras la ejecución de las secuencias de instrucciones en el programa origina que la CPU 106 realice los pasos del procesamiento que se describen en la presente memoria, se puede utilizar un conjunto de circuitos cableados en lugar de, o en combinación con las instrucciones del software para la realización de los procedimientos de la presente divulgación. De este modo, los sistemas y los métodos descritos no se limitan a ninguna combinación específica de hardware y software.

El código de programa de computadora adecuado se puede proporcionar para la realización de una o más funciones como se describen en la presente memoria. Además, el programa puede incluir elementos del programa como por ej., un sistema operativo 112, un sistema de gestión de la base de datos y «unidades del dispositivo» que permiten que el procesador haga interfaz con los dispositivos periféricos de la computadora (por ej., una pantalla de vídeo, un teclado, un mouse de computadora, etc.) a través del controlador de entrada/salida 110.

El término «medio legible por computadora», como se utiliza en la presente memoria, se refiere a cualquier medio no transitorio que proporciona o participa en el suministro de instrucciones al procesador del dispositivo de computación 100 (o cualquier otro procesador de un dispositivo que se describe en la presente memoria) para ejecución. Dicho medio puede tomar muchas formas, que incluyen pero sin estar limitadas, medios no volátiles y medios volátiles. Los medios no volátiles incluyen, por ejemplo, discos ópticos, magnéticos o magneto-ópticos, o una memoria de circuito integrado, como por ej., una memoria flash. Los medios volátiles incluyen una memoria dinámica de acceso aleatorio (DRAM), la cual constituye, típicamente, la memoria principal. Las formas comunes de los medios legibles por computadora incluyen, por ejemplo, un disquete, un disco flexible, un disco duro, una cinta magnética, cualquier otro medio magnético, un CD-ROM, DVD, cualquier otro medio óptico, tarjetas perforadas, cinta de papel, cualquier otro medio físico con patrones de orificios, una memoria RAM, una PROM, una EPROM o EEPROM (memoria ROM que puede ser programada, borrada y reprogramada eléctricamente), una FLASH-EEPROM, cualquier otra memoria, chip o cartucho, o cualquier otro medio no transitorio desde el cual se puede leer una computadora.

Varias formas de medios legibles por computadora pueden estar implicados en el transporte de una o más secuencias de una o más instrucciones a la CPU 106 (o cualquier otro procesador de un dispositivo que se describe en la presente memoria) para ejecución. Por ejemplo, las instrucciones pueden ser cargadas, inicialmente, en un disco magnético de una computadora remota (que no se muestra). La computadora remota puede cargar las instrucciones en su memoria dinámica y enviar las instrucciones a través de una conexión Ethernet, una línea de cable, o incluso una línea telefónica mediante el uso de un módem. Un dispositivo de comunicaciones local para un dispositivo de computación 100 (por ej., un servidor) puede recibir los datos en la línea de comunicaciones respectiva y colocar los datos sobre un bus del sistema para el procesador. El bus del sistema transporta los datos a la memoria principal, desde la cual el procesador recupera y ejecuta las instrucciones. Las instrucciones que recibe la memoria principal se pueden almacenar, opcionalmente, en la memoria, o bien antes, o después de la ejecución por parte del procesador. Además, las instrucciones pueden ser recibidas mediante un puerto de comunicación como señales eléctricas, electromagnéticas u ópticas, las cuales son formas ilustrativas de comunicaciones inalámbricas o de corrientes de datos que transportan varios tipos de información.

La figura 2 es un diagrama de flujo de un procedimiento 200 para diseñar y seleccionar una sonda (por ej., una MIP), de acuerdo con una realización ilustrativa. El procedimiento 200 incluye las etapas de determinar un conjunto de restricciones (etapa 202), identificar los cebadores mediante el uso del conjunto de restricciones (etapa 204), realizar una técnica de optimización para maximizar el porcentaje de capturas moleculares sobre el cromosoma específico y maximizar el número de sitios útiles sobre el cromosoma específico (etapa 206), y seleccionar una sonda basado en la técnica de optimización (etapa 208).

En la etapa 202, se determina un conjunto de restricciones. El conjunto de restricciones se puede determinar, por ejemplo, mediante la CPU 106 que utiliza software o aplicaciones que están implementadas a partir de ahí. En algunas

realizaciones, el software o las aplicaciones se pueden utilizar, además, mediante la CPU 106 para realizar cualquiera de una o más de las etapas subsiguientes en el procedimiento 200. Por ejemplo, el software y las aplicaciones pueden ser utilizadas por la CPU 106 para encontrar abundantes pares de cebadores en un genoma de referencia dado (por ej., HG19) basado en las restricciones determinadas, y para crear, de forma automática, el índice basado en la matriz del sufijo para el archivo del genoma.

En algunas realizaciones, el conjunto de restricciones puede hacer referencia, de forma alternativa, a las indicaciones algorítmicas. Por ejemplo, las restricciones (o indicaciones algorítmicas) pueden incluir una extensión del cebador izquierdo, una frecuencia mínima del par de cebador, una distancia máxima entre los cebadores (por ej., extensión del amplicón), una frecuencia total mínima y/o máxima del cebador, un contenido de GC mínimo por cebador en por ciento, una cantidad mínima de amplicones no idénticos en por ciento, una distribución de cebadores en el genoma, o cualquiera de sus combinaciones adecuadas. En una realización ilustrativa, las siguientes restricciones se pueden utilizar en el diseño de los pares de cebadores:

- extensión del cebador izquierdo: 18, 19, 20, 21 pares base (pb).
- frecuencia del par de cebador: 100, 250, 500, 2500, 5000, 10000.
- extensión del amplicón: 50-150 pb, por ej., menos de 85 pb.
- contenido de GC mínimo por cebador: 40 %.
- unicidad del amplicón (por ciento de secuencias diana de interés que son únicas): superior al 80 %.
- distribución de los cebadores en el genoma: se ejecutó iterativamente, con cada tamaño de cubo (tc) que oscila desde 1 a 50 %, y el relleno del cubo (rc) que oscila desde 1 a tc-1, en donde el tamaño del cubo (tc) se refiere al % de tc de la extensión del genoma, y cada cubo debe contener un % de rc de todos los aciertos.

En la etapa 204, un conjunto de cebadores se identifican mediante el uso del conjunto de restricciones que se determinan en la etapa 202. En particular, para cada diseño de cebador, cualquier combinación de los siguientes parámetros se puede proporcionar: las secuencias del cebador izquierdo (por ej., como así también el número de sus ocurrencias sobre las hebras positivas y negativas del genoma), la secuencias del cebador derecho (por ej., como así también el número de sus ocurrencias sobre las hebras positivas y negativas del genoma), la frecuencia del par (por ej., la secuencia del cebador izquierdo y la secuencia del cebador derecho apareados juntos con la extensión del amplicón que está limitada por un restricción) que incluye tanto pares únicos como no únicos, la frecuencia y el porcentaje de los amplicones que se producen de manera única, y la secuencias del amplicón a partir de los pares únicos y no únicos. En algunas realizaciones, cada par de cebador puede amplificar múltiples regiones en el genoma (por ej., más de cien, más de miles, más de decenas de miles, más de cientos de miles, o más de millones).

En algunas realizaciones, los pares de cebadores generados pueden identificar o predecir los sitios del amplicón sin permitir que ningún apareamiento erróneo ocurra, o bien en la secuencia del cebador izquierdo, o en la secuencia del cebador derecho (es decir, los brazos izquierdo o derecho). De forma alternativa, a fin de identificar o predecir sitios adicionales del amplicón, se puede permitir un pequeño número de apareamientos erróneos, como por ejemplo permitir:

- 1 apareamiento erróneo en el brazo izquierdo y 0 apareamientos erróneos en el brazo derecho,
- 0 apareamientos erróneos en el brazo izquierdo y 1 apareamiento erróneo en el brazo derecho,
- 1 apareamiento erróneo en el brazo izquierdo y 1 apareamiento erróneo en el brazo derecho,
- 2 apareamientos erróneos en el brazo izquierdo y 0 apareamientos erróneos en el brazo derecho, o
- 0 apareamientos erróneos en el brazo izquierdo y 2 apareamientos erróneos en el brazo derecho.

En algunas realizaciones, el esquema de predicción del amplicón que se describió anteriormente proporciona las coordenadas genómicas de los amplicones previstos. No obstante, en algunas realizaciones, puede ser computacionalmente intensivo para el esquema, que identifica los sitios del amplicón sin permitir que se produzcan apareamientos erróneos, que proporcione, además, las coordenadas genómicas de los amplicones previstos. En este caso, el esquema se puede dividir en dos partes. En una primera parte, los sitios del amplicón se identifican sin permitir que se produzcan apareamientos erróneos, y las coordenadas genómicas de los sitios del amplicón identificadas no se proporcionan. En una segunda parte, se identifican los sitios del amplicón que incluyen un pequeño número de apareamientos erróneos (por ej., el conjunto de apareamientos erróneos que se enumeró anteriormente), y se proporcionan las coordenadas genómicas de estos sitios del amplicón, como así también la coordenada genómica de los sitios del amplicón sin apareamiento erróneo. La separación del esquema en estas dos partes modulares puede ahorrar complejidad computacional. No obstante, en general, se entenderá que las dos partes se pueden combinar para proporcionar el conjunto de sitios del amplicón sin apareamiento erróneo, sitios del amplicón de apareamiento erróneo y sus coordenadas genómicas en una única función.

En algunas realizaciones, uno o más de los sitios del amplicón que se identifican en la etapa 204 se pueden eliminar (por ej., mediante una operación de filtración). Por ejemplo, los sitios del amplicón se pueden eliminar si los sitios tiene apareamientos erróneos que no producen, al menos, 3 pares base (pb) desde el extremo 5' de los cebadores izquierdos y desde el extremo 3' de los cebadores derechos. Los sitios del amplicón de esos cebadores que pasaron la operación de filtrado (en adelante se hace referencia como «cebadores candidatos») deben enriquecer el cromosoma de interés a la vez que focalizan múltiples regiones del genoma de referencia (por ej., típicamente 2500 o más). Además, en algunas

realizaciones, tanto las secuencias del brazo izquierdo como las del brazo derecho de los cebadores candidatos deben tener temperaturas de fusión (TM) que oscilan desde por debajo de 60 hasta por arriba de 60 según se computa mediante el modelo vecino más próximo de estabilidad de unión del ADN, en donde los parámetros de estabilidad empírica se acumulan de acuerdo con la secuencia de ácido nucleico. Ver, por ej., Santa Lucia and Hicks 2004. Finalmente, los cebadores candidatos deben tener alta tolerancia a los apareamientos erróneos que ocurren en los 3 pb en el extremo de 5' del brazo izquierdo y en el extremo de 3' del brazo derecho.

Después de la operación de eliminación (o filtración), los sitios restantes del amplicón se procesarán, posteriormente, para generar un conjunto de valores del parámetro para cada cebador candidato. En algunas realizaciones, se calculará la proporción del número de sitios del amplicón que proceden del cromosoma de interés y el número total de sitios del amplicón que han pasado la operación de filtrado. Para cada cebador candidato, la información de enriquecimiento (por ej., la proporción calculada), la información de los sitios del amplicón vinculados, y cualquier otro valor del parámetro se pueden guardar en una base de datos, como por ej., la base de datos 116.

En la etapa 206, se realiza una técnica de optimización para identificar un cebador con un rendimiento previsto óptimo. La técnica de optimización implica la evaluación de una función objetiva para cada cebador candidato. En particular, puede ser conveniente utilizar una función objetiva que maximice una proporción de los sitios capturados que están en el cromosoma de interés, en comparación con el número de sitios capturados que están sobre otros cromosomas. Además, la función objetiva puede maximizar un número de sitios a partir del cromosoma de interés. En algunas realizaciones, los cebadores candidatos pueden incluir, opcionalmente, cebadores que tengan una alta frecuencia de polimorfismos de nucleótidos únicos (SNP) en sus sitios del amplicón previstos. Debido a la presencia de cebadores de alto rango que focalizan secuencias de repetición similares, en algunas realizaciones, solo un subconjunto de los cebadores candidatos con la proporción más alta de sitios que proceden del cromosoma de interés pueden ser identificados de entre los cebadores con altas superposiciones, y los cebadores candidatos restantes se pueden dispersar para uso futuro.

