

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 685 465**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.08.2011 PCT/US2011/046981**

87 Fecha y número de publicación internacional: **09.02.2012 WO12019200**

96 Fecha de presentación y número de la solicitud europea: **08.08.2011 E 11745883 (6)**

97 Fecha y número de publicación de la concesión europea: **30.05.2018 EP 2601311**

54 Título: **Sistemas de ensayo para la determinación de la contribución de fuente en una muestra**

30 Prioridad:

**25.01.2011 US 201113013732  
06.08.2010 US 371605 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**09.10.2018**

73 Titular/es:

**Ariosa Diagnostics, Inc (100.0%)  
5941 Optical Court  
San Jose, CA 95138, US**

72 Inventor/es:

**SPARKS, ANDREW;  
STRUBLE, CRAIG;  
WANG, ERIC y  
OLIPHANT, ARNOLD**

74 Agente/Representante:

**LINAGE GONZÁLEZ, Rafael**

ES 2 685 465 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

Sistemas de ensayo para la determinación de la contribución de fuente en una muestra

### 5 **Campo de la invención**

Esta invención se refiere a métodos de ensayo para identificar una variación en el número de copias en muestras mixtas en un único ensayo.

### 10 **Antecedentes de la invención**

En la siguiente discusión se describirán determinados artículos y métodos con fines de antecedentes e introducción. Nada de lo contenido en el presente documento ha de interpretarse como una "admisión" de la técnica anterior. El solicitante se reserva expresamente el derecho a demostrar, cuando sea adecuado, que los artículos y métodos a los que se hace referencia en el presente documento no constituyen la técnica anterior en virtud de las disposiciones legales aplicables. Los avances recientes en pruebas diagnósticas se han centrado en mecanismos menos invasivos para determinar el riesgo, la presencia y el pronóstico de una enfermedad. Los procedimientos de diagnóstico para determinar anomalías genéticas se han convertido en técnicas convencionales para identificar enfermedades y trastornos específicos, así como proporcionar información valiosa sobre la fuente de la enfermedad y las opciones de tratamiento.

La identificación de ácidos nucleicos libres de células en muestras biológicas tales como sangre y plasma permite que se usen técnicas menos invasivas tales como extracción de sangre en la toma de decisiones clínicas. Por ejemplo, se ha encontrado ADN libre circulante de tumores sólidos malignos en la sangre periférica de pacientes con cáncer; los individuos que se han sometido a trasplante tienen ADN libre circulante del órgano trasplantado presente en su circulación sanguínea; y se han encontrado ARN y ADN fetal libres de células en la sangre y plasma de mujeres embarazadas. Además, la detección de ácidos nucleicos a partir de organismos infecciosos, tales como detección de carga viral o identificación genética de cepas específicas de un patógeno viral o bacteriano, proporciona importantes indicadores de diagnóstico y pronóstico. Por tanto, ácidos nucleicos libres de células de una fuente diferente de las propias células normales del paciente pueden proporcionar importante información médica, por ejemplo, sobre opciones de tratamiento, diagnóstico, pronóstico y similares.

La sensibilidad de tales pruebas depende a menudo de la identificación de la cantidad de ácido nucleico de las diferentes fuentes, y en particular la identificación de un bajo nivel de ácido nucleico de una fuente en el fondo de un nivel más alto de ácidos nucleicos de una segunda fuente. Detectar la contribución de las especies de ácido nucleico menores con respecto a ácidos nucleicos libres de células presentes en la muestra biológica puede proporcionar interpretación estadística exacta de los datos resultantes.

Chiu *et al*, 2008, PNAS, 20458-20463, dan a conocer el análisis de aneuploidía cromosómica fetal mediante secuenciación genómica en paralelo masiva de ADN en plasma materno, que es un método independiente del locus.

Por tanto, existe la necesidad de procedimientos para calcular una variación en el número de copias (VNC) en una o más regiones genómicas en una muestra biológica usando información sobre la contribución de ácidos nucleicos en la muestra. La presente invención aborda esta necesidad.

### 45 **Sumario de la invención**

Este sumario se proporciona para introducir una selección de conceptos en una forma simplificada que se describen adicionalmente a continuación en la descripción detallada. No se pretende que este sumario identifique características clave o esenciales del objeto reivindicado, ni se pretende que se use para limitar el alcance del objeto reivindicado. Otras características, detalles, utilidades y ventajas del objeto reivindicado serán evidentes a partir de la siguiente descripción detallada escrita incluyendo los aspectos ilustrados en los dibujos adjuntos y definidos en las reivindicaciones adjuntas.

Los métodos de la invención comprenden un único sistema de ensayo con la capacidad de determinar la contribución de una fuente principal y/o una fuente secundaria dentro de una muestra de un individuo y determinar la información del número de copias para una o más regiones genómicas dentro de una única fuente para determinar una diferencia de valor para la región genómica en comparación con la contribución de la fuente de los locus en la muestra mixta. La presente invención usa un único sistema de ensayo que usa tanto detección no polimórfica como polimórfica para determinar la contribución de fuente y variaciones en el número de copias (VNC) de una única fuente dentro de la muestra mixta. La determinación de la contribución de la fuente principal y/o secundaria dentro de la muestra puede proporcionar información sobre la presencia de suficiente material genético de ambas fuentes para permitir la identificación adecuada de regiones genómicas para la determinación de VNC en las muestras mixtas.

En un aspecto, el sistema de ensayo usa amplificación y detección de locus seleccionados en una muestra mixta de

un individuo para calcular la contribución de fuente e identificar el número de copias de una o más regiones genómicas. En un aspecto específico, la invención proporciona únicos sistemas de ensayo con la capacidad de determinar la contribución de ácidos nucleicos de una fuente principal y/o secundaria en la muestra mixta y la presencia o ausencia de VNC en una o más regiones genómicas de una única fuente en una muestra mixta. El sistema de ensayo puede detectar específicamente el número de copias de regiones genómicas presentes en dos o más fuentes dentro de la muestra mixta. Para la determinación de la contribución, los locus seleccionados de una fuente se distinguen de los locus seleccionados de al menos una otra fuente en la muestra mixta. Para la determinación de la variación en el número de copias de una región genómica, los locus seleccionados pueden, pero no necesitan, distinguirse en cuanto a contribución de fuente, puesto que puede detectarse la variación en el número de copias mediante comparación de los niveles de dos o más regiones genómicas dentro de una muestra mixta. Preferiblemente, los ácidos nucleicos libres de células analizados en el sistema de ensayo son ADN libre circulante (ADNlc).

Por tanto, en una primera implementación, la invención proporciona un único sistema de ensayo para 1) determinar la contribución de una fuente principal y/o una fuente secundaria en una muestra mixta usando datos de frecuencia derivados de dos o más locus informativos; 2) determinar la frecuencia de una o más regiones genómicas en la fuente principal y secundaria; y 3) identificar la presencia o ausencia de una VNC para una o más regiones genómicas en la fuente principal y/o secundaria. Preferiblemente, la identificación de VNC se basa en la comparación del número de copias de dos o más regiones genómicas de la fuente principal y/o secundaria en la muestra mixta.

Preferiblemente, los ácidos nucleicos analizados usando los sistemas de la invención son ácidos nucleicos libres de células. Más preferiblemente, los ácidos nucleicos analizados en el sistema de ensayo comprenden ADN libre circulante (ADNlc).

En otro aspecto específico, la invención proporciona un sistema de ensayo para el cálculo de la contribución de fuente y detección de la presencia o ausencia de VNC en una o más regiones genómicas dentro de una muestra mixta, comprendiendo el ensayo las etapas de introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en uno o más locus en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus informativos; ligar los oligonucleótidos hibridados para crear productos de ligación contiguos complementarios a los locus seleccionados; amplificar los productos de ligación contiguos para crear productos de amplificación; y detectar los productos de amplificación. La detección de los productos de amplificación se usa para calcular la contribución de fuente y el número de copias de una o más regiones genómicas en la muestra mixta.

Los conjuntos de oligonucleótidos fijos comprenden dos o más oligonucleótidos que se hibridan con regiones contiguas de las regiones genómicas o los locus informativos. En algunos aspectos preferidos, se interrogan conjuntos de locus para VNC, y son indicativos de una amplificación de una región genómica más grande, por ejemplo, todo o parte de un cromosoma. Preferiblemente, los sistemas de ensayo pueden distinguir el número de copias de estos locus entre una fuente principal y una fuente secundaria dentro de una muestra mixta. Los niveles de locus seleccionados pueden determinarse para una región genómica de interés y compararse con las cantidades de locus de una o más de otras regiones genómicas de interés y/o una o más regiones genómicas de referencia para detectar posibles VNC basándose en frecuencias de locus en la muestra mixta.

La detección de anomalía cromosómica en una muestra puede basarse en la detección de VNC para múltiples locus seleccionados ubicados en o asociados con un único cromosoma de una fuente secundaria y/o principal. Por tanto, en otro aspecto específico, la invención proporciona un único método de ensayo para 1) determinar la contribución de una fuente principal y/o una fuente secundaria en una muestra mixta usando datos de frecuencia derivados de dos o más locus informativos; 2) determinar la frecuencia de una o más regiones genómicas en la fuente principal y secundaria; y 3) identificar la presencia o ausencia de una anomalía cromosómica en la fuente principal y/o secundaria en la muestra mixta.

Por tanto, la invención proporciona un método de ensayo para el cálculo de la contribución de fuente y detección de la presencia o ausencia de una anomalía cromosómica en una muestra mixta usando un único ensayo, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un primer cromosoma; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un segundo cromosoma; introducir un tercer conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus informativos; ligar los oligonucleótidos hibridados para crear productos de ligación contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligación contiguos para crear productos de amplificación; y detectar los productos de amplificación. La detección de productos de amplificación

puede usarse para el cálculo de la contribución de fuente así como identificación de anomalías cromosómicas en la muestra mixta.

5 En un aspecto más específico, el método de ensayo de la invención se usa para identificar el porcentaje de contribución fetal y anomalías cromosómicas en una muestra materna. La invención proporciona un método de ensayo que comprende las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra materna en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un primer cromosoma; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra materna en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un segundo cromosoma; introducir un tercer conjunto de oligonucleótidos de secuencia fija en la muestra materna en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus informativos; ligar los oligonucleótidos hibridados para crear productos de ligación contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligación contiguos para crear productos de amplificación; y detectar los productos de amplificación. La detección de productos de amplificación puede usarse para el cálculo de contribución fetal así como identificación de anomalías cromosómicas en los ácidos nucleicos fetales en la muestra materna.

20 En determinados aspectos de la invención, los oligonucleótidos fijos se hibridan inmediatamente adyacentes en la región contigua, de modo que se ligan directamente durante la etapa de ligación del ensayo. En otros aspectos, sin embargo, puede haber un hueco de uno o más nucleótidos entre los extremos de los oligonucleótidos fijos tras la hibridación en la región contigua. Los oligonucleótidos fijos se unen mediante una combinación de, por ejemplo, extensión de cebador usando una polimerasa y ligación.

25 Cada conjunto de ácidos nucleicos de secuencia fija se diseña para hibridarse con al menos dos regiones diferentes en un locus seleccionado. En aspectos preferidos, se usan dos o más oligonucleótidos diferentes en un conjunto para hibridarse con estas regiones para proporcionar ácidos nucleicos adyacentes complementarios a los locus seleccionados. En algunos aspectos, sin embargo, un conjunto puede comprender una única sonda con dos o más regiones no adyacentes distintas que son complementarias a los locus seleccionados (por ejemplo, sondas candado), tal como se describe en más detalle en el presente documento. Los conjuntos de oligonucleótidos de secuencia fija pueden proporcionarse en el ensayo secuencial o simultáneamente en el ensayo.

35 En determinados aspectos preferidos, se usan oligonucleótidos en puente para aumentar la especificidad de los conjuntos de oligonucleótido y / o llenar un hueco entre oligonucleótidos de secuencia fija. Por consiguiente, otro aspecto específico de la invención proporciona un método de ensayo para el cálculo de la contribución de fuente y detección de la presencia o ausencia de VNC en una o más regiones genómicas dentro de una muestra mixta, comprendiendo el ensayo las etapas de introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en uno o más locus en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en al menos un locus informativos; introducir uno o más oligonucleótidos en puente en condiciones que permiten que los oligonucleótidos en puente se hibriden específicamente con regiones complementarias en los locus, en el que uno o más oligonucleótidos en puente son complementarios a una región de los locus entre e inmediatamente adyacentes a las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto; ligar los oligonucleótidos hibridados para crear productos de ligación contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligación contiguos para crear productos de amplificación; y detectar los productos de amplificación. Se usa detección de los productos de amplificación para calcular la contribución de fuente y el número de copias de una o más regiones genómicas en la muestra mixta.