La función objetiva para cada MIP candidata puede ser establecida, en algunas realizaciones, sobre la base de las siguientes matrices:

Conteos del sitio	Único	No único
0 apareamiento erróneo	A	B
1 apareamiento erróneo	C	D

Tabla 1: Conteo del sitio previsto sobre el cromosoma de interés

Conteos del sitio	Único	No único
0 apareamiento erróneo	E	F
1 apareamiento erróneo	G	H

Tabla 2: Conteo del sitio previsto a lo largo del genoma

En las matrices de sonda anteriores, las hileras que están marcadas como «0 apareamiento erróneo» indican las MIP con apareamientos perfectos en ambos brazos, y las hileras que están marcadas como «1 apareamiento erróneo» indican los cebadores que toleran como máximo 1 apareamiento erróneo en uno de sus brazos. En la tabla 1, la columna que está marcada como «única» corresponde al número de sitios sobre el cromosoma de interés que están alineados solo una vez a una MIP respectiva, donde la alineación ocurrió sobre el cromosoma de interés y sobre ningún

5 otro cromosoma. En la tabla 1, la columna que está marcada como «no única» corresponde al número de sitios sobre el cromosoma de interés que están alineados más de una vez a una MIP respectiva, donde la alineación ocurrió múltiples veces sobre el cromosoma de interés, sobre múltiples cromosomas, o ambos. En la tabla 2, la columna que está marcada como «única» corresponde al número de sitios a lo largo de todos los cromosomas (incluso todos los cromosomas 1-22 X, e Y, por ejemplo) que se alinean solo una vez a una MIP respectiva. En otras palabras, el valor E incluye una suma a lo largo de A_i múltiple, la cual representa el número de sitios únicos con cero apareamientos erróneos sobre el cromosoma i -th. De manera similar, la columna que está marcada como «no única» en la tabla 2 corresponde al número de sitios a lo largo de todos los cromosomas que están alineados más de una vez a una MIP respectiva, incluso las instancias donde la alineación ocurrió múltiples veces sobre el mismo cromosoma, sobre múltiples cromosomas, o ambos.

10 Varias funciones objetivas intuitivas se pueden deducir, con facilidad, a partir de estas matrices de sonda. En general, puede ser conveniente seleccionar una MIP que tenga un alto porcentaje de sitios que aparean las secuencias del brazo (o como máximo tienen un pequeño número de apareamientos erróneos, como por ej., un apareamiento erróneo). Además, puede ser conveniente seleccionar una MIP que esté vinculada con un alto porcentaje de sitios únicos. Esto puede estar representado por la selección de MIP que tienen valores para A, C, E o G (o cualquiera de sus combinaciones adecuadas) que son relativamente altos en comparación con B, D, F o H (o cualquiera de sus combinaciones adecuadas). Además, puede ser conveniente seleccionar una MIP que esté vinculada con un alto porcentaje de sitios que están sobre el cromosoma de interés. Esto puede estar representado por la selección de MIP que tienen valores para A, B, C o D (o cualquiera de sus combinaciones adecuadas) que son relativamente altos en comparación con E, F, G o H (o cualquiera de sus combinaciones adecuadas).

15 Por ejemplo, una función objetiva que maximiza o aumenta A/F puede producir menos lecturas ambiguas para los cebadores candidatos que toleran 0 apareamientos erróneos. En un segundo ejemplo, una función objetiva que maximiza o aumenta A/E puede producir lecturas específicas del cromosoma blanco (por ej., el cromosoma 21). Como tercer ejemplo, una función objetiva que maximiza o aumenta A/C selecciona cebadores que tienen sitios de apareamiento significativamente más perfectos que los que tienen 1 apareamiento erróneo, y como resultado de ello, representa una captura eficiente. Como cuarto ejemplo, una función objetiva que maximiza o aumenta A significa que un gran número de sitios están alineados de manera única al cromosoma de interés. Como quinto ejemplo, un cebador óptimo se puede seleccionar para maximizar una función objetiva de $(A+C) / (E+F+G+H)$. Para ilustrar mejor este concepto, a continuación se explican en detalle tres funciones objetivas ilustrativas.

20 **A. Número total de lecturas útiles del cromosoma de interés (P1)**

Una función objetiva ilustrativa para cada cebador o sonda candidata se puede definir como el número total de lecturas útiles a partir del cromosoma de interés (por ej., cromosoma 21):

$$P1 = f(A, B, C, \dots H; K_0, K_1) \tag{1}$$

35 donde K_0 es el coeficiente de captura promedio de 0 sitios de apareamiento erróneo y K_1 es el coeficiente de captura promedio de 1 sitio de apareamiento erróneo. Más específicamente:

$$P1 = K_0 A + K_1 C \tag{2}$$

40 $TMTC = K_0(E + F) + K_1(G + H) \tag{3}$

donde TMTC es el conteo total de etiquetas moleculares.

Dado que:

45
$$K_s = \frac{K_1}{K_0}, \text{ where } 0 < K_s < 1 \tag{4}$$

y que el valor de K_s se puede estimar a partir de los datos experimentales, la ecuación (2) se puede reescribir como:

$$P_1 = \frac{A + K_e C}{(E + F) + K_e (G + H)} \quad (5)$$

mediante el uso de valores que se extraen de las matrices de sonda en las tablas 1 y 2. El numerador de EQ 5 puede considerarse como un número ajustado de sitios utilizables (sobre el cromosoma de interés). El denominador de EQ 5 puede considerarse como un número de sitios sobre el genoma. El valor de P1 como se define en EQ 5 puede considerarse como una fracción de eficiencia o una fracción útil.

B. Número total de sitios eficaces sobre el cromosoma de interés (P2)

Otra función objetiva ilustrativa para cada cebador o sonda candidata se puede definir como el número total de sitios eficaces sobre el cromosoma de interés (por ej., cromosoma 21):

$$P_2 = g(A, B, C, \dots H; K_0, K_1) \quad (6)$$

donde K₀ y K₁ se definen en la ecuación (1). Más específicamente, P2 se puede definir como:

$$P_2 = A + K_e C \quad (7)$$

donde K₀ se define en la ecuación (4). Similar a P₁, el valor de P₂ se puede calcular, además, mediante el uso de valores que se extraen de las matrices de sonda en las tablas 1 y 2. El valor de P₂ puede considerarse como un número ajustado de sitios utilizables.

C. Función de rendimiento exhaustivo de la sonda

Una forma exhaustiva de evaluar una función objetiva para cada cebador o sonda candidata es:

$$P = P_1 * P_2 \quad (8)$$

Incorporar las ecuaciones (5) y (7), la ecuación (8) puede ser reescrita como:

$$P = \frac{(A + K_e C)^2}{(E + F) + K_e (G + H)} \quad (9)$$

Obsérvese que, como se describió anteriormente con relación a la ecuación (4), el valor de K_e se puede estimar mediante el uso de datos experimentales. Más en particular,

$$K_e = \frac{K_1}{K_0} = \frac{\text{molecular tag counts on 1 mismatch sites}}{\text{1 mismatch site count}} \div \frac{\text{molecular tag counts on 0 mismatch sites}}{\text{0 mismatch site count}} \quad (10)$$

(conteos de etiquetas moleculares sobre sitios con 1 apareamiento erróneo - conteo de sitios con 1 apareamiento erróneo, conteos de etiquetas moleculares sobre sitios con 0 apareamiento erróneo, conteo de sitios con 0 apareamiento erróneo)

El valor de P que se define en EQ 9 se puede utilizar como una puntuación compuesta para representar un rendimiento previsto general de un cebador o sonda candidata, y se puede considerar como un producto entre un número ajustado

de sitios utilizables y una fracción de eficiencia (o fracción útil). En un ejemplo, el valor para P que se define en EQ 9 puede ser comparado con un umbral predeterminado para evaluar si se selecciona un cebador o sonda candidata vinculada para posterior prueba o para diagnóstico. Por ejemplo, el umbral predeterminado puede ser un valor como por ej., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o cualquier otro número adecuado.

5 Cualquiera de los ejemplos antes descritos, y cualquier combinación de los ejemplos, se puede utilizar como una función objetiva. La selección de un cebador mediante la optimización de una función objetiva de esta manera tiene la ventaja de reducir una profundidad de lectura necesaria para obtener un número suficiente de lecturas útiles. En principio, la mejora en la profundidad de lectura escala en línea con una mejora en la reducción de ambigüedad.

10 En la etapa 208, un cebador se selecciona a partir del conjunto de cebadores candidatos basado en la técnica de optimización que se realiza en la etapa 206. Por ejemplo, el cebador seleccionado puede corresponder al cebador con el rendimiento previsto óptimo, es decir, el cebador que maximizó la función objetiva como se describe con relación a la etapa 206.

15 En una realización ilustrativa, un número de cebadores (por ej., MIP) están diseñados, sintetizados y probados. Para probar un cebador, se genera un valor para cada uno de uno o más de los números de conteos del sitio previsto en las tablas 1 y 2, basado en un procedimiento de búsqueda y conteo para sitios únicos y no únicos (sin apareamientos erróneo o con un apareamiento erróneo) sobre cada cromosoma de interés y a lo largo del genoma (incluso autosomas y cromosomas sexuales, por ejemplo). Cualquiera de los ejemplos antes descritos o combinaciones de los ejemplos anteriores, se pueden utilizar como una función objetiva, para representar una puntuación que se asigna a cada cebador y representa el rendimiento de ese cebador. Los cebadores se pueden comparar uno con otro basado en su

20 rendimiento, y se pueden seleccionar los cebadores con mejor rendimiento. Como ejemplo, la puntuación de un cebador se compara con un umbral predeterminado para determinar si el cebador se selecciona o no. En particular, cuando la función objetiva que está representada por EQ 9 se utiliza para representar la puntuación de cada cebador, el umbral se puede establecer hasta un número, como por ej., 2, 3, 4, 5, 6, 7, 8, 9 o 10.

25 En una realización ilustrativa, un cebador con una alta puntuación (por ej., MIP 003 en la tabla anterior) se seleccionó para posterior análisis. Los resultados del análisis posterior se describen a continuación con relación al ejemplo 3. La MIP 003 se seleccionó porque estaba vinculada con un valor alto para la puntuación como se evaluó de acuerdo con EQ 9. En particular, la MIP 003 tuvo una puntuación de 8,557. La tabla 3, a continuación, resume los números previstos de sitios únicos que están alineados a la MIP 003 mediante diferentes cromosomas sin apareamiento erróneo (A) y con un apareamiento erróneo (C).

Tabla 3

	Cromosoma	Número total de sitios únicos (A+C)	Sitios únicos sin apareamiento erróneo (A)	Sitios únicos con un apareamiento erróneo (C)
5	chr1	15986	5499	10487
	chr2	13621	4773	8848
	chr3	11126	3926	7200
	chr4	9154	3227	5927
10	chr5	9474	3279	6195
	chr6	9681	3326	6355
	chr7	10826	3812	7014
	chr8	7735	2652	5083
15	chr9	7733	2668	5065
	chr10	8786	3031	5755
	chr11	7857	2714	5143
	chr12	9506	3330	6176
20	chr13	4827	1725	3102
	chr14	5898	2019	3789
	chr15	6150	2146	4004
	chr16	7364	2439	4925
	chr17	9378	3315	6063
25	chr18	3899	1399	2500
	chr19	9360	3203	6157
	chr20	4684	1618	3066
	chr21	2100	783	1317
30	chr22	4092	1442	2650
	chroX	7363	2547	4816
	chroY	508	152	356

35 Se contempla que las etapas o las descripciones del procedimiento 200 se pueden utilizar con cualquier otra realización de esta divulgación. Además, las etapas y las descripciones que se describen con relación a la figura 2 se pueden llevar a cabo en órdenes alternativos o en paralelo con el propósito de esta divulgación. Por ejemplo, cada una de estas etapas se puede llevar a cabo en cualquier orden o en paralelo o sustancialmente de forma simultánea para reducir el retraso o incrementar la velocidad del sistema o el método. Más aún, debe observarse que el procedimiento 200 se puede llevar a cabo mediante el uso del dispositivo de computación 100 y, más en particular, de la CPU 106 del dispositivo de computación 100.

45 La figura 3 es un diagrama de flujo de un procedimiento 300 para pronosticar el estado de aneuploidía en un sujeto de prueba, de acuerdo con una realización ilustrativa. El procedimiento 300 incluye las etapas de recibir los datos de secuenciación para un sujeto de prueba (etapa 302), computar la métrica de captura del sitio para el sujeto de prueba (etapa 304), recibir las métricas de captura del sitio para un conjunto de sujetos de referencia (etapa 306), y pronosticar el estado de aneuploidía en el sujeto de prueba basado en la comparación de la métrica de captura del sitio para el sujeto de prueba con las métricas de captura del sitio para los sujetos de referencia (etapa 308). En algunas

realizaciones, la métrica de captura del sitio es un índice de eficiencia de captura del sitio (SCE), el cual es la relación entre el número de etiquetas identificatorias moleculares únicas y el número de lecturas para cada sitio individual. En algunas realizaciones, la métrica de captura del sitio es una medida de consistencia de la captura del sitio (SCC), la cual se calcula como el coeficiente de variabilidad de SCE para cada sitio individual. Por ejemplo, en el cromosoma 1 : 1-100, tenemos 100 lecturas alineadas y 99 etiquetas identificatorias moleculares únicas. En consecuencia, el SCE será del 99 %. En otro ejemplo, fuera de 100 muestras para el cromosoma 1 : 1-100, las 100 muestras tienen un SCE por encima del 90 %, mientras que fuera de las 100 muestras para el cromosoma 3 : 500-600, solo 50 muestras tienen un SCE por encima del 90 % y las restantes 50 muestras tienen un SCE inferior al 90 %. En consecuencia, la SCC indicará que el cromosoma 1 : 1-100 es un sitio más consistente en comparación con el cromosoma 3 : 500-600.

En la etapa 302, se reciben los datos de secuenciación para un sujeto de prueba. En particular, el sujeto de prueba tiene un estado de aneuploidía que se desconoce, y los datos de secuenciación recibidos se obtienen mediante la obtención de una muestra de ácido nucleico del sujeto de prueba y mediante el uso de una población de cebadores, como por ej., las sondas de inversión molecular causantes de la repetición (RO-MIP), para capturar un conjunto de sitios en la muestra de ácido nucleico. Como se describe en detalle con relación a la figura 5, cada RO-MIP incluye en secuencia un primer brazo de polinucleótidos de focalización, una primera etiqueta molecular de focalización única, un ligador polinucleotídico, una segunda etiqueta molecular de focalización única, y un segundo brazo de polinucleótidos de focalización. El primero y el segundo brazo de polinucleótidos de focalización son iguales a lo largo de las MIP en la población, mientras que la primera y la segunda etiqueta molecular de focalización únicas son distintas a lo largo de las MIP en la población. Los amplicones de las RO-MIP surgen de la captura de los sitios, y los amplicones son secuenciados para obtener los datos de secuenciación.

En la etapa 304, una fracción cromosómica se computa para el sujeto de prueba mediante una relación entre una suma de todos los eventos de captura únicos a partir de un cromosoma de interés (S1) y una suma de todos los eventos de captura únicos a partir de todos los cromosomas (S1+ S2). La fracción cromosómica proporciona una medida proporcional del cromosoma de interés en una muestra dada (es decir, la proporción de lecturas que proviene del cromosoma de interés). Un método de ejemplo de computación de la fracción cromosómica se describe con relación a las etapas 426, 428 y 430 en la figura 4.

En la etapa 306, se recibe un conjunto de fracciones cromosómicas para un conjunto de sujetos de referencia. En particular, los sujetos de referencia pueden corresponder a un grupo de personas que exhiben un estado de euploidía conocido. Por ejemplo, los sujetos pueden exhibir monosomía, disomía o trisomía para un cromosoma de interés particular. Las fracciones cromosómicas para los sujetos de referencia se computan de la misma manera que la que se describió con relación a la etapa 304, pero para cada sujeto de referencia. Como se describe en detalle con relación a la figura 4, las fracciones cromosómicas son representativas de la capacidad de la RO-MIP seleccionada para capturar las secuencias sobre el cromosoma de interés, en comparación con otros cromosomas.

En la etapa 308, la fracción cromosómica para el sujeto de prueba (que se computa en la etapa 304) se compara con las fracciones cromosómicas para los sujetos de referencia (que se obtienen en la etapa 306), y el estado de euploidía del sujeto de prueba se pronostica basado en esta comparación. En particular, una prueba estadística se puede utilizar para comparar la fracción cromosómica de prueba con la población de fracciones cromosómicas de referencia, y determinar si la fracción cromosómica de prueba pertenece a algún clúster de fracciones cromosómicas de referencia vinculado con el mismo estado de euploidía. Como se utiliza en la presente memoria, la «relación de prueba» puede incluir la fracción cromosómica.

La figura 4 es un diagrama de flujo de un procedimiento 400 para pronosticar el estado de euploidía en un sujeto de prueba, de acuerdo con una realización ilustrativa. En un ejemplo, el procedimiento 400 se puede utilizar para implementar la etapa 308 del procedimiento 300 que se muestra y se describe con relación a la figura 3. Como se describió en detalle con relación a la figura 3, una fracción cromosómica se puede utilizar para pronosticar un estado de euploidía en un sujeto de prueba que tiene un estado de euploidía desconocido.

El procedimiento 400 incluye las etapas de recibir los datos que se registran a partir de las muestras S con los genotipos conocidos (etapa 402) e inicializar un parámetro de iteración de la muestra s a 1 (etapa 404). Para cada muestra s, el procedimiento 400 incluye filtrar las lecturas de secuenciación para eliminar los artefactos conocidos (etapa 406), alinear las lecturas al genoma humano (etapa 408), establecer un parámetro de iteración del sitio k en 1 (etapa 412), y determinar un número de eventos de captura para el sitio k-th (etapa 414). Cuando todos los sitios K y todas las muestras S han sido consideradas, el procedimiento 400 incluye, además, las etapas de identificar un subconjunto de los sitios K que satisfacen, al menos, un criterio (etapa 424), computar una primera suma S1 de conteos de etiquetas a partir de un cromosoma de interés para una muestra de prueba t en las muestras S (etapa 426), computar una segunda suma S2 de conteos de etiquetas a partir de los cromosomas además del cromosoma de interés para la muestra de prueba t (etapa 428), computar una fracción cromosómica como $S1/(S1+S2)$ para la muestra de prueba t (etapa 430) y seleccionar un estado de euploidía para la muestra de prueba t mediante la comparación de la fracción cromosómica para la muestra de prueba t y un conjunto de fracciones cromosómicas de referencia (etapa 432).

En la etapa 402, se reciben los datos que se registran a partir del conjunto de muestras S, donde las muestras S se obtienen, cada una de ellas, de un sujeto diferente. Al menos una de las muestras S se obtiene de un sujeto de prueba, cuyo estado de euploidía puede ser desconocido. Las muestras pueden ser muestras de ácido nucleico que se aísla de los sujetos, y los datos pueden incluir los datos de secuenciación que se obtienen a partir de las muestras de ácido nucleico. En un ejemplo, los datos de secuenciación se obtienen mediante el uso de una población de RO-MIP para

amplificar un conjunto de sitios en la muestra de ácido nucleico para producir un conjunto de amplicones de las RO-MIP. Los amplicones de las RO-MIP se pueden secuenciar para obtener los datos de secuenciación que se recibieron en la etapa 402.

5 En la etapa 404, un parámetro de iteración de la muestra s se inicializa en 1. A medida que se procesan las muestras S , el parámetro de iteración de la muestra s se incrementa hasta que cada una de las muestras S se procesan para determinar el número de eventos de captura para cada sitio.

10 En la etapa 406, las lecturas de secuenciación para la muestra s se filtran para eliminar los artefactos conocidos. En un ejemplo, los datos que se reciben en la etapa 402 se pueden procesar para eliminar un efecto de interacción de sonda a sonda. En algunas realizaciones, los brazos de focalización de ligación y extensión de todas las RO-MIP se aparean a las lecturas de la secuencia apareada en el extremo. Las lecturas que no se aparearon con ambos brazos de las RO-MIP se determinan como inválidas y se descartan. Las secuencias del brazo para las restantes lecturas válidas se eliminan, y las etiquetas moleculares de ambos extremos de ligación y extensión también se pueden eliminar de las lecturas. Las etiquetas moleculares eliminadas se pueden conservar por separado para un procesamiento posterior en la etapa 414.

15 En la etapa 408, las lecturas recortadas resultantes se alinean al genoma humano. En algunas realizaciones, una herramienta de alineación se puede utilizar para alinear las lecturas a un genoma humano de referencia. En particular, se puede evaluar una puntuación de alineación para representar lo bien que una lectura específica se alinea a la referencia. Las lecturas anteriores con puntuaciones de alineación por encima de un umbral se pueden considerar en la presente memoria como alineaciones primarias, y se retienen. Por el contrario, las lecturas con puntuaciones de
20 alineación por debajo del umbral se pueden considerar, en la presente memoria, como alineaciones secundarias, y se descartan. Cualquier lectura que se alinee a las múltiples ubicaciones a lo largo del genoma de referencia se puede considerar, en la presente memoria, como múltiples alineaciones, y se descartan.

25 En la etapa 412, un parámetro de iteración del sitio k se inicializa en uno. En la etapa 414, se determina el número de eventos de captura para el sitio k -th, y el parámetro de iteración del sitio k es incrementado en la etapa 418 hasta que todos los sitios K hayan sido considerados.

Cuando todos los sitios K han sido considerados, el procedimiento 400 procede a la etapa 424 para identificar un subconjunto de los sitios K que satisface, al menos, un criterio. Por ejemplo, una medida de consistencia de captura del sitio se puede evaluar como un coeficiente de variación del número de eventos de captura a lo largo de las muestras S , y estos sitios que tienen altos coeficientes de variación se pueden descartar.

30 En la etapa 426, se computa una suma $S1$ de los conteos de etiquetas a partir del cromosoma de interés para una muestra de prueba, y en la etapa 428, se computa una suma $S2$ de los conteos de etiquetas a partir de los cromosomas además del cromosoma de interés para la muestra de prueba. En la etapa 430, se computa una fracción cromosómica como $S1/(S1+S2)$. La fracción cromosómica para la muestra de prueba se compara con un conjunto de fracciones cromosómicas de referencia (que han sido computadas a partir de los sujetos de referencia que tienen estados de
35 euploidía conocidos), y se realiza una prueba estadística para seleccionar un estado de euploidía que se pronostica para el sujeto de prueba en la etapa 432.

El orden de las etapas en la figura 4 se muestra para fines ilustrativos únicamente, y no es limitante.

Un experto en la técnica entenderá que las composiciones y los métodos que se describen en la presente memoria se pueden adaptar y modificar.

40 Esta divulgación se entenderá mejor a partir de los detalles experimentales que siguen a continuación. No obstante, un experto en la técnica apreciará que los métodos y los resultados específicos que se discuten son meramente ilustrativos de la divulgación, ya que esta se describe por completo en las realizaciones que siguen de aquí en adelante.

Ejemplos

Ejemplo 1: Diseño de la MIP y método para capturar secuencias diana de interés

45 *Construcción de la sonda*

Un oligonucleótido único que oscila en tamaño entre 80 y 105 pb (según la extensión del primero y del segundo brazo de polinucleótidos de focalización) se sintetiza como se muestra en la figura 5. Las cajas 6N se refieren a las secuencias de la etiqueta molecular que se utilizan para cuantificar los eventos de captura para cada secuencia diana de interés. En esta realización particular, en lugar del conteo de lecturas, se cuenta el número de secuencias únicas por sitio
50 capturado.

Reacción de captura del sitio

El primero y el segundo brazo de polinucleótidos de focalización; (en una concentración que se determina empíricamente) se mezcla con ADNcs que se extrae a partir de 1-2 ml de plasma en una reacción amortiguada de 20 ul. La mezcla se incuba en un termociclador a temperaturas que han sido optimizadas para hibridación de las sondas en el
55 molde (98 °C durante 3 min □ 85 °C durante 30 min., 66 °C durante 15 min). Durante esta incubación, las moléculas de la sonda hibridan al molde de ADNcs en ubicaciones cromosómicas específicas que son complementarias a la secuencia de la sonda (figura 6, medio). Los sitios que se predicen con mayor facilidad son aquellos con secuencias que

son exactamente complementarias al primero y al segundo brazo de polinucleótidos de focalización; (sitios invariables), pero los sitios que tienen una o más variantes en cualquier brazo también se focalizan, de algún modo, a una eficiencia más baja. La cantidad óptima de la MIP para cada reacción depende de tres variables principales: 1) el número de genomas que se utiliza como molde, el cual puede variar ampliamente entre los individuos, 2) el número total de sitios que son focalizados por la sonda específica, y 3) la relación de sitios invariables con los sitios variables.

Después de que se completa el programa de hibridación, se agrega una mezcla de 5 ul de enzimas y reactivos, y la mezcla se incuba a 66 °C durante 1 hora, a continuación a 72 °C durante 30 min, posteriormente se enfría hasta 4 °C. Vemos un incremento drástico en la especificidad al elevar tanto las temperaturas de hibridación como de extensión hasta 66 °C. Durante esta etapa, el hueco es llenado por la ADN polimerasa y la MIP es circularizada, de manera covalente, por la ADN ligasa (figura E, inferior). Solo las sondas que se hibridan al molde son circularizadas durante esta etapa, y las sondas lineales no son amplificables debido al hueco entre los sitios de unión al cebador.