50 Otro aspecto específico de la invención proporciona un método de ensayo usando oligonucleótidos en puente para calcular la contribución de fuente e identificar anomalías cromosómicas en una muestra mixta. Este método comprende las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un primer cromosoma; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus correspondientes a un segundo cromosoma; introducir un tercer conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en dos o más locus informativos; introducir uno o más oligonucleótidos en puente en condiciones que permiten que los oligonucleótidos en puente se hibriden específicamente con regiones complementarias en los locus, en el que uno o más oligonucleótidos en puente son complementarios a una región de los locus entre e inmediatamente adyacentes a las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto; introducir uno o más oligonucleótidos en puente en condiciones que permiten que los oligonucleótidos en puente se hibriden específicamente con regiones complementarias en los locus, en el que uno o más oligonucleótidos en puente son complementarios a una región de los locus entre e inmediatamente adyacentes a las regiones complementarias a

los oligonucleótidos de secuencia fija de cada conjunto; ligar los oligonucleótidos hibridados para crear productos de ligación contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligación contiguos para crear productos de amplificación; y detectar los productos de amplificación. La detección de productos de amplificación puede usarse para el cálculo de la contribución de fuente (tal como contribución fetal en una muestra materna), así como identificación de anomalías cromosómicas en la muestra mixta.

En determinados aspectos de la invención, los oligonucleótidos fijos se hibridan inmediatamente adyacentes a los oligonucleótidos en puente, de modo que todos se ligan directamente durante la etapa de ligación del ensayo. En otros aspectos, sin embargo, puede haber un hueco de uno o más nucleótidos entre el extremo de uno o ambos de los oligonucleótidos fijos y el oligonucleótido en puente tras la hibridación del oligonucleótido en puente. Los oligonucleótidos fijos y el oligonucleótido en puente se unen mediante una combinación de, por ejemplo, extensión de cebador usando una polimerasa y ligación.

Una característica importante es que los ensayos multiplexados de la invención permiten el análisis de 5 o más, preferiblemente 10 o más, preferiblemente 16 o más, preferiblemente 20 o más, preferiblemente 30 o más, preferiblemente 32 o más, preferiblemente 40 o más, preferiblemente 48 o más, preferiblemente 50 o más, preferiblemente 60 o más, preferiblemente 70 o más, preferiblemente 80 o más, preferiblemente 90 o más, y más preferiblemente 96 o más locus seleccionados simultáneamente. Estos locus seleccionados pueden ser locus diferentes de una única muestra, o pueden ser locus de dos o más muestras mixtas. En el último caso, al menos uno de los dos oligonucleótidos de secuencia fija usados para análisis de un locus seleccionado comprenderá un identificador de muestra (por ejemplo, un "índice de muestra") que permitirá que el locus se asocie con una muestra particular. Alternativamente, un índice de muestra puede añadirse durante la amplificación del producto de ligación usando un cebador que comprende el índice de muestra.

En aspectos preferidos, la interrogación de estos locus usa técnicas de amplificación universales que permiten la amplificación de múltiples locus en una única reacción de amplificación. Los ácidos nucleicos seleccionados para el cálculo de contribución y detección de VNC y/o anomalías cromosómicas en el sistema de ensayo de la invención pueden amplificarse usando métodos de amplificación universales tras la amplificación selectiva inicial de la muestra mixta. El uso de amplificación universal permite que múltiples de regiones de ácidos nucleicos de una única o múltiples muestras se amplifiquen usando un único o limitado número de cebadores de amplificación, y es especialmente útil en la amplificación de múltiples regiones seleccionadas en una única reacción. En un aspecto preferido, las regiones de cebador universal se usan en la determinación de secuencia de los productos de amplificación. En otro aspecto preferido, se usan las mismas regiones de cebador universal en los oligonucleótidos de secuencia fija usados para la detección de regiones genómicas y los oligonucleótidos de secuencia fija usados para la detección de polimorfismos.

Por tanto, en un aspecto específico de la invención, se introducen secuencias complementarias a cebadores para su uso en amplificación universal en los locus seleccionados durante o tras amplificación selectiva. Preferiblemente tales secuencias se introducen en los extremos de tales ácidos nucleicos seleccionados, aunque pueden introducirse en cualquier ubicación que permita la identificación del producto de amplificación a partir del procedimiento de amplificación universal.

En determinados aspectos preferidos, uno o ambos de los cebadores usados comprenden un índice de muestra u otro identificador. En un aspecto específico, se incluye un índice de muestra en uno o más de los cebadores universales. El índice de muestra se incorpora en los productos de amplificación, y entonces pueden combinarse productos de amplificación de diferentes muestras. El índice de muestra se detecta simultáneamente con la detección de la VNC o anomalía cromosómica y la detección de polimorfismo de modo que la VNC y el polimorfismo pueden asignarse adecuadamente a la muestra de origen.

En determinados aspectos, el sistema de ensayo multiplexa la interrogación de locus usando uno o más oligonucleótidos en puente comunes que son complementarios a regiones en dos o más locus interrogados, es decir un único oligonucleótido en puente puede usarse para dos o más conjuntos de oligonucleótidos fijos. Esto permite que el número de oligonucleótidos en puente usado en el sistema de ensayo multiplexado sea menos que el número de locus interrogados en el ensayo. En determinados aspectos específicos, el sistema de ensayo usa una agrupación de oligonucleótidos en puente que cada uno está diseñado para ser compatible con dos o más locus interrogados usando el sistema de ensayo de la invención.

Pueden determinarse las frecuencias de locus seleccionados para una región genómica de interés y compararse con las frecuencias de locus de una o más de otras regiones genómicas de interés y/o una o más regiones genómicas de referencia para detectar posibles VNC basándose en frecuencias de locus en la muestra mixta.

En algunos casos, la anomalía cromosómica se asocia con amplificación génica o expansión de locus en un cromosoma de interés. En otros casos, la anomalía cromosómica se asocia con una translocación que da como resultado la presencia de una porción extra de un cromosoma en el genoma. En aún otros casos, la anomalía cromosómica es una deleción.

En determinados aspectos preferidos, la anomalía cromosómica se asocia con aneuploidía de un cromosoma de interés. Por ejemplo, las anomalías cromosómicas más comunes en el feto son trisomía 21, 18, 13, X y/o Y o monosomía X. En aspectos preferidos específicos, los sistemas de ensayo de la invención se usan para detectar tales aneuploidías cromosómicas comunes en el ADN fetal de una muestra materna.

5 En el método de ensayo de la invención, los productos de amplificación se aíslan opcionalmente antes de la detección. Cuando se aíslan, se aíslan preferiblemente como moléculas individuales para ayudar en la posterior detección. Tras el aislamiento, los productos de amplificación pueden amplificarse adicionalmente para crear copias idénticas de todos o una porción de los productos de amplificación individuales antes de la detección.  
10 Alternativamente, los productos aislados de amplificación pueden amplificarse adicionalmente para crear copias idénticas de moléculas complementarias a todos o una porción de los productos de amplificación individuales antes de la detección.

15 Pueden emplearse diversos métodos de detección de VNC junto con la detección de los polimorfismos en el método de ensayo de la invención. En un aspecto general, el sistema de ensayo emplea un método para la determinación de una VNC en uno o más locus en una muestra mixta, que comprende las etapas de amplificar uno o más ácidos nucleicos seleccionados de una primera región genómica de interés en una muestra mixta; amplificar uno o más ácidos nucleicos seleccionados de un segundo locus de interés en la muestra mixta, determinar la frecuencia relativa de los locus seleccionados, comparar la frecuencia relativa de los locus seleccionados, e identificar la presencia o  
20 ausencia de una VNC basándose en las frecuencias relativas comparativas de los ácidos nucleicos seleccionados de los locus primero y segundo. Preferiblemente, el método de ensayo amplifica dos o más locus seleccionados de diferentes regiones genómicas, aunque los locus pueden ubicarse en la misma región genómica general para la confirmación de VNC que surgen de anomalías cromosómicas distintas de VNC de un único locus.

25 Más preferiblemente, los oligonucleótidos de secuencia fija no hibridados se retiran antes de la introducción de los oligonucleótidos en puente. En algunos aspectos, los oligonucleótidos en puente se introducen simultáneamente con la mezcla de ligación. En otros aspectos, los productos de hibridación de los oligonucleótidos de secuencia fija y los locus se aíslan antes de la introducción de los oligonucleótidos en puente.

30 En determinados aspectos específicos, el sistema de ensayo usa una agrupación de oligonucleótidos en puente que están diseñados cada uno para ser compatibles con dos o más locus interrogados usando el sistema de ensayo de la invención. En estos aspectos, los oligonucleótidos en puente usados en el ensayo multiplexado están diseñados preferiblemente para tener una  $T_f$  en un intervalo de  $\pm 5^\circ\text{C}$ , más preferiblemente en un intervalo de  $\pm 2^\circ\text{C}$ .

35 En determinados aspectos, los oligonucleótidos en puente están entre 2-45 nucleótidos de longitud. En un aspecto específico, los oligonucleótidos en puente están entre 3-9 nucleótidos de longitud. En aún otro aspecto específico, los oligonucleótidos están entre 10-30 nucleótidos de longitud.

40 Los locus interrogados para VNC pueden, en algunos casos, ser indicativos de una amplificación de una región genómica más grande, por ejemplo, todo o parte de un cromosoma. Preferiblemente, los sistemas de ensayo pueden distinguir el número de copias de estos locus entre una fuente principal y una fuente secundaria dentro de una muestra mixta.

45 En otro aspecto, la presente invención usa técnicas que permiten la identificación de tanto VNC como agentes infecciosos en una muestra mixta. Esto puede ser especialmente útil para monitorizar pacientes en los que el desenlace clínico puede ponerse en peligro por la presencia de un agente infeccioso. Por ejemplo, es probable que un paciente que se ha sometido a un trasplante esté tomando medicación inmunodepresora, y por tanto sea más propenso a infección en general. De manera similar, las mujeres embarazadas tienen cambios en su sistema inmunitario y, por tanto, pueden ser más susceptibles a infección con patógenos que pueden tener un efecto  
50 negativo en la madre y/o el feto. Además, determinados tipos de cáncer están asociados con agentes infecciosos (por ejemplo, cáncer de hígado asociado con infecciones de hepatitis B y C, cáncer de cuello uterino asociado con infección con virus del papiloma humano), y la identificación de los agentes infecciosos puede ser informativa en la predicción del desenlace clínico o la determinación del curso preferido de tratamiento médico para el paciente.

55 Por tanto, en determinados aspectos, la invención proporciona un método de ensayo para el cálculo de contribución de fuente, detección de la presencia o ausencia de VNC de una región genómica, y la presencia o ausencia de un agente infeccioso en una muestra mixta usando un único ensayo, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en uno o más locus en o asociados  
60 con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos fijos se hibriden específicamente con regiones complementarias en al menos un locus informativo; introducir un tercer conjunto de oligonucleótidos de secuencia fija en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en locus indicativos de un agente infeccioso; ligar los oligonucleótidos hibridados para crear un producto de ligación contiguo complementario a los locus; amplificar el  
65 producto de ligación contiguo para crear productos de amplificación; y detectar los productos de amplificación. La

detección de los productos de amplificación se correlaciona con el número de copias de la región genómica y la presencia o ausencia de un agente infeccioso en la muestra mixta.

5 En otro aspecto general, el sistema de ensayo emplea un método para determinar la presencia o ausencia de una anomalía cromosómica asociada con VNC en una región genómica, que comprende las etapas de amplificar uno o más locus seleccionados de un primer cromosoma de interés en una muestra mixta; amplificar uno o más locus seleccionados de un segundo cromosoma de interés en la muestra mixta, determinar la frecuencia relativa de las regiones seleccionadas de los cromosomas de interés primero y segundo, comparar la frecuencia relativa de las regiones seleccionadas de los cromosomas de interés primero y segundo, e identificar la presencia o ausencia de  
10 una anomalía basándose en las frecuencias relativas comparativas de las regiones seleccionadas. Preferiblemente, se seleccionan dos o más regiones de ácidos nucleicos de cada cromosoma, y más preferiblemente se seleccionan cinco o más locus de cada cromosoma.

15 En aún otro aspecto general, el sistema de ensayo emplea un método para la determinación de la presencia o ausencia de una aneuploidía en una muestra mixta de un individuo, que comprende las etapas de amplificar dos o más locus seleccionados en el ADNlc correspondiente a un primer cromosoma de interés en una muestra mixta; amplificar dos o más locus seleccionados en el ADNlc correspondiente a un segundo cromosoma de interés en la muestra mixta, determinar la frecuencia relativa de las regiones seleccionadas de los cromosomas de interés primero y segundo, comparar la frecuencia relativa de las regiones seleccionadas de los cromosomas de interés  
20 primero y segundo, e identificar la presencia o ausencia de una aneuploidía basándose en las frecuencias relativas comparativas de las regiones seleccionadas. En un aspecto específico, los locus de los cromosomas primero y segundo se amplifican en una única reacción, y preferiblemente en una única reacción contenida dentro de un único recipiente.