Amplificación del sitio capturado

Se agregan 20 ul de la mezcla de la MIP capturada a una mezcla de reacción de 50 ul que contiene polimerasa termoestable, dNTPS, amortiguador de PCR y cebadores universales que son complementarios al esqueleto de la sonda (figura F, superior). Cada muestra se amplifica mediante el uso de un cebador con una secuencia de «código de barra» distinta que permite la secuenciación múltiplex de una biblioteca de muestra agrupada. La reacción se somete a un número determinado empíricamente de ciclos de PCR hasta que una banda de amplicón limpio se puede observar mediante electroforesis. El producto de PCR se purifica mediante el uso de microesferas Ampure y se cuantifica mediante el uso de un fluorómetro Qubit.

Secuenciación de los sitios capturados

Los productos de PCR purificados se agrupan en una biblioteca de modo que todas las muestras están a una concentración igual, lo que asegura que el presupuesto de lectura se divide de manera equitativa a lo largo de las muestras. La biblioteca es secuenciada mediante el uso de secuenciación de extremo único o de extremo apareado, mediante el uso de 75-100 ciclos a fin de determinar la secuencia completa del hueco específico del sitio. Si se utiliza la secuenciación de extremo único, la lectura consistirá en el brazo de ligación seguido de la etiqueta molecular y la secuencia de hueco única que se llenó durante la etapa de extensión/ligación. La secuenciación en el brazo de extensión es innecesario porque la secuencia es conocida a partir de la sonda.

Ejemplo 2: Análisis de los datos de la secuencia y detección de aneuploidía en un sujeto

Los datos de secuenciación sin procesar se deben procesar a fin de que sean útiles en la detección de aneuploidía. Para comenzar, las lecturas de secuenciación se filtran para eliminar los artefactos conocidos como por ej., la interacción sonda a sonda, las secuencias del esqueleto o las secuencias del adaptador. Los brazos de ligación y extensión de la MIP (es decir, el primero y el segundo brazo de polinucleótidos de focalización) se aparean a las lecturas de secuencias, lo que permite un máximo de un apareamiento erróneo de par base en cada brazo. Las lecturas que no cumplen con este criterio se tratan como inválidas y se descartan. Las secuencias del brazo para las lecturas válidas se recortan para eliminar las partes no genómicas de los datos de la secuencia mediante la eliminación de la secuencia de los archivos de lectura que se utilizaron en las etapas de procesamiento subsiguientes. Al mismo tiempo, las etiquetas moleculares tanto del extremo de ligación como del extremo de extensión se mantienen por separado para el conteo de los eventos de captura en una etapa posterior. Las lecturas recortadas se alinean al genoma humano (hg19) con el programa de software bowtie2. Las lecturas alineadas se filtran con el programa de software Samtools para eliminar las lecturas que no se alinean como un par válido, o que se alinean de manera no única. Los parámetros de alineación y filtración se eligen cuidadosamente para mantener las lecturas alineadas de manera única solamente. Las lecturas alineadas y filtradas (en archivos de formato bam) se examinan para contar las etiquetas moleculares únicas para cada sitio focalizado con una secuencia de hueco de la MIP única (es decir, una secuencia diana de interés). Estos conteos son el número inicial de los eventos de hibridación de la MIP a blanco que son secuenciados en una plataforma de secuenciación de próxima generación (por ej., una celda de flujo Illumina HiSeq 2500). Un índice de eficiencia de captura del sitio (SCE) se calcula para cada sitio individual. Una medida de consistencia de la captura del sitio (SCC) se calcula como el coeficiente de variabilidad de SCE para cada sitio individual. Una parte de estos sitios se descartan basado en las medidas de variabilidad de eficiencia de captura del sitio a partir de los datos experimentales.

Esta etapa ayuda a reducir la variabilidad de muestra a muestra en la proporción cromosómica y a aumentar la puntuación Z de las muestras positivas que se calculan en una etapa posterior. Para una muestra dada, la suma (S1) de los conteos de las etiquetas moleculares únicas de los sitios restantes sobre el cromosoma de interés (por ej., cromosoma 21) se calcula y se almacena. La suma (S2) de los sitios de referencia restantes se calcula y se almacena. La relación entre las dos sumas (fracción cromosómica = $S1/(S1 + S2)$) se calcula como la medida proporcional del cromosoma de interés en una muestra dada. Mediante el uso de las medidas de fracción cromosómica, se calcula la puntuación Z del cromosoma de interés (por ej., el cromosoma 21) para cada muestra de prueba contra una colección de muestras de referencia con el conteo de los cromosomas fetales normales (es decir, feto normal cariotipo). La determinación de aneuploidía (por ej., trisomía y monosomía) o euploidía se realiza cuando el valor absoluto de la puntuación Z de una muestra dada está por encima de un cierto umbral.

Mediante el uso del método anterior, 48 muestras de mujeres embarazadas se ensayaron para el síndrome de Down (trisomía del cromosoma 21). El ensayo confirmó que dos muestras fueron positivas para el síndrome de Down, mientras que 46 muestras fueron negativas. La separación entre las muestras positivas y negativas fue particularmente

clara, con una puntuación Z mayor que 6 en ambas muestras positivas. Ver la figura 8.

Ejemplo 3: Detección de las trisomías 13, 18 y 21 en sujetos embarazados

Este ejemplo describe el uso del método que se describe en la presente memoria para discriminar (o distinguir) entre las mujeres embarazadas que portan trisomía 13, 18 y 21, y las mujeres embarazadas saludables.

5 **Determinación de los cortes de puntuación Z para detectar las trisomías 13, 18 y 21**

Un conjunto para capacitación de 48 muestras de plasma de mujeres embarazadas (estudio aprobado por la junta de revisión institucional) se utiliza para determinar los cortes de puntuación Z para detectar las trisomías 13, 18 y 21. Cada una de las muestras de plasma se toma de una mujer embarazada diferente. Las 48 muestras de plasma contienen 40 muestras saludables, cuatro muestras de trisomía 21, tres muestras de trisomía 18 y una muestra de trisomía 13. Los casos de trisomía se confirman mediante el análisis de cariotipo de las células de la placenta o del feto que se recolectan mediante CVS (muestra de vellosidades coriónicas) o amniocentesis.

Reacción de captura del sitio: El ADN en plasma que se extrajo de cada una de las 48 mujeres embarazadas se somete a una reacción de captura de sitio de la siguiente manera: el ADN en plasma extraído se mezcla con agua, amortiguador Ampligase (1x) y la RO-MIP a la concentración de sonda determinada empíricamente. La mezcla de reacción de captura del sitio se incuba en un ciclador térmico a 98 °C durante 3 min., a continuación 85 °C durante 30 min, posteriormente 66 °C durante 15 min. Después de la incubación, una mezcla maestra que contiene dNTP (0,6 mM), NAD (0,4x), betaína (0,3M), amortiguador Ampligase (1x), agua, Ampligase (5 unidades) y polimerasa Phusion HF (0,4 unidades) se agrega a la mezcla de reacción de captura del sitio. La mezcla combinada se incuba en un ciclador térmico a 66 °C durante 60 min, 72 °C durante 30 min, y se mantiene a 4 °C.

Amplificación del sitio capturado: Después de la incubación, se agregan 20 ml de la mezcla combinada a una mezcla maestra de PCR que contiene cebadores de indexado hacia adelante (500 nM) e inverso, amortiguador Phusion HF (1x), dNTPs (0,2 mM), agua y polimerasa Phusion HS (0,4 unidades). A cada muestra se le coloca un código de barra con un único índice de cebador inverso. La mezcla de reacción de PCR se incuba en un ciclador térmico a 98 °C durante 3 min, a continuación 20 ciclos a 98 °C durante 10 seg., 65 °C durante 20 seg., y 72 °C durante 30 seg., La mezcla de reacción de PCR se mantiene a 72 °C durante 5 min y a continuación a 4 °C.

Secuenciación de único extremo: Las bibliotecas amplificadas por PCR se purifican mediante el uso de microesferas AmpureXP y las muestras se agrupan a igual concentración (48 muestras por grupo). Las bibliotecas multiplexadas se cargan, cada una de ellas, sobre una celda de flujo única SR y se secuencian en modo Rapid Run sobre una HiSeq 2500 durante 106 ciclos.

Análisis de datos: Los datos de secuenciación se filtran y se alinean. Se calcula una proporción cromosómica (P_i) para los cromosomas 13, 18 y 21 dividiendo el número de etiquetas / identificadores moleculares únicos a partir de las lecturas que se alinean de manera única a cada cromosoma (es decir, el cromosoma 13, 18 y 21) (este numerador es análogo a la suma de «A» y «C») por el número total de etiquetas / identificadores moleculares únicos a partir de las lecturas que se alinean de manera única a los cromosomas 1 a 22 (este denominador es análogo a la suma de «A», «C», «E» y «G»). No obstante, mientras que los valores en la tabla 2 (E-H) incluyen todos los cromosomas a lo largo del genoma (incluso los cromosomas 1 a 22, X e Y), el denominador aquí puede no incluir las lecturas que se alinean a los cromosomas X o Y. En particular, el número de lecturas que se alinean a los cromosomas X o Y puede ser significativamente más grande que el número de lecturas que se alinean a otros cromosomas, incluso los cromosomas de interés. En este caso, incluso las lecturas que se alinean a los cromosomas X o Y en el denominador pueden reducir, significativamente, la relación resultante, y pueden introducir ruido y distorsión al cálculo de la estadística z. En consecuencia, cuando el cromosoma de interés es ni el cromosoma X ni el cromosoma Y (por ej., cuando se detecta una aneuploidía autosómica), las lecturas que se alinean a los cromosomas X o Y se pueden excluir tanto del numerador como del denominador. De forma alternativa, cuando el cromosoma de interés es un cromosoma sexual (por ej., cuando se detecta una aneuploidía cromosómica sexual), las lecturas que se alinean a los cromosomas X y/o Y se pueden incluir en el numerador y en el denominador.

Para las relaciones resultantes, la desviación media (x_i) y estándar (s_i) se calcula para las muestras no afectadas y se utilizan para calcular una puntuación Z ($(P_i - x_i)/s_i$) para cada muestra. El análisis de las puntuaciones Z revela que todas las muestras saludables tienen puntuaciones Z < 3,0 para los cromosomas 13, 18 y 21, mientras que las muestras positivas tienen puntuaciones Z > 3,0 para los cromosomas 13, 18 y 21. Por lo tanto, se determina que las muestras con una puntuación Z > 3,0 se denominarán positivas para las trisomías 13, 18 y 21. Las muestras con una puntuación Z < 3,0 se denominarán negativas para las trisomías 13, 18 y 21.

Determinación de sensibilidad y especificidad para detectar las trisomías 13, 18 y 21

Un conjunto de prueba de 422 muestras de mujeres embarazadas (estudio aprobado por la junta de revisión institucional) se utiliza para determinar la sensibilidad y la especificidad de la prueba. Cada muestra es de una mujer embarazada diferente. Estas 422 muestras contienen 387 muestras saludables, 21 muestras T21, 9 muestras T18 y 5 muestras T13. Los casos de trisomía se confirman mediante el análisis de cariotipo de las células de la placenta o del feto por CVS o amniocentesis.

Reacción de captura del sitio: El ADN en plasma que se extrajo de cada una de las 422 mujeres embarazadas se somete a una reacción de captura de sitio de la siguiente manera: el ADN en plasma extraído se mezcla con agua,

amortiguador Ampligase (1x) y la RO-MIP a la concentración de sonda determinada empíricamente. La mezcla de reacción de captura se incuba en un ciclador térmico a 98 °C durante 3 min, a continuación 85 °C durante 30 min, posteriormente 66 °C durante 15 min. Después de la incubación, una mezcla maestra que contiene dNTP (0,6 mM), NAD (0,4x), betaína (0,3M), amortiguador Ampligase (1x), agua, Ampligase (5 unidades), y polimerasa Phusion HF (0,4 unidades) se agrega a la mezcla de reacción de captura del sitio. La mezcla de reacción combinada se incuba en un ciclador térmico a 66 °C durante 60 min, 72 °C durante 30 min y se mantiene a 4 °C.