25 Preferiblemente, el sistema de ensayo detecta la presencia o ausencia de locus en muestras que pueden obtenerse fácilmente de un sujeto, tal como sangre, plasma, suero y similares. En un aspecto general, el sistema de ensayo usa detección de regiones seleccionadas en ADNlc en una muestra mixta. En un aspecto más específico, el sistema de ensayo usa detección de regiones seleccionadas en ADNlc de una muestra mixta de un individuo para identificar la presencia o ausencia de VNC en una región genómica y la presencia o ausencia de un polimorfismo en uno o más  
30 locus. Puede determinarse el número de copias dentro de una región genómica basándose en la detección de cantidades de locus seleccionados y comparación con las cantidades de locus seleccionados de otra región genómica y/o con las cantidades de locus seleccionados de una región genómica de referencia. En un aspecto particular, la razón de las frecuencias del ácido nucleico se comparan con una razón media de referencia que se ha determinado para una población estadísticamente significativa de sujetos genéticamente "normales", es decir sujetos  
35 que no tienen una VNC asociada con los locus interrogados particulares en el sistema de ensayo.

En un aspecto preferido de la invención, los productos de amplificación correspondientes a los ácidos nucleicos seleccionados se aíslan como moléculas individuales para análisis de los locus seleccionados. Estos productos de amplificación individuales se aíslan entre sí, y preferiblemente se aíslan físicamente (por ejemplo, sobre un sustrato  
40 o en recipientes individuales). Las moléculas individuales pueden amplificarse adicionalmente tras el aislamiento para hacer copias idénticas, múltiples del producto de amplificación, una porción del mismo, o un ácido nucleico complementario al producto de amplificación o una porción del mismo.

En un aspecto preferido, los productos de amplificación individuales se analizan mediante determinación de  
45 secuencia. En otros aspectos, los productos de amplificación individuales se analizan usando técnicas de hibridación.

Una característica de la presente invención es que el número de copias de los locus seleccionados puede detectarse usando métodos de detección no polimórfica, es decir, métodos de detección que no dependen de la presencia o  
50 ausencia de un polimorfismo particular para identificar la región de ácido nucleico seleccionada. En un aspecto preferido, los sistemas de detección de ensayo usan métodos de detección no polimórfica para "contar" los números relativos de locus seleccionados presentes en una muestra mixta. Estos números pueden usarse para determinar si, estadísticamente, es probable que una muestra mixta tenga una VNC en una región genómica en una fuente principal y/o secundaria dentro de la muestra mixta. De manera similar, estos números pueden usarse para  
55 determinar, si estadísticamente, ácidos nucleicos de la fuente principal y/o fuente secundaria tienen uno o más polimorfismos. Tal información puede usarse para identificar un trastorno genético o patología particular, para confirmar un diagnóstico o recaída de una enfermedad o trastorno, para determinar el pronóstico de una enfermedad o trastorno, para ayudar a determinar posibles opciones de tratamiento, etc.

60 En algunos aspectos, los métodos para la determinación de aneuploidía usados por el sistema de ensayo miden la variación en el número de copias de múltiples locus seleccionados de dos o más cromosomas en una muestra. Los niveles de los diferentes locus seleccionados correspondientes a cromosomas específicos pueden cuantificarse y compararse individualmente para determinar la presencia o ausencia de una aneuploidía cromosómica en una o  
65 más fuentes de células en una muestra mixta. Las regiones cuantificadas individualmente pueden someterse a un cálculo de normalización o los datos pueden someterse a exclusión atípica antes de la comparación para determinar la presencia o ausencia de una aneuploidía en una muestra mixta.

En otros aspectos, las frecuencias relativas de los locus seleccionados se usan para determinar una frecuencia del cromosoma de los cromosomas primero y segundo de interés, y la presencia o ausencia de una aneuploidía se basa en las frecuencias del cromosoma comparadas de los cromosomas primero y segundo de interés.

5 En aún otros aspectos, las frecuencias relativas de los locus seleccionados se usan para determinar una frecuencia del cromosoma de un cromosoma de interés y un cromosoma de referencia, y la presencia o ausencia de una aneuploidía se basa en las frecuencias del cromosoma comparadas del cromosoma de interés y el cromosoma de referencia.

10 Puesto que el método de ensayo de la invención está configurado preferiblemente como un sistema muy multiplexado, pueden analizarse múltiples locus de un único cromosoma o múltiples dentro de una muestra de un individuo y/o múltiples muestras simultáneamente. En tales sistemas multiplexados, las muestras pueden analizarse por separado, o pueden agruparse inicialmente en grupos de dos o más para el análisis de números más grandes de muestras. Cuando se obtienen datos agrupados, tales datos se identifican preferiblemente para las diferentes muestras antes del análisis de aneuploidía. En algunos aspectos, sin embargo, los datos agrupados pueden analizarse para determinar posibles VNC, y se analizaron posteriormente muestras individuales del grupo si los resultados iniciales indican que se detecta una posible aneuploidía dentro del grupo agrupado.

15 20 En determinados aspectos, los sistemas de ensayo usando uno o más índices que proporcionan información sobre muestras específicas. Por ejemplo, puede usarse un índice en amplificación selectiva o universal que es indicativa de una muestra de la que se amplificó ácido nucleico.

25 En un particular aspecto, los locus seleccionados se aíslan antes de la detección. Los locus seleccionados pueden aislarse de la muestra mixta usando cualquier medio que aisle selectivamente los ácidos nucleicos particulares presentes en la muestra mixta para el análisis, por ejemplo, hibridación, amplificación u otra forma de aislamiento basado en secuencias de los ácidos nucleicos de la muestra mixta. Tras el aislamiento, los ácidos nucleicos seleccionados se distribuyen individualmente en un formato de detección adecuado, por ejemplo, sobre una micromatriz o en una célula de flujo, para la determinación de la secuencia y/o cantidades relativas de cada ácido nucleico seleccionado en la muestra mixta. Las cantidades relativas de los ácidos nucleicos detectados son indicativas del número de copias de cromosomas que corresponden a los ácidos nucleicos seleccionados presentes en la muestra mixta.

30 35 Tras el aislamiento y la distribución de los ácidos nucleicos seleccionados en un formato adecuado, las secuencias seleccionadas se identifican, por ejemplo, mediante determinación de secuencia de la secuencia seleccionada.

40 En un aspecto específico, la invención proporciona un método de ensayo para la detección de la presencia o ausencia de una aneuploidía fetal, que comprende las etapas de proporcionar una muestra mixta que comprende ADNc materno y fetal, amplificar dos o más locus seleccionados de un primer y segundo cromosoma de interés en la muestra mixta, determinar la frecuencia relativa de las regiones seleccionadas de los cromosomas de interés, comparar la frecuencia relativa de los locus seleccionados de los cromosomas primero y segundo de interés, e identificar la presencia o ausencia de una aneuploidía fetal basándose en las frecuencias relativas comparativas de los locus seleccionados.

45 50 En algunos aspectos específicos, las frecuencias relativas de los locus de una región genómica se calculan individualmente, y las frecuencias relativas de los locus individuales se comparan para determinar la presencia o ausencia de una anomalía cromosómica. En otros aspectos específicos, las frecuencias relativas de los locus seleccionados se usan para determinar una frecuencia del cromosoma de un primer y segundo cromosoma de interés y un cromosoma de referencia, y la variación en el número de copias para el cromosoma o una región genómica del cromosoma se basa en las frecuencias del cromosoma comparadas de los cromosomas primero y segundo de interés.

55 60 La muestra mixta usada para el análisis puede obtenerse o derivarse de cualquier muestra que contenga el ácido nucleico de interés que va a analizarse usando el método de ensayo de la invención. Por ejemplo, una muestra mixta puede ser de cualquier fluido materno que comprende tanto ácidos nucleicos libres de células maternas como fetales, incluyendo pero sin limitarse a plasma materno, suero materno o sangre materna. Una muestra mixta de un paciente de trasplante sería cualquier fluido o tejido que contenga ácidos nucleicos libres de células tanto de las células del donante como las células del paciente. Una muestra mixta de un paciente con un tumor maligno contendría ácidos nucleicos libres de células del tejido sano y normal del paciente así como ácidos nucleicos libres de células de las células cancerosas.

65 Aunque se usa preferiblemente el sistema de ensayo para detectar ADNc en una muestra mixta, en determinados aspectos el ADN de interés que va a analizarse usando el método de ensayo de la invención comprende ADN directamente de los diferentes tipos de células en vez de una muestra mixta que contiene ADN de los tipos de células principales y secundarias. Tales muestras pueden obtenerse a partir de diversas fuentes según el ADN

diana. Por ejemplo, pueden derivarse células fetales para el análisis de muestras tales como líquido amniótico, placenta (por ejemplo, las vellosidades coriónicas), y similares. Pueden obtenerse muestras de órganos de donantes en un individuo mediante biopsia. Pueden aislarse organismos infecciosos directamente de un individuo y analizarse tras el aislamiento. Puede extraerse ADN de tejidos o células cancerosas y usarse para el análisis.

5 Otra característica de la invención es que la gran mayoría de los ácidos nucleicos aislados de la muestra mixta y detectados en el sistema de ensayo proporcionan información relevante a la presencia, cantidad y/o naturaleza polimórfica de un locus particular en la muestra mixta. Esto garantiza que la mayoría de ácidos nucleicos analizados en el método de ensayo de la invención son informativos.

10 En algunos aspectos, un conjunto de múltiples locus seleccionados se interroga para cada región genómica, y la cantidad del conjunto de regiones seleccionadas presentes en la muestra mixta se suma individualmente para determinar la frecuencia relativa de una región genómica en una muestra mixta. Esto incluye determinación de la frecuencia del locus para el ADN materno y fetal combinado presente en la muestra mixta. Preferiblemente, la determinación no requiere una distinción entre el ADN de fuentes diferentes, aunque en determinados aspectos esta información puede obtenerse además de la información de frecuencias relativas en la muestra como un todo.

15 En aspectos preferidos, se detectan ácidos nucleicos seleccionados correspondientes a locus informativos y se suman para determinar la frecuencia relativa de una región genómica en la muestra mixta. Frecuencias que son mayores de lo esperado para locus de una primera región genómica cuando se comparan con los locus de un segundo locus en una muestra mixta son indicativas de una VNC de la primera región genómica en la muestra mixta.

20 La comparación de regiones genómicas puede ser una comparación de parte o todo un cromosoma. Por ejemplo, la región genómica detectada para VNC puede ser un cromosoma completo en el feto (por ejemplo, cromosomas 18 y 21), donde la probabilidad de que ambos sean aneuploides es mínima. Esto puede ser una comparación de cromosomas en los que uno es supuestamente aneuploide (por ejemplo, cromosoma 21) y el otro actúa como un cromosoma de referencia (por ejemplo, un autosoma tal como el cromosoma 2). En aún otros aspectos, la comparación puede usar dos o más cromosomas que son supuestamente aneuploides y uno o más cromosomas de referencia.

25 En un aspecto, el método de ensayo de la invención analiza múltiples ácidos nucleicos que representan locus seleccionados en cromosomas de interés, y la frecuencia relativa de cada locus seleccionado de la muestra se analiza para determinar una frecuencia del cromosoma relativa para cada cromosoma de interés particular en la muestra. La frecuencia cromosómica de dos o más cromosomas o porciones de los mismos se compara entonces para determinar estadísticamente si existe una anomalía cromosómica.

30 En otro aspecto, el método de ensayo de la invención analiza múltiples copias de un conjunto de locus seleccionados en cromosomas de interés, y la frecuencia relativa de cada uno de los locus seleccionados de la muestra se analiza y se cuantifica independientemente para determinar una frecuencia para cada locus seleccionado en la muestra. La suma de los locus en la muestra se compara para determinar estadísticamente si existe una VNC para uno o más locus en una región genómica de una fuente en una muestra mixta.

35 En otro aspecto, se analizan subconjuntos de locus en cada cromosoma para determinar si existe una anomalía cromosómica. La frecuencia de locus puede sumarse para un cromosoma particular, y las sumas de los locus usados para determinar aneuploidía. Este aspecto de la invención suma las frecuencias de los locus individuales en cada región genómica y entonces compara la suma de los locus en una región genómica de un cromosoma frente a una región genómica de otro cromosoma para determinar si existe una anomalía cromosómica. Los subconjuntos de locus pueden escogerse al azar pero con suficientes números de locus para proporcionar un resultado estadísticamente significativo para determinar si existe una anomalía cromosómica. Pueden realizarse múltiples análisis de diferentes subconjuntos de locus dentro de una muestra mixta para proporcionar más poder estadístico. En otro aspecto, pueden seleccionarse locus particulares que se sabe que tienen menos variación entre muestras, o limitando los datos usados para la determinación de frecuencia cromosómica, por ejemplo, ignorando los datos de los locus con frecuencia muy alta o muy baja dentro de una muestra.

40 En un aspecto particular, las cantidades medidas de uno o más locus particulares se normalizan para justificar diferencias en la cantidad de locus en la muestra. Esto puede realizarse normalizando una variación conocida a partir de fuentes tales como el sistema de ensayo (por ejemplo, temperatura, diferencias de lotes de reactivo), biología subyacente de la muestra (por ejemplo, contenido en ácido nucleico), diferencias de operario, o cualquier otra variable.