Amplificación del sitio capturado: Después de la incubación, se agregan 20 ml de la mezcla combinada a una mezcla maestra de PCR que contiene cebadores de indexado hacia adelante (500 nM) e inverso, amortiguador Phusion HF (1x), dNTPs (0,2 mM), agua y polimerasa Phusion HS (0,4 unidades). A cada muestra se le coloca un código de barra con un único índice de cebador inverso. La mezcla de reacción de PCR se incuba en un ciclador térmico a 98 °C durante 3 min, a continuación 20 ciclos a 98 °C durante 10 seg., 65 °C durante 20 seg., y 72 °C durante 30 seg., La mezcla de reacción de PCR se mantiene a 72 °C durante 5 min y a continuación a 4 °C

Secuenciación de único extremo: Las bibliotecas amplificadas se purifican mediante el uso de microesferas AmpureXP y las muestras se agrupan a igual concentración (48 muestras por grupo). Hay 9 grupos en total: los primeros ocho grupos tienen 47 muestras de prueba y 1 muestra de control por grupo y el noveno grupo tiene 46 muestras de prueba y 1 muestra control. Las bibliotecas multiplexadas se cargan, cada una de ellas, sobre una celda de flujo única SR y se secuencian en modo Rapid Run sobre una HiSeq 2500 durante 106 ciclos.

Análisis de datos: El ADN de cada una de las 422 mujeres embarazadas se someten a la reacción de captura del sitio, a la amplificación del sitio capturado y a la secuenciación de único extremo mediante el uso de 106 ciclos, como con el conjunto para capacitación. Los datos de la secuencia de cada muestra se utilizan para calcular una puntuación Z para los cromosomas 13, 18 y 21 (figuras 10-12) Mediante el uso de los cortes que se determinan por el conjunto de capacitación, la especificidad (por ej., el índice negativo cierto o el número de negativos ciertos divididos por la suma de negativos ciertos y falsos positivos) es > 99,9 % para las trisomías 13 (5/5), 18 (9/9) y 21 (21/21). La sensibilidad (por ej., el índice positivo cierto o el número de positivos ciertos dividido por la suma de positivos ciertos y falsos negativos) es > 99,2 % para el cromosoma 13, > 99,9 % para el cromosoma 18 y > 99,5 % para el cromosoma 21.

El gráfico en la figura 10 muestra el rendimiento de la prueba del cebador en la detección de la trisomía 13. El eje y en la figura 10 representa la estadística z, y el eje x representa las cinco muestras de la trisomía 13 a la izquierda (en ningún orden particular) y las restantes 545 muestras a la derecha (en el orden de estadística z decreciente). Como se muestra en la figura 10, las cinco muestras de la trisomía 13 tuvieron estadística z más alta que 3,0 y se identificaron correctamente como muestras positivas mediante el uso del cebador, que provoca una sensibilidad 5/5. Un pequeño número de muestras saludables tuvo estadística z mayor que 3,0 que provoca una especificidad del 99,2 %.

La figura 11 es similar a la figura 10, salvo que la figura 11 muestra el rendimiento de la prueba del cebador en la detección de la trisomía 18. Las nueve muestras de la trisomía 18 tuvieron estadística z más alta que 3,0 y se identificaron correctamente como muestras positivas mediante el uso del cebador, que provoca una sensibilidad 9/9. Además, casi todas las muestras saludables se identificaron correctamente como muestras negativas, que provocan un índice de especificidad más grande que 99,9 %. La fig. 12 es, además, similar a la figura 10, salvo que la figura 12 muestra el rendimiento de la prueba del cebador en la detección de la trisomía 21. Las 21 muestras de la trisomía 21 se identificaron correctamente como muestras positivas mediante el uso del cebador, que provoca una sensibilidad 21/21. Además, casi todas las muestras saludables se identificaron correctamente como muestras negativas, que provocan un índice de especificidad mayor que 99,8 %.

Los resultados que se muestran en las figuras 10-12 indican que los sistemas y los métodos de la presente divulgación proporcionan una herramienta útil para la selección de un cebador con alta sensibilidad y especificidad en la detección de aneuploidía. Además, otros métodos tienen niveles similares de rendimiento (según se mide por sensibilidad y/o especificidad, por ejemplo). Por ejemplo, se puede obtener un rendimiento similar mediante el uso de secuenciación por bombardeo con microproyectiles. Sin embargo, la presente divulgación tiene varias ventajas sobre otros enfoques porque la presente divulgación utiliza solo un único cebador, y por lo tanto, puede ser más económico, más simple y más eficiente que los de los otros enfoques.

A los fines ilustrativos, los ejemplos que se proporcionan mediante esta divulgación ponen énfasis esencialmente sobre un número de diferentes realizaciones ilustrativas de los sistemas y los métodos para determinar las variantes del número de copias, las anomalías cromosómicas o las micro supresiones. Sin embargo, se entiende que se pueden realizar variantes en la forma y el diseño general de una o más realizaciones sin cambiar significativamente las funciones y las operaciones de la presente divulgación. Más aún, debe observarse que las características y las limitaciones que se describen en cualquiera de las realizaciones se pueden aplicar a cualquier otra realización en la presente memoria, y las descripciones y los ejemplos que se refieren a una realización se pueden combinar con cualquier otra realización de una manera adecuada. Además, las figuras y los ejemplos que se proporcionan en la divulgación están destinadas a ser solo ilustrativas y no limitantes. Debe observarse que los sistemas y/o los métodos que se describieron anteriormente se pueden aplicar, o se pueden utilizar de acuerdo con otros sistemas y/u otros métodos, incluso sistemas y/o métodos que pueden o no estar directamente relacionados con la determinación de las variantes del número de copias.

REIVINDICACIONES

1. Un método de detección sistemática para la aneuploidía en un feto, que comprende:

- a) obtener una muestra de ácido nucleico que se aísla a partir de una muestra de sangre materna;
- 5 b) capturar una pluralidad de secuencias diana de interés en la muestra de ácido nucleico que se obtiene en la etapa a) mediante el uso de una o más poblaciones de sondas de inversión molecular (MIP) para producir una pluralidad de replicones, en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes: primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización; en donde los pares del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en la pluralidad de secuencias diana de interés; en donde la primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;
- 10 c) secuenciar una pluralidad de amplicones de MIP que se amplían a partir de los replicones que se obtienen en la etapa b);
- d) determinar el número de eventos de captura de cada una de una primera población de amplicones de la pluralidad de amplicones que se proporcionan en la etapa c) basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la primera población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- 20 e) determinar el número de eventos de captura de cada una de una segunda población de amplicones de la pluralidad de amplicones que se proporciona en la etapa c), basado en el número de las etiquetas moleculares únicas de cada MIP que amplificó un replicón, en donde la primera población de amplicones se determina mediante la secuencia de la secuencia diana de interés;
- f) determinar, para cada secuencia diana de interés a partir de la cual la primera población de amplicones se produjo, una métrica de captura del sitio, basado al menos en parte en el número de eventos de captura que se determinaron en la etapa d);
- 25 g) identificar un primer subconjunto de la métrica de captura del sitio que se determinó en la etapa f) que satisface, al menos, un criterio;
- h) determinar, para cada secuencia diana de interés a partir de la cual la segunda población de amplicones se produjo, una métrica de captura del sitio, basado al menos en parte en el número de eventos de captura que se determinaron en la etapa e);
- 30 i) identificar un segundo subconjunto de la métrica de captura del sitio que se determinó en la etapa h) que satisface, al menos, un criterio;
- 35 j) normalizar una primera medida que se determina a partir del primer subconjunto de métricas de captura del sitio que se identifica en la etapa g) mediante una segunda medida que se determina a partir del segundo subconjunto de métricas de captura del sitio que se identifica en la etapa i) para obtener una relación de prueba;
- k) comparar la relación de prueba con una pluralidad de relaciones de referencia que se computan basado en las muestras de ácido nucleico de referencia que se aislaron a partir de los sujetos de referencia, que se sabe exhiben euploidía o aneuploidía, para determinar la presencia o la ausencia de una relación de prueba que indica la aneuploidía en el feto.

2. El método de la reivindicación 1, en donde la muestra de ácido nucleico es ADN o ARN.

3. El método de cualquiera de las reivindicaciones 1 o 2, en donde la métrica de captura del sitio es un índice de eficiencia de captura del sitio (SCE) o es una medida de consistencia de captura del sitio (SCC).

45 4. El método de cualquiera de las reivindicaciones 1-3, en donde cada uno de los replicones de las MIP que se proporcionan en la etapa b) es producido por:

- i) el primero y el segundo brazo de polinucleótidos de focalización, respectivamente, que hibridan a la primera y a la segunda región en la muestra de ácido nucleico, respectivamente, en donde la primera y la segunda región flanquean una secuencia diana de interés; y
- 50 ii) después de la hibridación, utilizan una mezcla de ligación/extensión para extender y ligar la región con huecos entre los dos brazos de polinucleótidos de focalización para formar moléculas de ácido nucleico circulares monocatenarias.

55 5. El método de cualquiera de las reivindicaciones 1-4, en donde el método comprende, antes de la etapa de secuenciación de c), una reacción de PCR de indexado, en donde la reacción de PCR de indexado introduce en cada uno de los amplicones de las MIP los siguientes componentes: un par de cebadores de indexado, un único código de

barras de muestra y un par de adaptadores de secuenciación, preferiblemente en donde los amplicones de las MIP con código de barra comprenden en secuencia los siguientes componentes: un primer adaptador de secuenciación, un primer cebador de secuenciación, la primera etiqueta molecular de focalización única, el primer brazo de polinucleótidos de focalización, el ácido nucleico capturado, el segundo brazo de polinucleótidos de focalización, la segunda etiqueta molecular de focalización única, un único código de barra de muestra, un segundo cebador de secuenciación, un segundo adaptador de secuenciación.

6. El método de cualquiera de las reivindicaciones 1-5, en donde la relación de prueba y las relaciones de referencia son fracciones cromosómicas, preferiblemente, en donde las fracciones cromosómicas se definen mediante una relación entre una suma de todos los eventos de captura únicos a partir de un cromosoma de interés (S1) y una suma de todos los eventos de captura únicos a partir de todos los cromosomas (S1+S2).

7. Un método de detección sistemática para la aneuploidía en un feto, que comprende:

- a) obtener una muestra de ADN genómico a partir de una muestra de sangre materna;
- b) agregar la muestra de ADN genómico en cada pocillo de una placa de múltiples pocillos, en donde cada pocillo de la placa de múltiples pocillos comprende una mezcla de sonda, en donde la mezcla de sonda comprende una población de sondas de inversión molecular (MIP) y un amortiguador; en donde cada una de las MIP en la población de MIP comprende en secuencia los siguientes componentes: primer brazo de polinucleótidos de focalización, primera etiqueta molecular única, ligador polinucleotídico, segunda etiqueta molecular única, segundo brazo de polinucleótidos de focalización; en donde los pares del primero y del segundo brazo de polinucleótidos de focalización en cada una de las MIP son idénticos, y son sustancialmente complementarios a la primera y a la segunda región en el ácido nucleico que, respectivamente, flanquean cada secuencia en una pluralidad de secuencias diana de interés; en donde la primera y la segunda etiqueta molecular de focalización única en cada una de las MIP en combinación son distintas en cada una de las MIP;
- c) incubar la muestra de ADN genómico con la mezcla de sonda para que las MIP capturen la pluralidad de secuencias diana de interés;
- d) agregar una mezcla de extensión/ligación a la muestra de c) para que las MIP y la pluralidad de secuencias de interés formen una pluralidad de amplicones de las MIP, en donde la mezcla de extensión/ligación comprende una polimerasa, una pluralidad de dNTP, una ligasa y un amortiguador;
- e) agregar una mezcla de exonucleasa a los amplicones de las MIP de focalización y control para eliminar las sondas en exceso o el ADN genómico en exceso;
- f) agregar una mezcla de PCR de indexado a la muestra de e) para agregar un par de cebadores de indexado, un único código de barras de muestra y un par de adaptadores de secuenciación a la pluralidad de amplicones;
- g) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una primera población de amplicones con código de barras que se proporcionan en la etapa f) basado en el número de las etiquetas moleculares de focalización únicas, en donde la primera población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés.
- h) utilizar un método de secuenciación masiva paralela para determinar el número de lecturas de secuenciación de una segunda población de amplicones con código de barras que se proporcionan en la etapa f) basado en el número de las etiquetas moleculares de focalización únicas, en donde la segunda población de amplicones con código de barras se identifica mediante la secuencia de la secuencia diana de interés;
- i) computar una métrica de captura del sitio basado, al menos, en parte en el número de las primeras lecturas de secuenciación que se determinaron en la etapa g) y una pluralidad de métricas de captura de sonda de control basado, al menos, en parte en los números de segundas lecturas de secuenciación que se determinaron en la etapa h);
- j) identificar un subconjunto de métricas de captura del sitio de la población de amplicones de las MIP que tienen métricas de captura de la sonda de control que satisfacen, al menos, un criterio;
- k) normalizar la métrica de captura del sitio mediante un factor que se computa a partir del subconjunto de métricas de captura de sonda de control que satisfacen el, al menos, un criterio, para obtener una métrica de captura del sitio normalizada de prueba;
- l) comparar la métrica de captura del sitio normalizada de prueba con una pluralidad de métricas de captura del sitio normalizadas de referencia que se computan, basado en las muestras de ADN genómico de referencia que se obtienen a partir de los sujetos de referencia que exhiben genotipos conocidos mediante el uso de los mismos sitios específicos y de control, la población específica, el subconjunto de poblaciones de control en las etapas b)-h) para determinar la presencia o ausencia de una métrica de captura del sitio normalizada de prueba que indica la aneuploidía en el feto.