45 En determinados aspectos específicos, determinar el porcentaje relativo de ácidos nucleicos a partir de la fuente secundaria en una muestra mixta puede ser beneficioso para realizar el sistema de ensayo, ya que proporcionará información importante sobre la presencia estadística relativa de locus que puede ser indicativa de variación en el número de copias dentro de la fuente secundaria en esa muestra. Determinar los locus contribuidos a la muestra mixta de la fuente secundaria puede proporcionar información usada para calcular las diferencias estadísticamente

significativas en frecuencias para regiones genómicas de interés. Por tanto, tales locus podían proporcionar dos formas de información en el ensayo, puede usarse información alélica para determinar el porcentaje de contribución de células secundarias en una muestra mixta y puede usarse una suma de la información alélica para determinar la frecuencia relativa global de ese locus en una muestra mixta. La información alélica no se necesita para determinar la frecuencia relativa global de ese locus.

En otro aspecto específico, el método de ensayo de la invención puede usarse para determinar un posible mosaicismo en una población de células, y si deben llevarse a cabo pruebas de confirmación adicionales para confirmar la identificación de mosaicismo en la fuente principal y/o secundaria. En determinados casos, la determinación del porcentaje de ácidos nucleicos de la fuente secundaria en una muestra mixta podría ayudar en la cuantificación del nivel estimado de mosaicismo. El mosaicismo podría confirmarse posteriormente usando otros métodos de prueba que podrían distinguir aneuploidía de mosaico completo o parcial en tejido o células específicos.

En aún otro aspecto específico, el método de ensayo de la invención puede usarse para determinar la contaminación en una muestra, representando la especie secundaria una especie contaminante.

En otro aspecto, los oligonucleótidos para un ácido nucleico dado seleccionado pueden conectarse en los extremos específicos no secuenciales de modo que puede unirse a los mismos una sonda circular o unimolecular. En este aspecto, el extremo 3' y el extremo 5' de la sonda circular se unen al locus seleccionado y al menos una región de amplificación universal está presente en la secuencia específica no seleccionada de la sonda circular.

Una característica importante del ensayo es que los productos de amplificación pueden analizarse directamente sin la necesidad de regiones polimórficas de la muestra mixta inicial. Por tanto, la invención actual permite la detección tanto de VNC como polimorfismos de una muestra materna sin una etapa de enriquecimiento polimórfico de intervención antes de la determinación de secuencia de los locus seleccionados.

Otra característica importante del ensayo es que se determinan tanto VNC como la contribución de fuente usando un enfoque dirigido de amplificación y detección seleccionados. Esto permite que la mayor parte de la información recogida en el ensayo sea útil para la determinación de la VNC y/o contribución de fuente, y obvia la necesidad de generar lecturas de secuencia que deben alinearse con una secuencia de referencia.

Estos y otros aspectos, características y ventajas se proporcionarán en más detalle tal como se describe en el presente documento.

### 35 **Breve descripción de las figuras**

La figura 1 es un diagrama de flujo simplificado de las etapas generales usadas en los sistemas de ensayo de la invención.

40 La figura 2 ilustra un primer esquema general para un sistema de ensayo basado en ligación de la invención.

La figura 3 ilustra un segundo esquema general para un sistema de ensayo basado en ligación de la invención.

45 La figura 4 es un tercer esquema general para un sistema de ensayo basado en ligación de la invención.

La figura 5 ilustra el rendimiento de genotipado que se obtuvo usando un sistema de ensayo de la invención.

50 La figura 6 es un gráfico que ilustra resultados de una determinación de porcentaje fetal usando un ensayo de la invención.

La figura 7 ilustra los elementos usados para una detección de aneuploidía y polimorfismo para dos cohortes de muestras maternas.

55 La figura 8 es un resumen de información del paciente y de la muestra y datos para un subconjunto de una segunda cohorte de sujetos gestantes.

La figura 9 ilustra la detección de aneuploidía del cromosoma 21 lograda usando un aspecto de la invención para una primera cohorte.

60 La figura 10 ilustra la detección de aneuploidía del cromosoma 18 lograda usando un aspecto de la invención para una primera cohorte.

La figura 11 ilustra la detección de aneuploidía del cromosoma 21 lograda usando un aspecto de la invención para una segunda cohorte.

65 La figura 12 ilustra la detección de aneuploidía del cromosoma 18 lograda usando un aspecto de la invención para

una segunda cohorte.

**Definiciones**

- 5 Se pretende que los términos usados en el presente documento tengan el significado simple y común tal como entienden los expertos habituales en la técnica. Se pretende que las siguientes definiciones ayuden al lector a comprender la presente invención, pero no se pretende que varíen o de otro modo limiten el significado de tales términos a menos que se indique específicamente.
- 10 El término “ácido nucleico amplificado” es cualquier molécula de ácido nucleico cuya cantidad se ha aumentado al menos dos veces mediante cualquier método de replicación o amplificación de ácido nucleico realizado *in vitro* en comparación con su cantidad de partida en una muestra mixta.
- 15 El término “producto de amplificación” tal como se usa en el presente documento se refiere al producto resultante de una reacción de amplificación usando el producto de ligación contigua como una molde, o el producto resultante de una reacción de amplificación usando una molécula complementaria al producto de ligación contigua como una molde.
- 20 El término “anomalía cromosómica” se refiere a cualquier variación genética que afecta a todo o parte de un cromosoma mayor que un único locus. Las variantes genéticas pueden incluir pero no se limita a cualquier VNC tal como amplificaciones o deleciones, translocaciones, inversiones y mutaciones. Los ejemplos de anomalías cromosómicas incluyen, pero no se limitan a, síndrome de Down (trisomía 21), síndrome de Edwards (trisomía 18), síndrome de Patau (trisomía 13), síndrome de Klinefelter (XXY), síndrome de triple X, síndrome XYY, trisomía 8, trisomía 16, síndrome de Turner, translocación robertsoniana, síndrome de DiGeorge y síndrome de Wolf-Hirschhorn.
- 25 Los términos “complementario” o “complementariedad” se usan en referencia a moléculas de ácido nucleico (es decir, una secuencia de nucleótidos) que se relacionan mediante normas de apareamiento de bases. Nucleótidos complementarios son, generalmente, A y T (o A y U), o C y G. Se dice que dos moléculas de ARN o ADN monocatenario son sustancialmente complementarias cuando los nucleótidos de una hebra, alineados óptimamente y con inserciones o deleciones de nucleótidos apropiadas, se aparean con al menos de aproximadamente el 90% a aproximadamente el 95% de complementariedad, y más preferiblemente desde aproximadamente el 98% hasta aproximadamente el 100% de complementariedad, e incluso más preferiblemente con el 100% de complementariedad. Alternativamente, existe complementariedad sustancial cuando una hebra de ARN o ADN se hibrida en condiciones de hibridación selectivas con su complemento. Las condiciones de hibridación selectivas incluyen, pero no se limitan a, condiciones de hibridación rigurosas. Las condiciones de hibridación rigurosas incluirán normalmente concentraciones de sal de menos de aproximadamente 1 M, más habitualmente menos de aproximadamente 500 mM y preferiblemente menos de aproximadamente 200 mM. Las temperaturas de hibridación son generalmente de al menos aproximadamente 2°C a aproximadamente 6°C menores que las temperaturas de fusión ( $T_f$ ).
- 30
- 35
- 40 El término “variación en el número de copias” o “VNC” tal como se usa indistintamente en el presente documento son alteraciones del ADN de un genoma que da como resultado una célula que tiene un número anómalo de copias de uno o más locus en el ADN. Las VNC que son clínicamente relevantes pueden limitarse a un único gen o incluir un conjunto contiguo de genes. Una VNC también puede corresponderse con regiones relativamente grandes del genoma que se han deleccionado, invertido o duplicado en determinados cromosomas, hasta una que incluye una o más copias adicionales de un cromosoma completo. El término VNC tal como se usa en el presente documento no se refiere a ninguna información relacionada con la secuencia, sino más bien con la cantidad o “recuentos” de regiones genéticas presentes en una muestra.
- 45
- 50 El término “índice de corrección” se refiere a un índice que puede contener nucleótidos adicionales que permiten la identificación y corrección de amplificación, secuenciación u otros errores experimentales incluyendo la detección de deleción, sustitución o inserción de una o más bases durante la secuenciación así como cambios de nucleótidos que pueden producirse fuera de la secuenciación tales como síntesis de oligonucleótido, amplificación, y cualquier otro aspecto del ensayo. Estos índices de corrección pueden ser índices independientes que son secuencias diferentes, o pueden incrustarse dentro de otras regiones para ayudar a confirmar la exactitud de las técnicas experimentales usadas, por ejemplo, un índice de corrección puede ser un subconjunto de secuencias usado para amplificación universal o un subconjunto de nucleótidos de un locus de muestra.
- 55
- 60 El término “herramienta de diagnóstico” tal como se usa en el presente documento se refiere a cualquier composición o ensayo de la invención usado en combinación como, por ejemplo, en un sistema con el fin de llevar a cabo una prueba o ensayo de diagnóstico en la muestra de un paciente.
- 65 El término “rasgo de la enfermedad” se refiere a un rasgo monogénico o poligénico asociado con un estado patológico, por ejemplo, una enfermedad, trastorno, síndrome o predisposición.

El término “región genómica” tal como se usa en el presente documento se refiere a cualquier región de uno o más locus que se encuentran normalmente en de manera contigua en un genoma. Una región genómica puede variar en tamaño hasta e incluir un cromosoma completo.

- 5 El término “hibridación” significa generalmente la reacción por la que se produce el apareamiento de hebras complementarias de ácido nucleico. El ADN es habitualmente bicatenario, y cuando las hebras se separan se volverán a hibridar en las condiciones adecuadas. Pueden formarse híbridos entre ADN-ADN, ADN-ARN o ARN-ARN. Pueden formarse entre una hebra corta y una hebra larga que contienen una región complementaria a la corta. También pueden formarse híbridos imperfectos, pero cuanto más imperfectos sean, menos estables serán (y será menos probable que se formen).

15 El término “locus informativo” tal como se usa en el presente documento se refiere a un locus que es homocigoto para una fuente de células y heterocigoto para una segunda fuente de células en un cromosoma particular o porción de un cromosoma interrogado con fines de determinar una VNC de todo o parte de ese cromosoma. Los locus informativos para su uso en el sistema de ensayo de la invención incluyen locus usados para interrogación de un cromosoma de referencia así como locus usados para interrogación de un cromosoma que es supuestamente aneuploide en una fuente de células. Los locus informativos también pueden distinguir el número de copias de locus en fuentes de células de diferentes individuos dentro de un único individuo (por ejemplo, detección de células del donante del trasplante en un receptor de trasplante o detección de un ADN fetal dentro de una muestra materna mixta).

Los términos “locus” y “locus” tal como se usa en el presente documento se refieren a un locus de ubicación conocida en un genoma.

- 25 El término “fuente principal” se refiere a una fuente de ácidos nucleicos en una muestra de un individuo que es representativa del material genómico predominante en ese individuo.

30 El término “muestra materna” tal como se usa en el presente documento se refiere a cualquier muestra tomada de un mamífero gestante que comprende tanto material genómico libre de células fetal como materno (por ejemplo, ADN). Preferiblemente, se obtienen muestras maternas para su uso en la invención mediante medios relativamente no invasivos, por ejemplo, flebotomía u otras técnicas convencionales para extraer muestras periféricas de un sujeto.

35 El término “temperatura de fusión” o  $T_f$  se define comúnmente como la temperatura a la que una población de moléculas de ácido nucleico bicatenario se disocia a medias en hebras individuales. La ecuación para calcular la  $T_f$  de ácidos nucleicos se conoce bien en la técnica. Tal como se indica mediante referencias convencionales, puede calcularse una estimación sencilla del valor de  $T_f$  mediante la ecuación:  $T_f = 81,5 + 16,6(\log_{10}[\text{Na}^+]) - 0,41(\%[\text{G}+\text{C}]) - 675/n - 1,0m$ , cuando un ácido nucleico está en disolución acuosa que tiene concentraciones catiónicas de 0,5 M o menos, el contenido en (G+C) está entre el 30% y el 70%, n es el número de bases, y m es el porcentaje de apareamientos erróneos de pares de bases (véase, por ejemplo, Sambrook J *et al.*, Molecular Cloning, A Laboratory Manual, 3ª ed., Cold Spring Harbor Laboratory Press (2001)). Otras referencias incluyen cálculos más sofisticados, que consideran las características estructurales así como secuenciales para el cálculo de  $T_f$ .

45 “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 porta una matriz de sitios que contienen 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 elementos de la matriz; es decir, los sitios están separados espacialmente. La matriz o micromatriz también puede comprender una estructura interrogable no plana con una superficie tal como una perla o un pocillo. Los oligonucleótidos o polinucleótidos de la matriz pueden estar unidos covalentemente al soporte sólido, o pueden estar unidos de manera no covalente. La tecnología de micromatrices convencional se revisa en, por ejemplo, Schena, Ed., Microarrays: A Practical Approach, IRL Press, Oxford (2000). “Análisis de matriz”, “análisis mediante matriz” o “análisis mediante micromatriz” se refiere a análisis, tales como, por ejemplo, análisis de secuencia, de una o más moléculas biológicas usando una micromatriz.

- 55 El término “fuente secundaria” se refiere a una fuente de ácidos nucleicos dentro de un individuo que está presente en cantidades limitadas y que puede distinguirse de la fuente principal debido a diferencias en su constitución genómica y/o expresión. Los ejemplos de fuentes secundarias incluyen, pero no se limitan a, células fetales en una mujer embarazada, células cancerosas en un paciente con un tumor maligno, células de un órgano de donante en un paciente de trasplante, ácidos nucleicos de un organismo infeccioso en un huésped infectado, y similares.

60 El término “muestra mixta” tal como se usa en el presente documento se refiere a cualquier muestra que comprende material genómico libre de células (por ejemplo, ADN) de dos o más tipos celulares de interés, siendo uno una fuente principal y siendo el otro una fuente secundaria dentro de un único individuo. Las muestras mixtas a modo de ejemplo incluyen una muestra materna (por ejemplo, sangre, suero o plasma maternos que comprenden tanto ADN materno como fetal), y una muestra somática derivada periféricamente (por ejemplo, sangre, suero o plasma que comprende diferentes tipos celulares, por ejemplo, células hematopoyéticas, células mesenquimatosas y células

circulantes de otros sistemas de órganos). Las muestras mixtas incluyen muestras con material genómico tanto de una fuente principal como una secundaria en un individuo, que puede ser por ejemplo, células somáticas normales y atípicas, o células que comprenden genomas de dos individuos diferentes, por ejemplo, una muestra tanto con material genómico materno como fetal o una muestra de un paciente de trasplante que comprende células tanto del donante como del receptor.

El término “rasgo monogénico” tal como se usa en el presente documento se refiere a cualquier rasgo, normal o patológico, que se asocia con una mutación o polimorfismo en un único gen. Tales rasgos incluyen rasgos asociados con una enfermedad, trastorno o predisposición provocados por una disfunción en un único gen. Los rasgos también incluyen características no patológicas (por ejemplo, presencia o ausencia de moléculas de la superficie celular en un tipo celular específico).

El término alelo “no materno” significa un alelo con un polimorfismo y/o mutación que se encuentra en un alelo fetal (por ejemplo, un alelo con una mutación o SNP *de novo*) y/o un alelo paterno, pero que no se encuentra en el alelo materno.

Por “no polimórfico”, cuando se usa con respecto a la detección de locus seleccionados, quiere decirse una detección de tal locus, que puede contener uno o más polimorfismos, pero en el que la detección no depende de la detección del polimorfismo específico dentro de la región. Por tanto, un locus seleccionado puede contener un polimorfismo, pero la detección de la región usando el sistema de ensayo de la invención se basa en la aparición de la región en lugar de la presencia o ausencia de un polimorfismo particular en esa región.

Tal como se usa en el presente documento “nucleótido” se refiere a una combinación de base-azúcar-fosfato. Los nucleótidos son unidades monoméricas de una secuencia de ácido nucleico (ADN y ARN). El término nucleótido incluye ribonucleósidos trifosfato ATP, UTP, CTG, GTP y desoxirribonucleósidos trifosfato tales como dATP, dCTP, dTTP, dUTP, dGTP, dTTP, o derivados de los mismos. Tales derivados incluyen, por ejemplo, [ $\alpha$ S]dATP, 7-deaza-dGTP y 7-deaza-dATP, y derivados de nucleótidos que confieren resistencia a nucleasas en la molécula de ácido nucleico que las contienen. El término nucleótido tal como se usa en el presente documento también se refiere a didesoxirribonucleósidos trifosfato (ddNTP) y a sus derivados. Los ejemplos ilustrados de didesoxirribonucleósidos trifosfato incluyen, pero no se limitan a, ddATP, ddCTP, ddGTP, ddITP y ddTTP.

Según la presente invención, un “nucleótido” puede no estar marcado o marcado de manera detectable mediante técnicas bien conocidas. Los marcadores fluorescentes y su unión a oligonucleótidos se describen en muchas revisiones, incluyendo Haugland, Handbook of Fluorescent Probes and Research Chemicals, 9<sup>a</sup> ed., Molecular Probes, Inc., Eugene OR (2002); Keller y Manak, DNA Probes, 2<sup>a</sup> ed., Stockton Press, Nueva York (1993); Eckstein, ed., Oligonucleotides and Analogues: A Practical Approach, IRL Press, Oxford (1991); Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26:227-259 (1991); y similares. Otras metodologías aplicables a la invención se dan a conocer en la siguiente muestra de referencias: Fung *et al.*, patente estadounidense n.º 4.757.141; Hobbs, Jr., *et al.*, patente estadounidense n.º 5.151.507; Cruickshank, patente estadounidense n.º 5.091.519; Menchen *et al.*, patente estadounidense n.º 5.188.934; Begot *et al.*, patente estadounidense n.º 5.366.860; Lee *et al.*, patente estadounidense n.º 5.847.162; Khanna *et al.*, patente estadounidense n.º 4.318.846; Lee *et al.*, patente estadounidense n.º 5.800.996; Lee *et al.*, patente estadounidense n.º 5.066.580; Mathies *et al.*, patente estadounidense n.º 5.688.648; y similares. También puede llevarse a cabo el marcaje con puntos cuánticos, tal como se da a conocer en las siguientes patentes y publicaciones de patente: patentes estadounidenses n.ºs 6.322.901; 6.576.291; 6.423.551; 6.251.303; 6.319.426; 6.426.513; 6.444.143; 5.990.479; 6.207.392; 2002/0045045; y 2003/0017264. Los marcadores detectables incluyen, por ejemplo, isótopos radiactivos, marcadores fluorescentes, marcadores quimioluminiscentes, marcadores bioluminiscentes y marcadores enzimas. Los marcadores fluorescentes de nucleótidos pueden incluir, pero no se limitan a, fluoresceína, 5-carboxifluoresceína (FAM), 2'7'-dimetoxi-4'5'-dicloro-6-carboxifluoresceína (JOE), rodamina, 6-carboxi-rodamina (R6G), N,N,N',N'-tetrametil-6-carboxi-rodamina (TAMRA), 6-carboxi-X-rodamina (ROX), ácido 4-(4'-dimetilaminofenilazo)benzoico (DABCIL), Cascade Blue, Oregon Green, Texas Red, cianina y ácido 5-(2'-amoetil)aminonaftaleno-1-sulfónico (EDANS). Los ejemplos específicos de nucleótidos marcados de manera fluorescente incluyen [R6G]dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP y [dROX]ddTTP disponibles de Perkin Elmer, Foster City, Calif. FluoroLink DeoxiNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink FluorX-dCTP, FluoroLink Cy3-dUTP, y FluoroLink Cy5-dUTP disponibles de Amersham, Arlington Heights, IL; fluoresceína-15-dATP, fluoresceína-12-dUTP, tetrametil-rodamina-6-dUTP, IR770-9-dATP, fluoresceína-12-ddUTP, fluoresceína-12-UTP y fluoresceína-15-2'-dATP disponibles de Boehringer Mannheim, Indianápolis, IN; y nucleótidos marcados con cromosomas, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, Cascade Blue-7-UTP, Cascade Blue-7-dUTP, fluoresceína-12-UTP, fluoresceína-12-dUTP, Oregon Green 488-5-dUTP, rodamina Green-5-UTP, rodamina Green-5-dUTP, tetrametilrodamina-6-UTP, tetrametilrodamina-6-dUTP, Texas Red-5-UTP, Texas Red-5-dUTP y Texas Red-12-dUTP disponibles de Molecular Probes, Eugene, OR.

Los términos “oligonucleótidos” u “oligonucleótidos” tal como se usan en el presente documento se refieren a oligómeros lineales de monómeros de ácido nucleico natural o modificado, incluyendo desoxirribonucleótidos,

ribonucleótidos, formas anoméricas de los mismos, monómeros de ácidos nucleicos peptídicos (PNA), monómeros de ácido nucleico bloqueado (LNA), y similares, o una combinación de los mismos, que pueden unirse específicamente a un polinucleótido de cadenas sencillas mediante un patrón regular de interacciones entre monómeros, tales como apareamiento de bases de tipo Watson-Crick, apilamiento de base, apareamiento de bases de tipo Hoogsteen o Hoogsteen inverso, o similares. Habitualmente, los monómeros se unen mediante enlaces fosfodiéster o análogos de los mismos para formar oligonucleótidos que oscilan en tamaño entre unas pocas unidades monoméricas, por ejemplo, 8-12, y varias decenas de unidades monoméricas, por ejemplo, 100-200 o más. Pueden prepararse moléculas adecuadas de ácido nucleico mediante el método de fosforamidita descrito por Beaucage y Carruthers (Tetrahedron Lett., 22:1859-1862 (1981)), o mediante el método triéster según Matteucci, *et al.* (J. Am. Chem. Soc., 103:3185 (1981)), o mediante otros métodos químicos tales como usar un sintetizador de oligonucleótidos automatizado comercial.

El término "rasgo poligénico" tal como se usa en el presente documento se refiere a cualquier rasgo, normal o patológico, que se asocia con una mutación o polimorfismo en más de un único gen. Tales rasgos incluyen rasgos asociados con una enfermedad, trastorno, síndrome o predisposición provocados por una disfunción en dos o más genes. Los rasgos también incluyen características no patológicas asociadas con la interacción de dos o más genes.

Tal como se usa en el presente documento el término "polimerasa" se refiere a una enzima que une nucleótidos individuales entre sí en una hebra larga, usando otra hebra como una molde. Hay dos tipos generales de polimerasa-ADN polimerasas, que sintetizan ADN, y ARN polimerasas, que sintetizan ARN. Dentro de estas dos clases, hay numerosos subtipos de polimerasas, según qué tipo de ácido nucleico puede funcionar como molde y qué tipo de ácido nucleico se forma.

Tal como se usa en el presente documento "reacción en cadena de la polimerasa" o "PCR" se refiere a una técnica para replicar un fragmento específico de ADN seleccionado *in vitro*, incluso en presencia de exceso de ADN inespecífico. Se añaden cebadores al ADN seleccionado, donde los cebadores inician la copia del ADN seleccionado usando nucleótidos y, normalmente, Taq polimerasa o similares. Al ciclar la temperatura, el ADN seleccionado se desnaturaliza de manera repetitiva y se copia. Una única copia del ADN seleccionado, incluso si se mezcla con otro ADN aleatorio, puede amplificarse para obtener billones de réplicas. La reacción en cadena de la polimerasa puede usarse para detectar y medir cantidades muy pequeñas de ADN y para crear fragmentos personalizados de ADN. En algunos casos, pueden usarse métodos de amplificación lineal como alternativa a PCR.

El término "polimorfismo" tal como se usa en el presente documento se refiere a cualquier cambio genético o variante de secuencia en un locus, incluyendo pero sin limitarse a polimorfismos de nucleótido único (SNP), diferencias de metilación, repeticiones cortas en tándem (STR), polimorfismos de un único gen, mutaciones puntuales, repeticiones de trinucleótidos, indels y similares.

Generalmente, un "cebador" es un oligonucleótido usado para, por ejemplo, cebar la extensión, ligación y/o síntesis de ADN, tales como en la etapa de síntesis de la reacción en cadena de la polimerasa o en las técnicas de extensión de cebador usadas en determinadas reacciones de secuenciación. También puede usarse un cebador en técnicas de hibridación como medios para proporcionar complementariedad de un locus a un oligonucleótido de captura para la detección de un locus específico.

El término "herramienta de investigación" tal como se usa en el presente documento se refiere a cualquier composición o ensayo de la invención usado para investigación científica, de naturaleza académica o comercial, incluyendo el desarrollo de opciones terapéuticas farmacéuticas y/o biológicas. No se pretende que las herramientas de investigación de la invención sean terapéuticas o se sometan a aprobación regulatoria; más bien, se pretende que las herramientas de investigación de la invención faciliten la investigación y ayuden en tales actividades de desarrollo, incluyendo cualquier actividad realizada con la intención de producir información para respaldar un sometimiento regulatorio.

El término "índice de muestra" se refiere generalmente a una serie de nucleótidos únicos (es decir, cada índice de muestra es único para una muestra en un sistema de ensayo multiplexado para el análisis de múltiples muestras). El índice de muestra puede, por tanto, usarse para ayudar en la identificación de locus para la multiplexación de diferentes muestras en un único recipiente de reacción, de modo que cada muestra puede identificarse basándose en su índice de muestra. En un aspecto preferido, hay un único índice de muestra para cada muestra en un conjunto de muestras, y las muestras se agrupan durante la secuenciación. Por ejemplo, si se agrupan doce muestras en una única reacción de secuenciación, hay al menos al menos doce índices de muestras único de manera que cada muestra se marca de manera única.

El término "locus seleccionado" tal como se usa en el presente documento se refiere a un locus correspondiente a locus interrogados, por ejemplo, para el número de copia, la presencia o ausencia de uno o más polimorfismos, presencia o ausencia de un organismo infeccioso, etc. Tales locus seleccionados pueden aislarse directamente y amplificarse de la muestra para detección, por ejemplo, basándose en la hibridación y/u otras técnicas basadas en secuencias, o pueden amplificarse usando la muestra como una molde antes de la detección de la secuencia. Pueden seleccionarse regiones de ácidos nucleicos para su uso en los sistemas de ensayo de la presente invención.

basándose en la variación del nivel de ADN entre individuos, basándose en la especificidad para un cromosoma particular, basándose en el contenido en CG y/o condiciones de amplificación requeridas de los locus seleccionados, u otras características que serán evidentes para un experto en la técnica tras leer la presente divulgación.