8. El método de cualquiera de las reivindicaciones 1-7, en donde la muestra de sangre es una muestra de sangre

entera, una muestra de plasma, o una muestra de suero, preferiblemente una muestra de plasma.

9. El método de cualquiera de las reivindicaciones 1-8, en donde

- 5 a) la extensión del primero y/o del segundo brazo de polinucleótidos de focalización es entre 14 y 30 pares base, y/o en donde cada uno de los brazos de polinucleótidos de focalización tiene una temperatura de fusión de entre 45 °C y 80 °C; y/o en donde cada uno de los brazos de polinucleótidos de focalización tiene un contenido de GC de entre 30 % y 80 %, preferiblemente entre 30 % y 70 % y/o
- b) en donde la extensión de la primera y/o de la segunda etiqueta molecular única está entre 4 y 15 pares; y/o en donde cada una de las etiquetas moleculares únicas tiene una temperatura de fusión de entre 45 °C y 80 °C; y/o
- 10 c) en donde el ligador polinucleotídico tiene una extensión de entre 20 y 1.000 pares base, y/o en donde el ligador polinucleotídico tiene una temperatura de fusión de entre 45 °C y 80 °C, y/o en donde el ligador polinucleotídico tiene un contenido de GC de entre 30 % y 80 %, preferiblemente entre 30 % y 70 %.

10. El método de cualquiera de las reivindicaciones 1-9, en donde el ligador polinucleotídico comprende al menos un cebador de amplificación, preferiblemente, comprende un cebador de amplificación hacia adelante y un cebador de amplificación inverso.

15

11. El método de la reivindicación 10, en donde:

- a) la secuencia del cebador de amplificación hacia adelante comprende la secuencia de nucleótidos de 5' - CTCAGCTTCCCGATTACGG- 3' (SEC ID NO: 1); y/o
- 20 b) la secuencia del cebador de amplificación inverso comprende la secuencia de nucleótidos de 5'- GCACGATCCGACGGTAGTGT-3' (SEC ID NO: 2).

12. El método de cualquiera de las reivindicaciones 1-11, en donde:

- a) el ligador polinucleotídico comprende la secuencia de nucleótidos de 5'-CTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT-3' (SEC ID NO: 3)

13. El método de cualquiera de las reivindicaciones 1-12, en donde el primer brazo de polinucleótidos de focalización comprende la secuencia de nucleótidos de

25

5'-CACTGCACTCCAGCCTGG - 3' (SEC ID NO: 4); y/o

el segundo brazo de polinucleótidos de focalización comprende la secuencia de 5'- GAGGCTGAGGCAGGAGAA - 3' (SEC ID NO: 5).

14. El método de cualquiera de las reivindicaciones 1-13, en donde la MIP comprende la secuencia de nucleótidos de 5'- CACTGCACTC-CAGCCTGG(N1-6)CTCAGCTTCCCGATTACGGGCACGATCCGACGGTAGTGT(N7-12)GAGGCTGAGGCAG-GAGAA-3' (SEC ID NO: 6), en donde (N1-6) representa la primera etiqueta molecular única y (N7-12) representa la segunda etiqueta molecular única.

30

15. El método de cualquiera de las reivindicaciones 1-14, en donde la etapa de secuenciación tiene una profundidad de lectura de entre 6-8 millones de lecturas.