5 Los términos “secuenciación”, “determinación de secuencia” y similares tal como se usa en el presente documento se refieren generalmente a cualquiera y a todos los métodos bioquímicos que pueden usarse para determinar el orden de bases de nucleótido en un ácido nucleico.

10 El término “contribución de fuente” tal como se usa en el presente documento se refiere a la contribución relativa de dos o más fuentes de ácidos nucleicos dentro de un individuo. La contribución de una sola se determina generalmente como porcentaje de los ácidos nucleicos de una muestra, aunque puede usarse cualquier medición relativa.

15 El término “se une específicamente”, “unión específica” y similares tal como se usa en el presente documento, cuando se refiere a un componente de unión (por ejemplo, una sonda o cebador de ácido nucleico, anticuerpo, etc.) que da como resultado la generación de una señal positiva estadísticamente significativa en las condiciones de ensayo designadas. Normalmente la interacción dará como resultado posteriormente una señal detectable que es al menos al menos el doble de la desviación estándar de cualquier señal generada como resultado de interacciones no deseadas (fondo).

20 El término “estado” tal como se usa en el presente documento en relación con un gen se refiere al estado de secuencia de los alelos de un gen particular, incluyendo las regiones codificantes y las regiones no codificantes que afectan a la traducción y/o expresión de proteínas de ese gen. El estado de un gen asociado con una enfermedad dominante autosómica tal como acondroplasia (por ejemplo, el gen que codifica para el receptor del factor de crecimiento de fibroblastos) o enfermedad de Huntington (por ejemplo, el gen Huntingtin), o para una enfermedad ligada a X en el caso de un feto masculino, puede clasificarse como afectado, es decir, un alelo posee mutación/mutaciones que es/son causante(s) de las enfermedades o el trastorno, o no afectado, es decir, ambos alelos carecen de tal(es) mutación/mutaciones. El estado de un gen asociado con una enfermedad recesiva autosómica o un gen materno asociado con un trastorno recesivo ligado a X, puede clasificarse como afectado, es decir, ambos alelos poseen mutación/mutaciones causante(s) de las enfermedades o trastorno; portador, es decir un alelo posee mutación/mutaciones causante(s) de la enfermedades o trastorno; o no afectado, es decir, ambos alelos carecen de tal(es) mutación/mutaciones. El estado de un gen también puede indicar la presencia o ausencia de un alelo particular asociado con el riesgo de desarrollar una enfermedad poligénica, por ejemplo, un polimorfismo que es protector frente a una enfermedad o trastorno particular o un polimorfismo asociado con un riesgo potenciado para una enfermedad o trastorno particular.

**Descripción detallada de la invención**

40 Los sistemas de ensayo y métodos descritos en el presente documento pueden emplear, a menos que se indique lo contrario, técnicas convencionales y descripciones de biología molecular (incluyendo técnicas recombinantes), biología celular, bioquímica, tecnología de micromatriz y secuenciación, que están dentro de la habilidad de los practicantes en la técnica. Tales técnicas convencionales incluyen síntesis de matriz polimérica, hibridación y ligación de oligonucleótidos, secuenciación de oligonucleótidos y detección de hibridación usando un marcador. Pueden tenerse ilustraciones específicas de técnicas adecuadas mediante referencia a los ejemplos en el presente documento. Sin embargo, evidentemente también pueden usarse procedimientos convencionales equivalentes. Tales técnicas convencionales y descripciones pueden encontrarse en manuales de laboratorio convencionales tales como Green, *et al.*, eds., *Genome Analysis: A Laboratory Manual Series* (vol. I-IV) (1999); Weiner, *et al.*, eds., *Genetic Variation: A Laboratory Manual* (2007); Dieffenbach, Dveksler, eds., *PCR Primer: A Laboratory Manual* (2003); Bowtell y Sambrook, *DNA Microarrays: A Molecular Cloning Manual* (2003); Mount, *Bioinformatics: Sequence and Genome Analysis* (2004); Sambrook y Russell, *Condensed Protocols from Molecular Cloning: A Laboratory Manual* (2006); y Sambrook y Russell, *Molecular Cloning: A Laboratory Manual* (2002) (todos de Cold Spring Harbor Laboratory Press); Stryer, L., *Biochemistry* (4ª ed.) W.H. Freeman, Nueva York (1995); Gait, “*Oligonucleotide Synthesis: A Practical Approach*” IRL Press, Londres (1984); Nelson y Cox, *Lehninger, Principles of Biochemistry*, 3ª ed., W. H. Freeman Pub., Nueva York (2000); y Berg *et al.*, *Biochemistry*, 5ª ed., W.H. Freeman Pub., Nueva York (2002). Antes de que se describan las presentes composiciones, herramientas de investigación y métodos, ha de entenderse que esta invención no se limita a los métodos, composiciones, dianas y usos específicos descritos, puesto que pueden variar, evidentemente. También debe entenderse que la terminología usada en el presente documento es con el fin de describir aspectos particulares únicamente y no se pretende que limite el alcance de la presente invención, que se limitará únicamente mediante las reivindicaciones adjuntas.

60 Debe indicarse que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas en singular “un”, “una” y “el” incluyen referentes en plural a menos que el contexto dicte claramente lo contrario. Por tanto, por ejemplo, la referencia a “un locus” se refiere a uno, más de uno, o mezclas de tales regiones, y la referencia a “un ensayo” incluye la referencia a etapas y métodos equivalentes conocidos por los expertos en la técnica, y así sucesivamente.

Cuando se proporciona un intervalo de valores, debe entenderse que cada valor de intervención entre el límite superior e inferior de ese intervalo, y cualquier otro valor declarado o de intervención en ese intervalo declarado, está abarcado dentro de la invención. Cuando el intervalo declarado incluye límites superiores e inferiores, también deben incluirse intervalos que excluyen cualquiera de los límites incluidos en la invención.

5 A menos que se declare expresamente, se pretende que los términos usados en el presente documento tengan el significado simple y común tal como entienden los expertos habituales en la técnica. Se pretende que las siguientes definiciones ayuden al lector a comprender la presente invención, pero no se pretende que varíen o de otro modo limiten el significado de tales términos a menos que se indique específicamente.

10 En la siguiente descripción, se exponen numerosos detalles específicos para proporcionar un entendimiento más concienzudo de la presente invención. Sin embargo, será evidente para un experto en la técnica que la presente invención puede practicarse sin uno o más de estos detalles específicos. En otros casos, no se han descrito características bien conocidas y procedimientos bien conocidos para los expertos en la técnica con el fin de evitar complicar la invención.

### 15 La invención en general

20 La presente invención proporciona métodos de ensayo individuales con la capacidad de detectar variaciones en el número de copias, polimorfismos y ácidos nucleicos asociados con estados patológicos (por ejemplo, ácidos nucleicos de patógenos, o asociados con cáncer, diabetes, enfermedad de Alzheimer, y similares) en una muestra mixta de un único individuo. El ensayo permite la identificación de variación genética en una o más fuentes secundarias en una muestra mixta usando información sobre el número de copias de locus seleccionados en la muestra e información sobre el porcentaje contribución de ácidos nucleicos de la fuente principal y la una o más

25 fuentes secundarias en la muestra. Estos métodos son útiles para cualquier muestra mixta que contiene material genómico (por ejemplo, ADN) de una fuente principal y una fuente secundaria que están presentes en un único individuo.

30 El uso de locus seleccionados en los métodos de ensayo de la invención proporciona detección directa de locus para la determinación de variación en el número de copias en una o más fuentes dentro de la muestra mixta. Una clara ventaja de la invención es que los locus seleccionados correspondientes a la variación en el número de copias y/o polimorfismos, pueden analizarse además usando una variedad de técnicas de detección y cuantificación, incluyendo pero sin limitarse a técnicas de hibridación, PCR digital y técnicas de determinación de secuenciación de alto rendimiento. Las sondas pueden diseñarse contra cualquier número de locus seleccionados para cualquier

35 cromosoma. Aunque la amplificación antes de la identificación y cuantificación de los locus seleccionados no es obligatoria, puede usarse amplificación limitada de la muestra mixta antes de la detección para expandir el número global de ácidos nucleicos dentro de los materiales de partida.

40 La figura 1 es un diagrama de flujo simplificado de las etapas generales usadas en los sistemas de ensayo de la invención. La figura 1 muestra el método 100, donde en una primera etapa 110, se proporciona una muestra de ácido nucleico mixta el análisis. La muestra mixta puede prepararse a partir de prácticamente cualquier muestra puesto que tales técnicas las conocen los expertos en la técnica (véase, por ejemplo, Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, 4ª ed., capítulo 2, Burtis, C. Ashwood E. y Bruns, D, eds. (2006); Chemical Weapons Convention Chemicals Analysis: Sample Collection, Preparation and Analytical Methods, Mesilaakso, M., ed., (2005); Pawliszyn, J., Sampling and Sample Preparation for Field and Laboratory, (2002); Venkatesh Iyengar, G., *et al.*, Element Analysis of Biological Samples: Principles and Practices (1998); Drielak, S., Hot Zone Forensics: Chemical, Biological, and Radiological Evidence Collection (2004); Wells, D., High Throughput Bioanalytical Sample Preparation (Progress in Pharmaceutical and Biomedical Analysis) (2002)). Según el tipo de muestra mixta elegida, pueden realizarse etapas de procesamiento y/o purificación adicionales para obtener fragmentos de ácido nucleico de un tamaño o pureza deseados, usando métodos de procesamiento incluyendo pero sin limitarse a sonicación, nebulización, purificación en gel, sistemas de purificación de PCR, escisión de nucleasas, o una combinación de estos métodos. En un aspecto preferido, se usan muestras que comprenden ADN libre circulante (ADNlc).

55 En la etapa 120, se introduce un primer conjunto de oligonucleótidos de secuencia fija en la muestra de ácido nucleico mixta, en condiciones que permiten que el primer conjunto de oligonucleótidos de secuencia fija se hibriden con la muestra de ácido nucleico mixta. El primer conjunto de oligonucleótidos de secuencia fija comprende secuencias de ácido nucleico que son complementarias a uno o más locus seleccionados en la muestra mixta, que tal como se describirán en detalle en el presente documento son útiles para determinar variaciones en el número de copias y/o anomalías cromosómicas. Las secuencias de ácido nucleico que pueden determinar variaciones en el

60 número de copias y/o anomalías cromosómicas incluyen secuencias que permiten la identificación de anomalías cromosómicas tales como amplificaciones o deleciones, aneuploidías, translocaciones o inversiones.

65 En la etapa 130, se introducen un segundo conjunto de oligonucleótidos de secuencia fija en la muestra de ácido nucleico mixta y primer conjunto de oligonucleótidos de secuencia fija en condiciones que permiten que el segundo conjunto de oligonucleótidos de secuencia fija se hibride con la muestra de ácido nucleico mixta. El segundo conjunto de oligonucleótidos de secuencia fija comprende secuencias de ácido nucleico que son complementarias a

uno o más locus seleccionados en la muestra mixta, que pueden detectar polimorfismos. Pueden incluirse opcionalmente etapas de lavado entre las etapas 120 y 130, y 130 y 140.

5 En la etapa 140, se ligan los conjuntos primero y segundo de oligonucleótidos de secuencia fija que se han hibridado con regiones adyacentes de los locus seleccionados en la muestra mixta, y en la etapa 150, se amplifican los oligonucleótidos ligados. Entonces se detectan y se analizan los oligonucleótidos ligado y amplificados, lo que permite la determinación de variaciones en el número de copias o anomalías cromosómicas e identificación de polimorfismos en la etapa 160.

10 Los conjuntos de ácidos nucleicos de secuencia fija están diseñados para hibridarse con al menos dos regiones diferentes en un locus seleccionado. En aspectos preferidos, dos o más oligonucleótidos diferentes se usan para hibridarse con estas regiones para proporcionar ácidos nucleicos adyacentes complementarios al locus seleccionado. En algunos aspectos, sin embargo, puede usarse una única sonda que comprende dos o más regiones no adyacentes distintas que son complementarias a los locus seleccionados incluyendo sondas precirculares tales como las denominadas "sondas candado" o "sondas de inversión molecular (MIP, *molecular inversión probes*)".

20 La presente invención proporciona un sistema mejorado sobre técnicas más aleatorias tales como secuenciación en paralelo de manera masiva, secuenciación aleatoria, y el uso de PCR digital aleatoria que se han usado por otros para detectar VNC. Estos enfoques mencionados anteriormente se basan en la secuenciación de toda o una población estadísticamente significativa de fragmentos de ADN en una muestra, seguido por mapeo de estos fragmentos o de otro modo asociación de los fragmentos a sus cromosomas apropiados. Los fragmentos identificados se comparan entonces se comparan entre sí o frente alguna otra referencia (por ejemplo, estructura cromosómica normal) para determinar VNC en cromosomas particulares. Estos métodos son inherentemente ineficaces en comparación con la presente invención, puesto que los cromosomas primarios de interés sólo constituyen una minoría de los datos que se generan a partir de la detección de tales fragmentos de ADN en las muestras mixtas.

30 El método de la presente invención proporciona detección dirigida de locus seleccionados, lo que proporciona información sobre tanto el contenido del locus seleccionado (es decir, presencia de una región polimórfica) e información sobre la frecuencia del locus seleccionado en una muestra (con o sin detectar ningún polimorfismo posible en esa región). Esta característica clave proporciona la capacidad de detectar tanto el número de copias de locus seleccionados como la presencia o ausencia de polimorfismos en locus seleccionados como un único conjunto de datos a partir del rendimiento de un ensayo multiplexado de la invención.

35 Las técnicas que dependen de un muestreo muy amplio de ADN en una muestra proporcionan una cobertura muy amplia del ADN analizado, pero están de hecho muestreando el ADN contenido dentro de una muestra en una base 1X o menos (es decir, submuestreo). En cambio, la amplificación de locus seleccionados usados en los presentes ensayos proporciona profundidad de cobertura de locus particulares de interés, y proporciona un "supermuestreo" de tales locus seleccionados con una cobertura de secuencia promedio de preferiblemente 2X o más, más preferiblemente cobertura de secuencia de 100X o más, incluso más preferiblemente cobertura de secuencia de 1000X o más de los locus seleccionados (incluyendo de la una o más fuentes secundarias) presentes en la muestra mixta inicial.

45 Una ventaja clara de la invención es que los productos de amplificación que resultan a partir de los ensayos pueden analizarse usando una variedad de técnicas de detección y cuantificación, incluyendo pero sin limitarse a, técnicas de hibridación, PCR digital y técnicas de determinación de secuenciación de alto rendimiento.

50 Los métodos de la invención proporcionan un uso más eficiente y económico de los datos, y la gran mayoría de secuencias analizadas tras una amplificación de muestra dan como resultado información afirmativa sobre la identidad de secuencia y frecuencia de locus seleccionados en la muestra mixta. Por tanto, al contrario que las técnicas que se basan en secuenciación en paralelo de manera masiva o "recuento" digital aleatorio de regiones del cromosoma y posterior identificación de datos relevantes a partir de tales recuentos, el sistema de ensayo de la invención proporciona un uso mucho más eficaz de la colección de datos que los enfoques aleatorios enseñados por otros en la técnica.

### Métodos de ensayo

60 El método de ensayo de la invención usa un esquema general tal como se describió anteriormente, aunque pueden emplearse muchas configuraciones y variaciones diferentes, unas pocas de las cuales se describen a continuación y más de las cuales se ejemplifican en la solicitud estadounidense n.º 61/371605 presentada el 6 de agosto de 2010.

65 La figura 2 ilustra un primer esquema general para un sistema de ensayo basado en ligación de la invención. Los oligonucleótidos de secuencia fija 201, 203 comprenden regiones de cebador universal 209 y 211, respectivamente, y regiones complementarias al locus seleccionado 205 y 207, respectivamente. Sin embargo, además, el sistema de ensayo en la figura 2 emplea una región de índice de muestra 221 en el primer oligonucleótido de secuencia fija 201.

En determinados aspectos, todos o una porción de las secuencias de los locus seleccionados se detectan directamente usando las técnicas descritas, por ejemplo, mediante determinación de secuencia o técnicas de hibridación. En el ejemplo de la figura 2, se asocia un índice de muestra con el primer oligonucleótido de secuencia fija 201. La detección de los índices puede identificar una secuencia de una muestra específica en un sistema de ensayo muy multiplexado.

En la etapa 202, los oligonucleótidos de secuencia fija 201, 203 se introducen en la etapa 202 en la muestra mixta 200 y se permite que se unan específicamente al locus seleccionado 215. Tras la hibridación, los oligonucleótidos de secuencia fija no hibridados se separan preferiblemente del resto de la muestra genética (mediante, por ejemplo, lavado, no mostrado). Entonces se introduce un oligonucleótido en puente y se permite que se hibride en la etapa 204 con la región del locus 215 entre los oligonucleótidos primero 201 y segundo 203 de secuencia fija. Los oligonucleótidos unidos se ligan en la etapa 206 para crear un ácido nucleico contiguo que se extiende y es complementario al locus de interés. En determinados aspectos de la invención, los oligonucleótidos en puente son de entre 2-45 nucleótidos de longitud. En un aspecto específico, los oligonucleótidos en puente son de entre 3-9 nucleótidos de longitud. En aún otro aspecto específico, los oligonucleótidos son de entre 10-30 nucleótidos de longitud.

Tras la ligación, el producto de ligación se eluye a partir de la molde de ADNg. Se introducen los cebadores universales 217, 219 en la etapa 208 para amplificar los oligonucleótidos ligados primero y segundo de secuencia fija para crear 210 productos de amplificación 223 que comprenden la secuencia del locus de interés. Estos productos 223 se aíslan, se detectan, se identifican y se cuantifican para proporcionar información respecto a la presencia y cantidad de los locus seleccionados en la muestra mixta. Preferiblemente, los productos de amplificación se detectan y se cuantifican mediante determinación de secuencia. En aspectos específicos, es deseable determinar las secuencias de tanto el índice de muestra como los productos de amplificación, por ejemplo, para proporcionar identificación de la muestra así como del locus. Los índices tales como el índice de muestra mostrado en este caso visualizados en la invención pueden asociarse con los primeros oligonucleótidos de secuencia fija, los segundos oligonucleótidos de secuencia fija o ambos. Alternativamente o además, pueden asociarse índices con cebadores que se usan para amplificar los oligonucleótidos ligados primero y segundo de secuencia fija, que también sirve para incorporar índices en los productos de amplificación.

En aspectos preferidos y tal como se muestra en la figura 2, se usan índices representativos de la muestra mixta a partir de la cual puede aislarse un ácido nucleico para identificar la fuente de los locus seleccionados en un sistema de ensayo multiplexado. En tales aspectos, los ácidos nucleicos se identifican de manera única con el índice de muestra. Entonces pueden combinarse oligonucleótidos identificados de manera única en un único recipiente de reacción con ácidos nucleicos de otras muestras mixtas antes de la secuenciación. En un caso de este tipo, los datos de secuenciación los segrega el índice de muestra único para determinar la frecuencia de cada locus diana para cada muestra mixta y para determinar si hay una anomalía cromosómica en una muestra individual.

En aspectos de la invención que usan índices de muestra, los oligonucleótidos de secuencia fija preferiblemente se diseñan de modo que los índices de muestra que comprenden identificar información se ubican entre las regiones de cebador universal 209 y 211 y las regiones complementarias a los locus seleccionados en la muestra 205 y 207. Alternativamente, los índices y las secuencias de amplificación universal pueden añadirse a los oligonucleótidos ligados primero y segundo de secuencia fija (y el oligonucleótido en puente, si está presente) incluyendo estos índices en los cebadores usados para amplificar los productos de ligación para muestras diferentes. En cualquier caso, los índices se codifican preferiblemente en el sentido de 5' de las secuencias específicas del locus pero en el sentido de 3' de los cebadores universales de manera que se conservan tras la amplificación.

La figura 3 ejemplifica métodos del sistema de ensayo en el que uno o más oligonucleótidos en puente se emplean y ejemplifica cómo pueden detectarse e identificarse polimorfismos. En la figura 3, se usan dos conjuntos fijos de secuencia oligonucleótidos que comprenden sustancialmente los mismos cebadores universales 309, 311 y las regiones específicas de secuencia 305, 307, pero comprenden diferentes índices de muestra, 321, 323 en los oligonucleótidos de secuencia fija del conjunto donde los diferentes índices corresponden a diferentes secuencias de base para el único polimorfismo de nucleótido presente en una muestra particular. Las reacciones de ligación se llevan a cabo con material de la misma muestra mixta 300, pero en tubos diferentes con los diferentes conjuntos de oligonucleótido específicos de alelo. Se usan los oligonucleótidos en puente correspondientes a dos posibles nucleótidos para este SNP en los locus seleccionados 313, 333 para detectar el locus seleccionado en cada reacción de ligación. Se incorporan dos índices de alelo 321, 323 que son indicativos de los alelos polimórficos particulares en los productos de amplificación de modo que la determinación de secuencia de la secuencia real de los oligonucleótidos ligados primero y segundo en puente no se necesitan necesariamente, aunque las secuencias de todos los productos de ligación pueden todavía determinarse para identificar polimorfismos y/o proporcionar confirmación.

Cada uno de los oligonucleótidos de secuencia fija comprende una región complementaria al locus seleccionado 305, 307, y regiones de cebador universal 309, 311 usadas para amplificar los diferentes locus seleccionados tras la selección y/o aislamiento iniciales de los locus seleccionados de la muestra mixta. Las regiones de cebador universal se ubican en los extremos de los oligonucleótidos de secuencia fija 301, 303 y 323 que flanquean los

índices y las regiones complementarias al ácido nucleico de interés, conservando así las secuencias específicas de ácido nucleico y los índices de muestra en los productos de cualquier método de amplificación universal. Los oligonucleótidos de secuencia fija 301, 303, 323 se introducen en la etapa 302 a una alícuota de la muestra genética 300 y se permite que se unan específicamente a los locus seleccionados 315 ó 325. Tras la hibridación, los oligonucleótidos de secuencia fija no hibridados se separan preferiblemente del resto de la muestra genética mediante, por ejemplo, lavado (no mostrado).

Los oligonucleótidos en puente correspondientes a un A/T SNP 313 o un G/C SNP 333 se introducen y se permite que se unan en la etapa 304 a la región del locus seleccionado 315 ó 325 entre las regiones complementarias de ácido nucleico primera 305 y segunda 307 de los oligonucleótidos de secuencia fija. Alternativamente, los oligonucleótidos en puente 313, 333 pueden introducirse en la muestra simultáneamente con los oligonucleótidos de secuencia fija. Los oligonucleótidos unidos se ligan en la etapa 306 en la única mezcla de reacción para crear una expansión de ácido nucleico contigua y complementaria al locus seleccionado.

Tras la ligación, las diferentes reacciones pueden combinarse preferiblemente para las etapas de amplificación universal y detección. Se introducen los cebadores universales 317, 319 en las reacciones combinadas en la etapa 308 para amplificar las regiones de molde ligadas y crear en la etapa 310 oligonucleótidos ligados primero y segundo de secuencia fija y productos de oligonucleótido en puente 327, 329 que comprenden la secuencia del locus seleccionado que representa ambos SNP en el locus seleccionado. Estos productos de ligación 327, 329 se detectan y se cuantifican mediante determinación de secuencia del producto de ligación, a través del índice de muestra y/o la región del producto que contiene el SNP en el locus seleccionado.

En una configuración alternativa de los métodos de los sistemas de ensayo de la invención, el oligonucleótido en puente puede hibridarse con una región que no es directamente adyacente a la región complementaria a uno o ambos de los oligonucleótidos de secuencia fija, y es necesaria una etapa intermedia que requiere extensión de uno o más de los oligonucleótidos antes de la ligación. Por ejemplo, tal como se ilustra en la figura 4, cada conjunto de oligonucleótidos contiene preferiblemente dos oligonucleótidos 401, 403 de secuencia fija y uno o más oligonucleótidos en puente 413. Cada uno de los oligonucleótidos de secuencia fija comprende una región complementaria al locus seleccionado 405, 407, y secuencias de cebador, preferiblemente secuencias de cebador universal, 409, 411, es decir, regiones de oligonucleótido complementarias a cebadores universales. Las secuencias de cebador 409, 411 se ubican en o cerca de los extremos de los oligonucleótidos de secuencia fija 401, 403, y por tanto, conservan las secuencias específicas de ácido nucleico en los productos de cualquier método de amplificación universal. Los oligonucleótidos de secuencia fija 401, 403 se introducen en la etapa 402 a la muestra mixta 400 y se permite que se unan específicamente a las porciones complementarias del locus de interés 415. Tras la hibridación, los oligonucleótidos de secuencia fija no hibridados se separan preferiblemente del resto de la muestra genética (no mostrado).

El oligonucleótido en puente se introduce entonces en la etapa 404 y se permite que se una a la región del locus seleccionado 415 entre los oligonucleótidos primero 401 y segundo 403 de secuencia fija. Alternativamente, el oligonucleótido en puente puede introducirse simultáneamente con los oligonucleótidos de secuencia fija. En este aspecto a modo de ejemplo, el oligonucleótido en puente se hibrida con una región directamente adyacente a la primera región de oligonucleótido de secuencia fija 405, pero se separa mediante uno o más nucleótidos de la región complementaria del segundo oligonucleótido de secuencia fija 407. Tras la hibridación de la secuencia fija y los oligonucleótidos en puente, el oligonucleótido en puente 413 se extiende en la etapa 406, por ejemplo, usando una polimerasa y dNTP, para rellenar el hueco entre el oligonucleótido en puente 413 y el segundo oligonucleótido de secuencia fija 403. Tras la extensión, los oligonucleótidos unidos se ligan en la etapa 408 para crear una expansión de ácido nucleico contigua y complementaria al locus de interés 415. Tras la ligación se introducen, cebadores universales 417, 419 en la etapa 410 para amplificar los oligonucleótidos ligados primero y segundo en puente para crear en la etapa 412 productos de amplificación 423 que comprenden la secuencia del locus seleccionado de interés. Los productos de amplificación 423 se aíslan, se detectan y se cuantifican opcionalmente para proporcionar información sobre la presencia y cantidad del/de los locus seleccionado(s) en la muestra mixta.