35

100

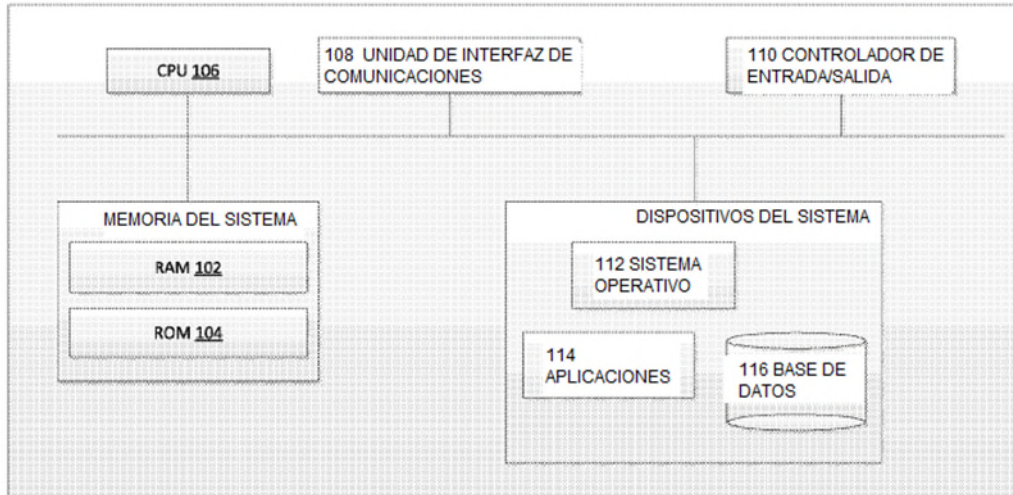


FIG. 1

200

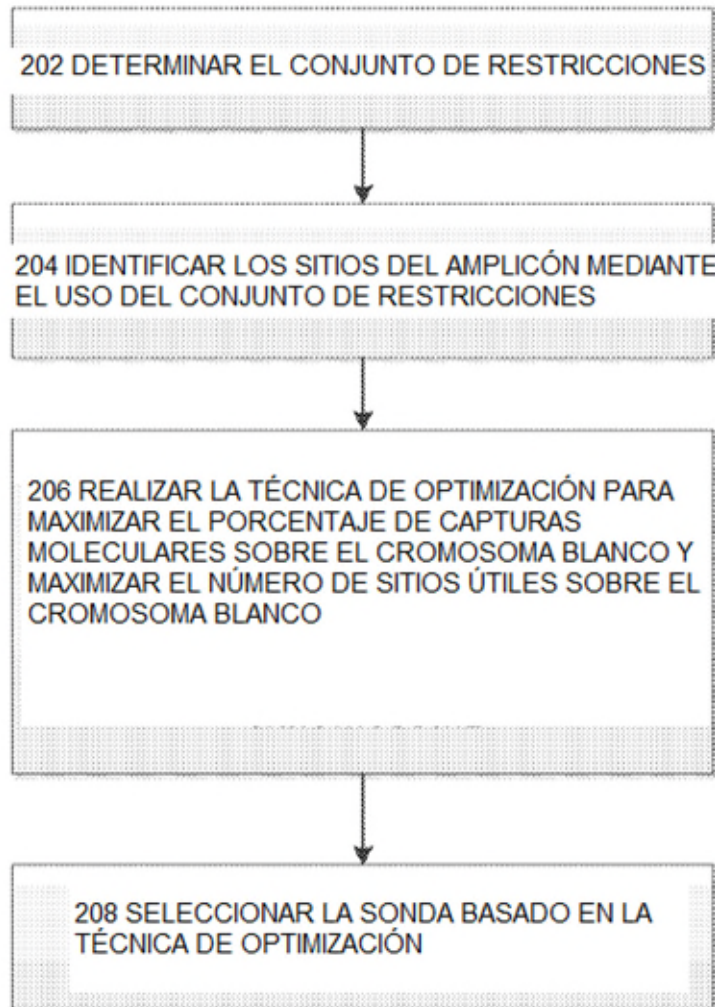


FIG. 2

300

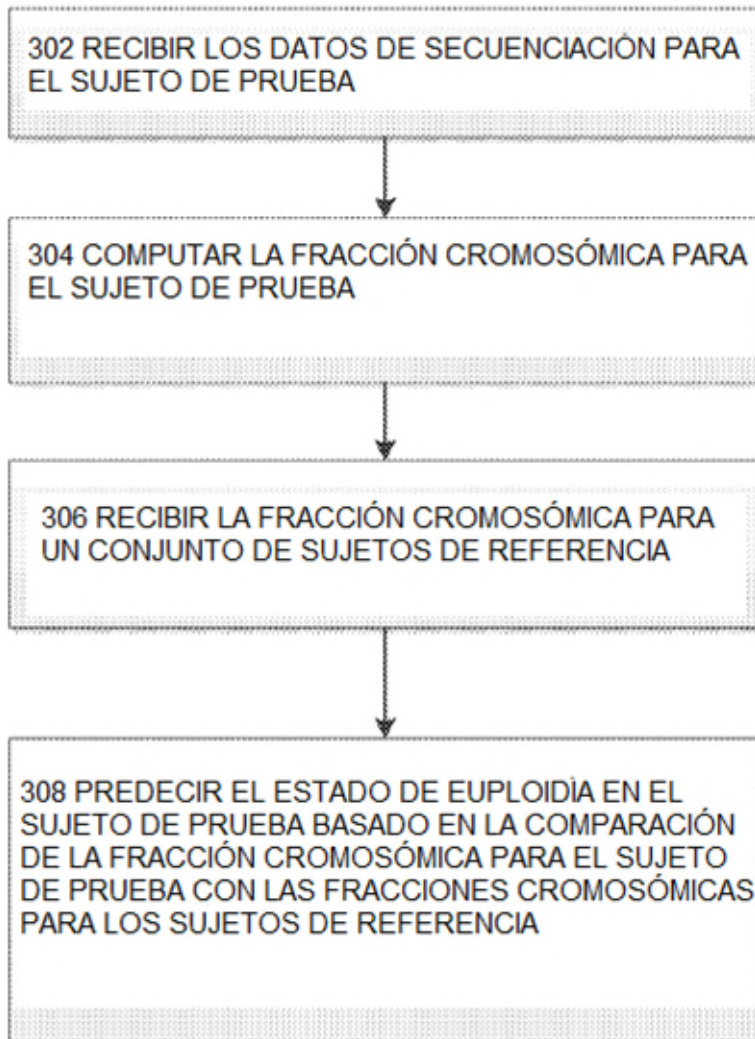


FIG. 3

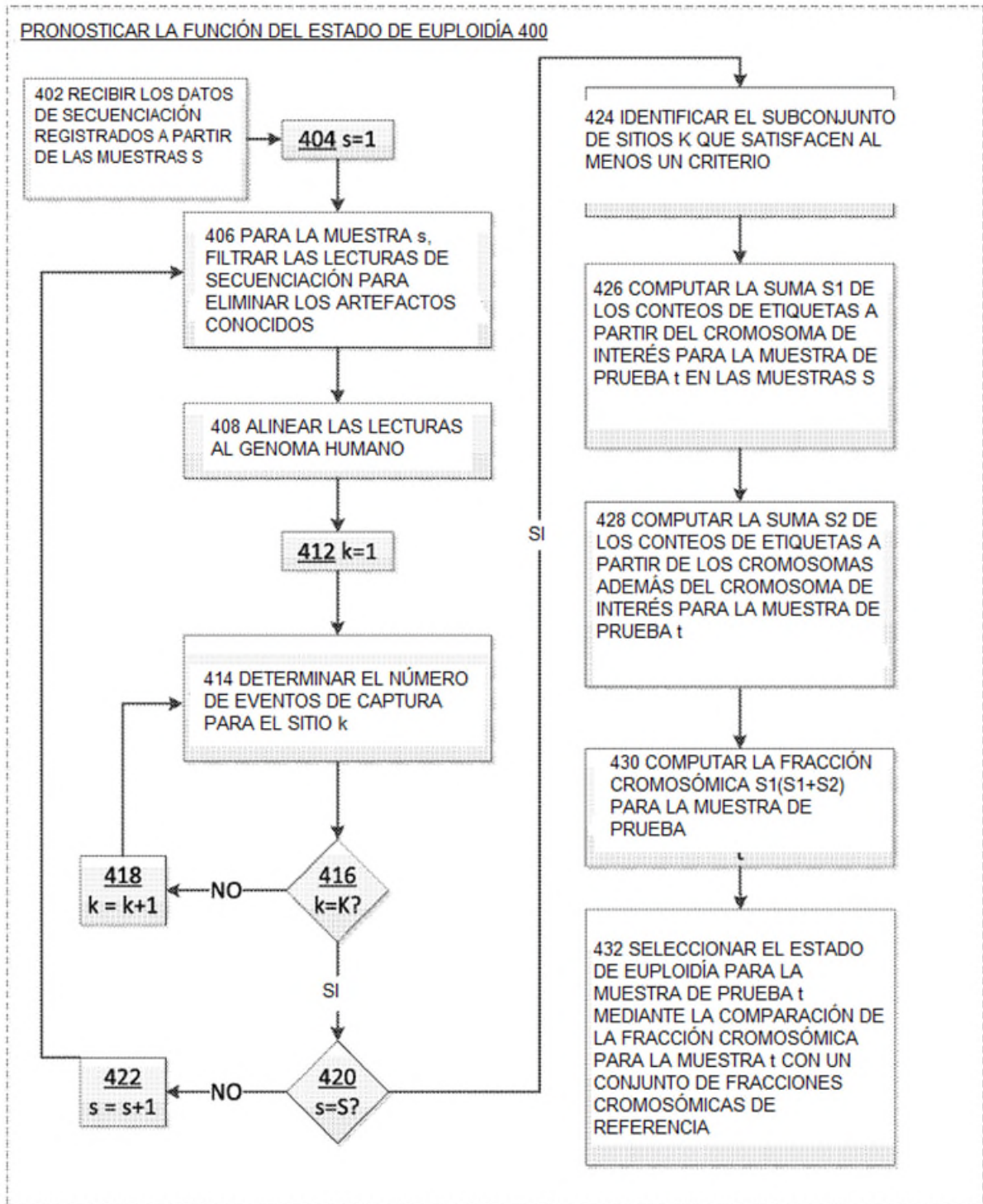


FIG. 4

500

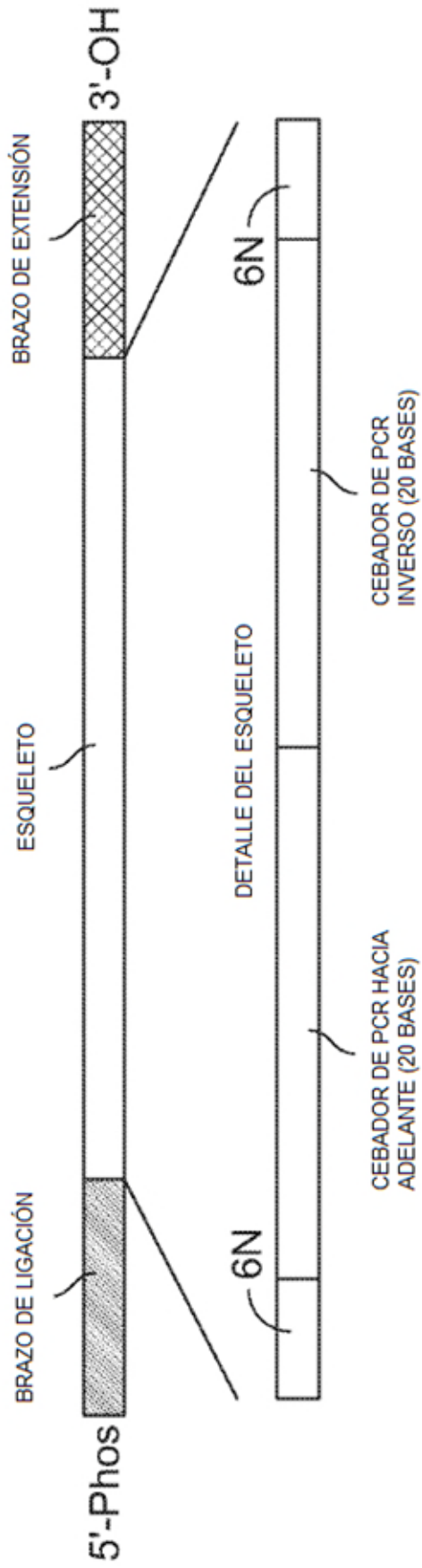


FIG. 5

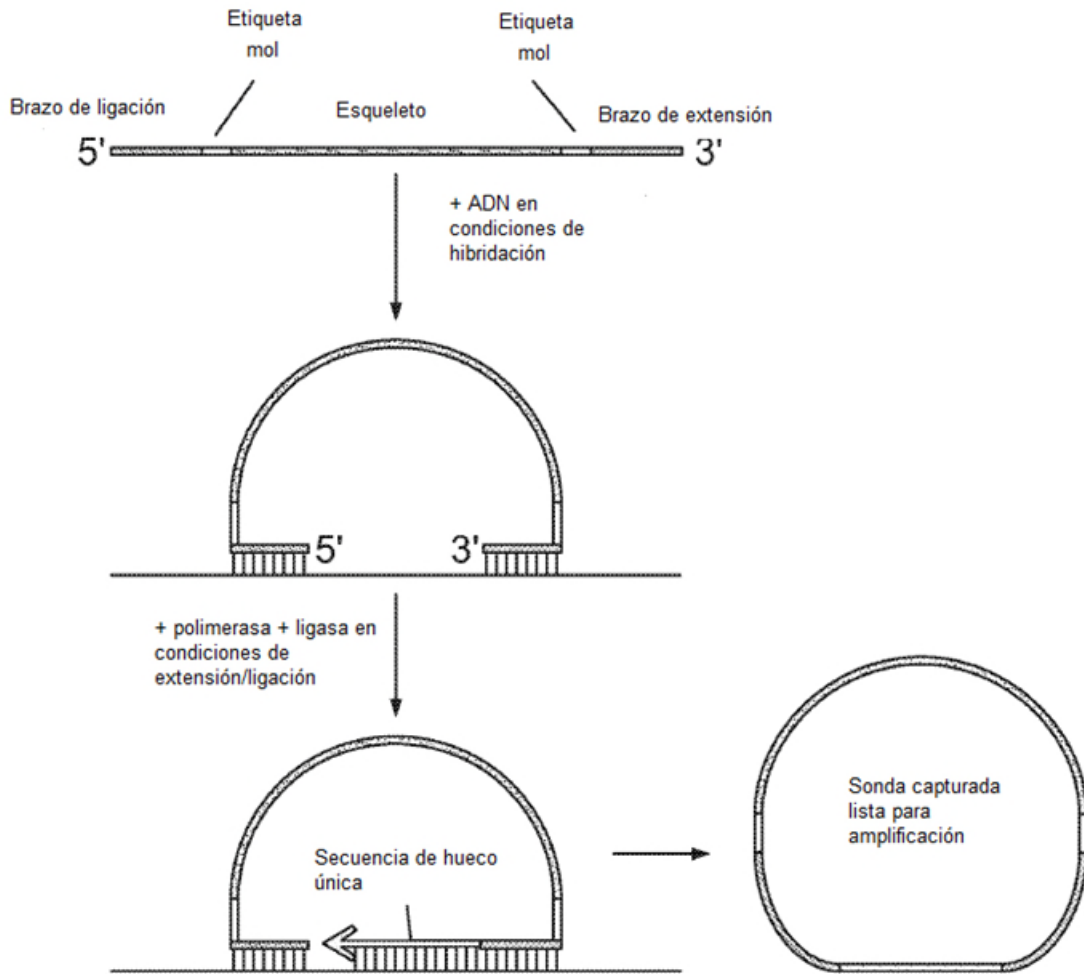


FIG. 6

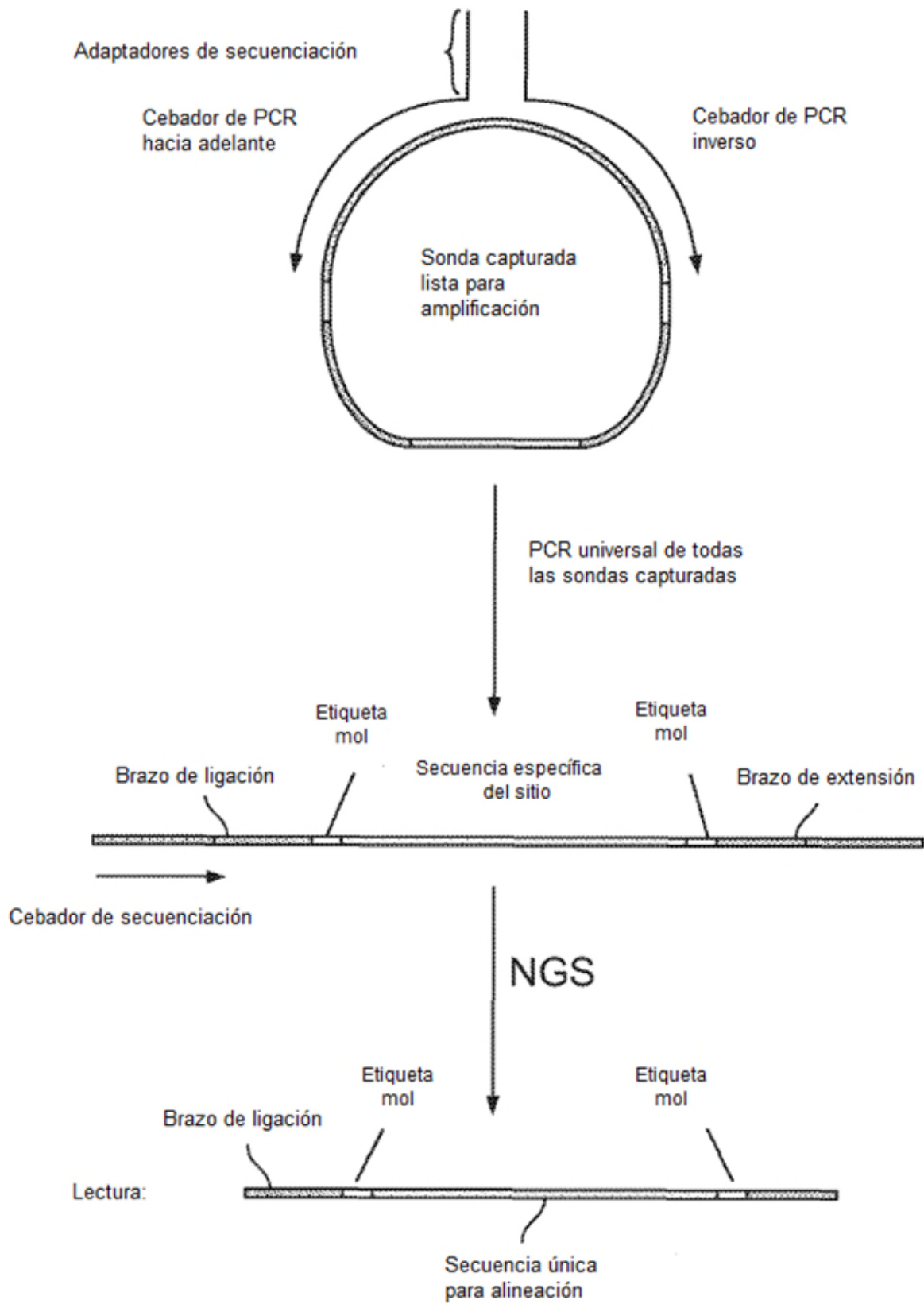


FIG. 7

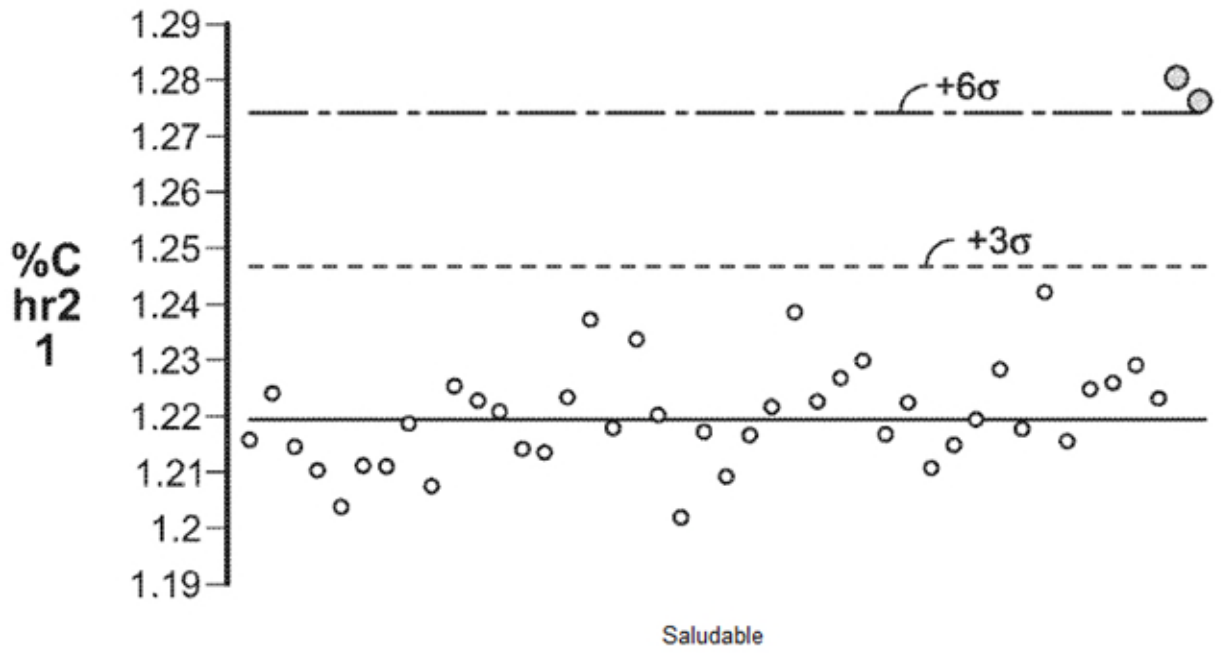


FIG. 8

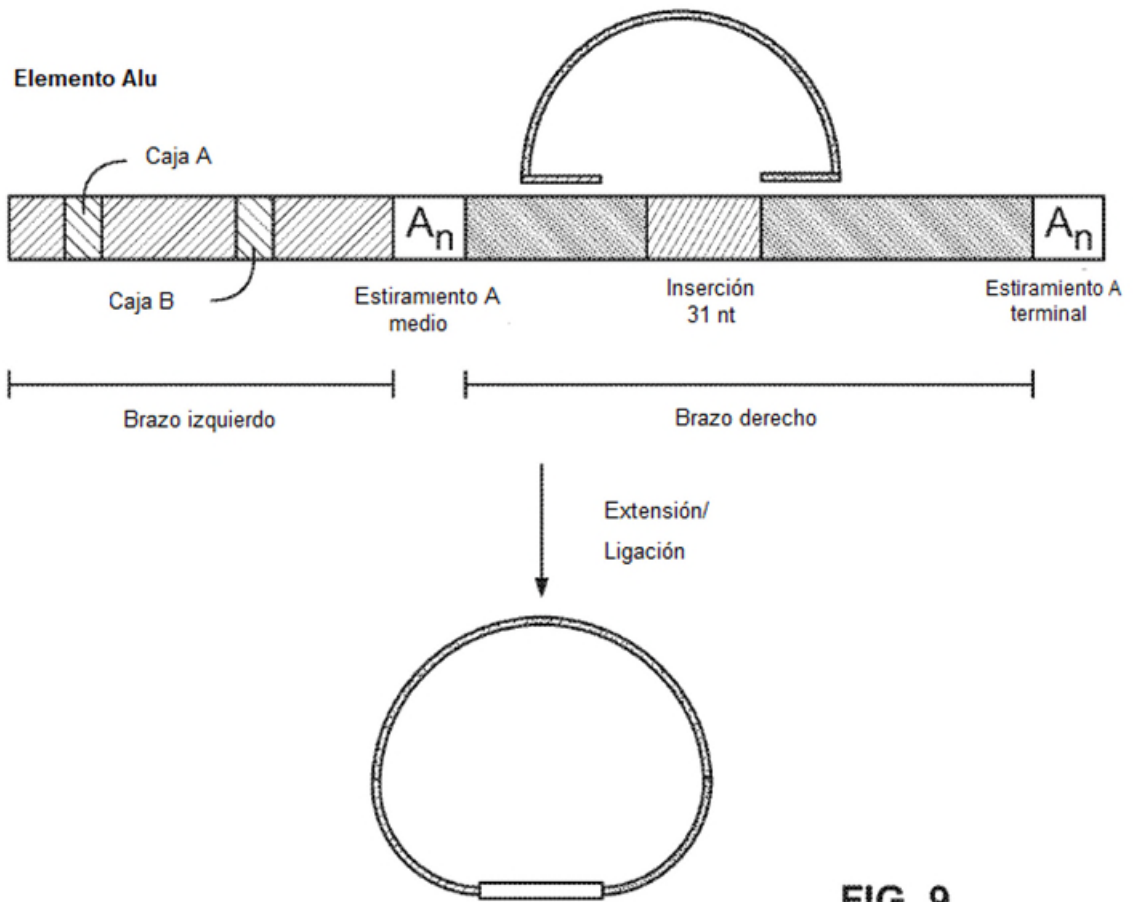
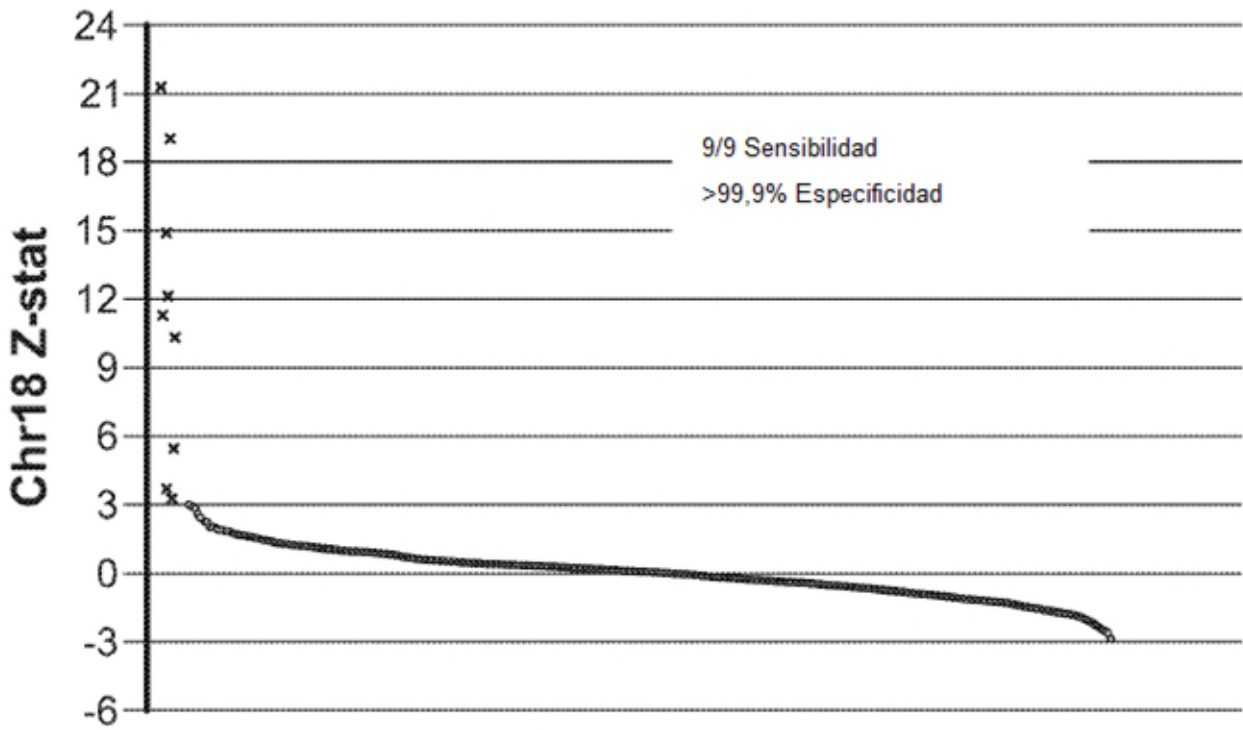
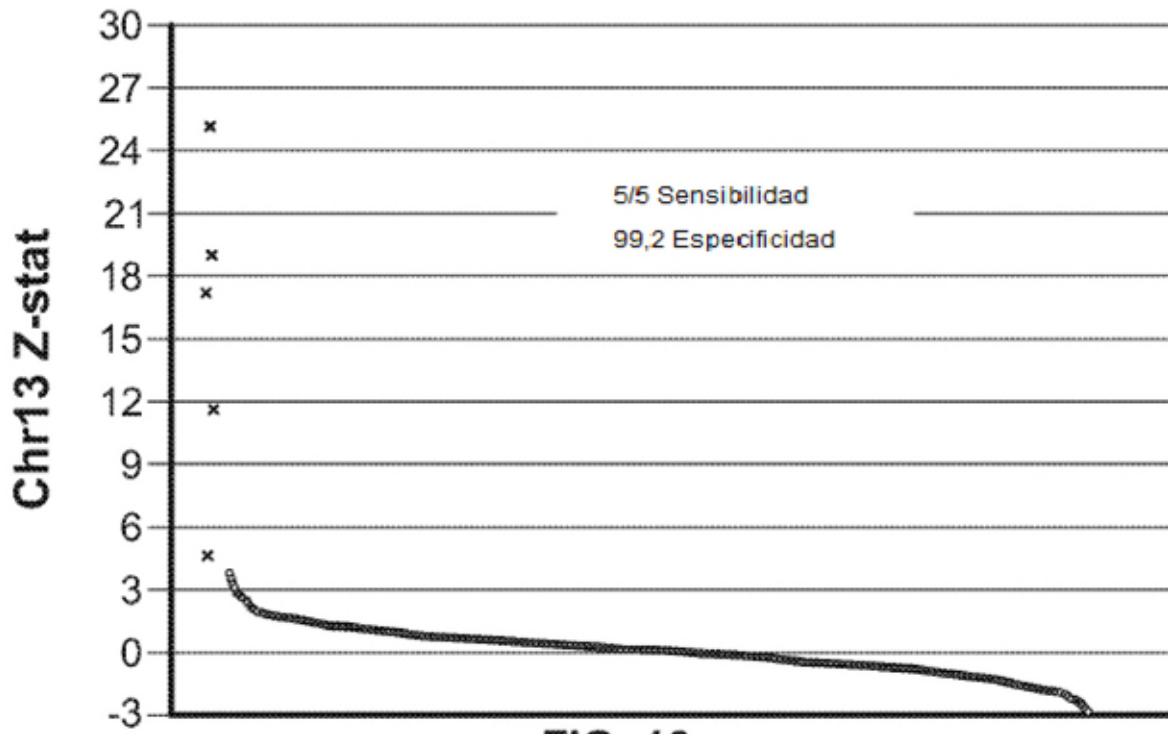


FIG. 9



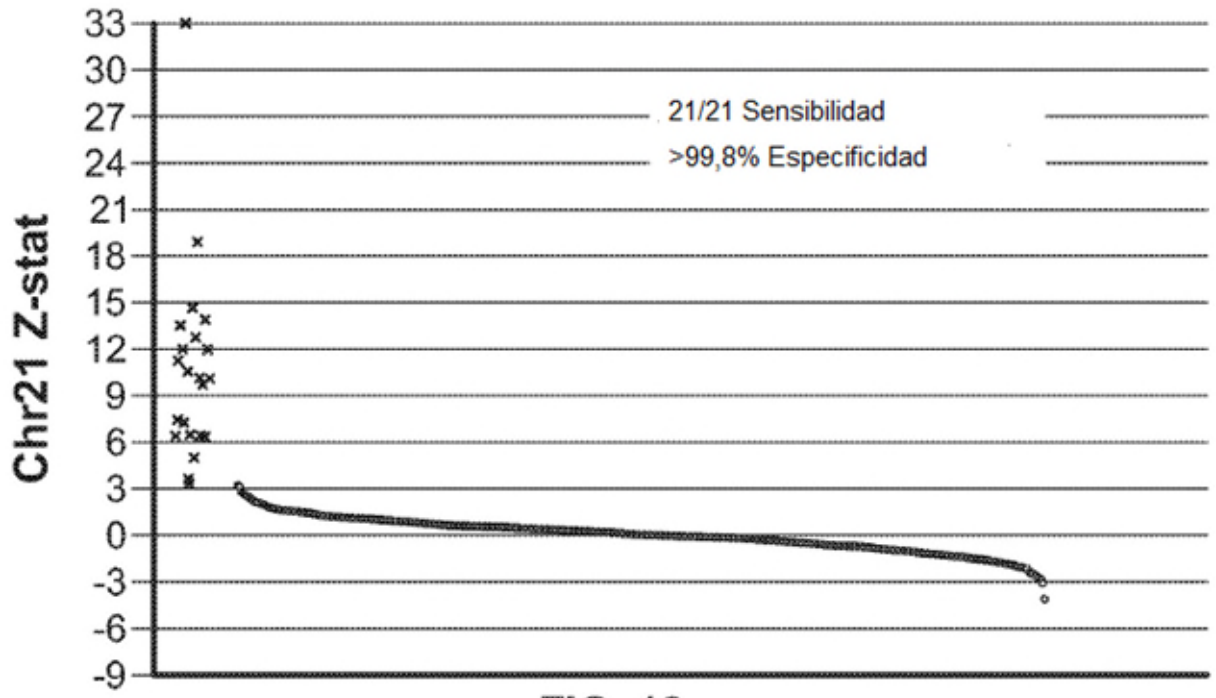


FIG. 12