#### **Detección de variaciones en el número de copias**

La presente invención proporciona métodos para identificar una variación en el número de copias en uno o más locus y la presencia o ausencia de uno o más polimorfismos. Esto puede realizarse usando métodos de amplificación para la identificación de locus correspondientes a cromosomas específicos de locus de interés correspondientes a secuencias de genes individuales.

Los sistemas de ensayo usan sondas de ácido nucleico diseñadas para identificar, y preferiblemente aislar, locus seleccionados en una muestra mixta. Determinadas sondas identifican secuencias de interés en locus seleccionados interrogados para el número de copias (es decir, frecuencia de locus), y otras sondas identifican secuencias que corresponden a polimorfismos de interés (es decir, contenido de locus) en ácidos nucleicos correspondientes a una fuente principal o fuente secundaria en una muestra mixta.

En aspectos específicos, el método de ensayo de la invención emplea una o más etapas de amplificación selectiva

- (por ejemplo, usando uno o más cebadores que se hibridan específicamente con un locus seleccionado) para aislar, amplificar o analizar sustancialmente todos los locus seleccionados analizados. Esto está en contraste directo con el enfoque de amplificación aleatorio usado por otros que emplea, por ejemplo, secuenciación en paralelo de manera masiva, puesto que tales técnicas de amplificación implican generalmente amplificación aleatoria de todo o una
- 5 porción sustancial del genoma. Además, la muestra inicial puede enriquecerse opcionalmente usando métodos tales como amplificación general para aumentar el número de copias de ácidos nucleicos en la muestra mixta. Preferiblemente, las etapas de hibridación, ligación y amplificación usadas para identificar los locus de interés se realizan directamente sobre la muestra mixta.
- 10 En un aspecto general, el usuario de la invención analiza múltiples locus seleccionados en diferentes cromosomas. Cuando se analizan múltiples locus para una muestra, una realización preferida es amplificar todos los locus seleccionados para cada muestra en un recipiente de reacción. Las frecuencias de los múltiples locus seleccionados se usan para determinar si existe una anomalía cromosómica, y opcionalmente identificar polimorfismos en los locus seleccionados.
- 15 En aspectos preferidos, pueden amplificarse múltiples locus seleccionados de dos o más muestras en un único recipiente de reacción, y analizarse la información simultáneamente en un único conjunto de datos, por ejemplo, mediante determinación de secuencia. Entonces se analizan los datos resultantes para separar los resultados para la muestra diferente y se usan para determinar la presencia de ausencia de VNC y la contribución de fuente en
- 20 muestras individuales.
- En un aspecto, se identifican anomalías cromosómicas en el método de ensayo de la invención usando múltiples locus seleccionados en múltiples cromosomas, y se compara la frecuencia de los locus seleccionados en los múltiples cromosomas para identificar un aumento de probabilidad de aneuploidía basándose en las frecuencias de
- 25 razón de los múltiples locus en los cromosomas. Puede realizarse normalización o estandarización de las frecuencias para uno o más locus seleccionados.
- En otro aspecto, el sistema de ensayo suma las frecuencias de los locus seleccionados en dos o más cromosomas y entonces compara la suma de los locus seleccionados en un cromosoma frente a otro cromosoma para determinar si existe una aneuploidía cromosómica. En otro aspecto, el sistema de ensayo analiza subconjuntos de locus
- 30 seleccionados frecuencias en dos o más cromosomas para determinar si existe una aneuploidía cromosómica para uno de los dos cromosomas. La comparación puede realizarse o bien dentro del mismo cromosoma o bien diferentes.
- 35 En determinados aspectos, los datos usados para determinar la frecuencia de los locus seleccionados pueden excluir datos atípicos que parecen deberse a error experimental, o que tienen niveles elevados o reducidos basándose en un sesgo genético idiopático dentro de una muestra particular. En un ejemplo, los datos usados para la suma pueden excluir regiones de ADN con una frecuencia particularmente elevada en una o más muestras. En otro ejemplo, los datos usados para la suma pueden excluir locus seleccionados que se encuentran en una
- 40 abundancia particularmente baja en una o más muestras.
- En otro aspecto, pueden elegirse subconjuntos de locus seleccionados para dar un resultado estadísticamente significativo cuando se determina si existe una anomalía cromosómica. Pueden realizarse múltiples análisis de diferentes subconjuntos de locus seleccionados dentro de una muestra mixta para dar más poder estadístico. Por
- 45 ejemplo, si hay 100 locus seleccionados para el cromosoma 21 y 100 locus seleccionados para el cromosoma 18, podría realizarse una serie de análisis que evalúa menos de 100 regiones para cada uno de los cromosomas. En este ejemplo, no se excluyen selectivamente locus seleccionados.
- La cantidad de diferentes ácidos nucleicos detectable en determinados cromosomas puede variar según varios
- 50 factores, incluyendo representación general de locus en diferentes fuentes de células en muestras mixtas, tasas de degradación de los diferentes ácidos nucleicos que representan diferentes locus en muestras mixtas, métodos de preparación de muestra, y similares. Por tanto, en otro aspecto, la cantidad de locus particulares en un cromosoma se suma para determinar la cantidad de locus para diferentes cromosomas en la muestra. Las frecuencias de locus se suman para un cromosoma particular, y la suma de los locus se usa para determinar aneuploidía. Este aspecto
- 55 de la invención suma las frecuencias de los locus individuales en cada cromosoma y entonces compara la suma de los locus en un cromosoma frente a uno o más de otros cromosomas para determinar si existe una anomalía cromosómica.
- Los ácidos nucleicos analizados usando los métodos de ensayo de la invención se amplifican preferiblemente de
- 60 manera selectiva y se aíslan opcionalmente de la muestra mixta usando cebadores específicos para el locus de interés (por ejemplo, para un locus de interés en una muestra mixta). Los cebadores para amplificación selectiva pueden elegirse por diversos motivos, pero están diseñados preferiblemente para 1) amplificar de manera eficaz una región del cromosoma de interés; 2) tener un intervalo predecible de expresión a partir de fuentes maternas y/o
- 65 fetales en diferentes muestras mixtas; y 3) ser distintivos del cromosoma particular, es decir, no amplificar regiones homólogas en otros cromosomas. Las siguientes son técnicas a modo de ejemplo que pueden emplearse en el sistema de ensayo o la invención.

El método de ensayo de la invención detecta tanto aneuploidías como anomalías cromosómicas específicas a través de la identificación y cuantificación de locus de interés específicos. Tales anomalías cromosómicas incluyen, pero no se limitan a, mutaciones de delección, mutaciones de inserción, polimorfismos del número de copias, variantes del número de copias, síndrome de delección del cromosoma 22q11, síndrome de delección 11q en el cromosoma 11, síndrome de delección 8p en el cromosoma 8, y similares. Generalmente, se analizan al menos dos locus seleccionados presentes en los mismos o diferentes cromosomas, y al menos uno de los locus seleccionados se asocia con la anomalía alélica fetal. Las secuencias de los dos locus seleccionados y el número de copias de los dos locus seleccionados se comparan entonces para determinar si la anomalía cromosómica está presente, y si lo está, la naturaleza de la anomalía.

Aunque gran parte de la descripción contenida en el presente documento describe la detección de aneuploidía mediante el recuento de la abundancia de locus seleccionados en uno o más cromosomas aneuploides supuestos y la abundancia de locus seleccionados en uno o más cromosomas normales, pueden usarse las mismas técnicas para detectar variaciones en el número de copias donde tal variación en el número de copias se produce sólo en una porción de un cromosoma. En la detección de variaciones en el número de copias, se comparan múltiples locus seleccionados dentro de la supuesta variación en la ubicación del número de copias con múltiples locus seleccionados fuera de la supuesta variación en la ubicación del número de copias. Por ejemplo, puede detectarse un síndrome de delección del cromosoma 22q11 en un feto en una muestra materna seleccionando dos o más locus dentro de la delección 22q11 y dos o más locus fuera de la delección 22q11. Los locus fuera de la delección 22q11 pueden estar en otra región del cromosoma 22 o puede estar en un cromosoma completamente diferente. La abundancia de cada locus se determina mediante los métodos descritos en esta solicitud.

En algunos aspectos, puede usarse una amplificación universal para amplificar los locus seleccionados. En algunos aspectos, los locus seleccionados para cada muestra se valoran en una única reacción en un único recipiente. En otros aspectos, pueden valorarse locus de múltiples muestras en una única reacción en un único recipiente.

Determinados aspectos de la invención pueden detectar una delección, incluyendo los límites de tales delecciones. En algunos aspectos, pueden usarse al menos 24 locus seleccionados dentro de la región de la supuesta delección y pueden usarse al menos 24 locus seleccionados fuera de la región de la supuesta delección. En otro aspecto, pueden usarse al menos 48 locus seleccionados dentro de la región de la supuesta delección y pueden usarse al menos 48 locus seleccionados fuera de la región de la supuesta delección. En otro aspecto, pueden usarse al menos 96 locus seleccionados dentro de la región de la supuesta delección y pueden usarse al menos 96 locus seleccionados fuera de la región de la supuesta delección. En otro aspecto, pueden usarse al menos 48 locus seleccionados dentro de la región de la supuesta delección y pueden usarse al menos 384 locus seleccionados fuera de la región de la supuesta delección. En un aspecto preferido, pueden usarse al menos 384 locus seleccionados dentro de la región de la supuesta delección y pueden usarse al menos 384 locus seleccionados fuera de la región de la supuesta delección. Se suman los locus seleccionados dentro de la región de la supuesta delección y los locus seleccionados fuera de la región de la supuesta delección. Estas sumas se comparan entonces entre sí para determinar la presencia o ausencia de una delección. Opcionalmente, se calcula una razón de las sumas y esa razón puede compararse con una razón promedio creada a partir de una población normal. Cuando la razón para uno o más locus seleccionados se sitúa estadísticamente fuera de una razón esperada, se detecta una delección. El umbral para identificar positivamente una delección puede ser dos veces o más, preferiblemente cuatro o más veces la variación calculada en la población normal. Cuando se demanda una pluralidad de locus seleccionados dentro y fuera la región de la supuesta delección, pueden identificarse límites de la delección.

### **Polimorfismos asociados con enfermedades o predisposiciones**

El método de ensayo de la invención también puede usarse para detectar polimorfismos, tales como los asociados con una enfermedad o mutación autosómica dominante o recesiva. Dada la naturaleza multiplexada de los sistemas de ensayo de la invención, la detección tiene lugar en el mismo ensayo que la detección de anomalías cromosómicas. Por tanto, un único sistema de ensayo puede proporcionar información de diagnóstico sobre diferentes clases de mutaciones genéticas. Por consiguiente, puesto que los sistemas de ensayo preferidos de la invención están muy multiplexados y pueden interrogar ciento o incluso miles de locus seleccionados dentro de una muestra mixta, en determinados aspectos es deseable interrogar la muestra para loci marcadores dentro de la muestra mixta, por ejemplo, locus asociados con riesgo genético o que identifican la presencia o ausencia de organismos infecciosos. Por tanto, los sistemas de ensayo proporcionan detección de tales loci marcadores junto con la detección de locus seleccionados para la determinación del número de copias en una muestra mixta.

Por tanto, el método de ensayo de la invención puede usarse para detectar polimorfismos en una muestra mixta, donde tales polimorfismos están asociados con genes asociados con trastornos autosómicos recesivos, mutaciones asociadas con trastornos autosómicos dominantes; polimorfismos asociados con riesgo de desarrollar una enfermedad y/o progresión de una enfermedad (por ejemplo, metástasis) e indicadores de pronóstico.

En otros aspectos específicos, el método de ensayo de la invención puede usarse para detectar polimorfismos o mutaciones fetales en una muestra materna, donde tales mutaciones o polimorfismos están asociados con

trastornos poligénicos tales como cardiopatía coronaria, diabetes, hipertensión, defectos cardíacos congénitos y epilepsia. Los ejemplos incluyen mutaciones en genes asociados con predisposiciones tales como mutaciones en genes con susceptibilidad al cáncer, (por ejemplo mutaciones en BRCA1 o II o en p53); polimorfismos asociados con riesgo aumentado de desarrollar enfermedades de aparición tardía, tales como el polimorfismo del gen apoE3 asociado con riesgo de Alzheimer.

Además de la detección de anomalías cromosómicas y mutaciones de un único gen o polimorfismos asociados con enfermedad, trastornos o predisposiciones monogénicos o poligénicos, los sistemas de ensayo de la invención pueden identificar agentes infecciosos en la muestra mixta.

### **Amplificación seleccionada**

Pueden usarse numerosos métodos de amplificación selectiva para proporcionar los ácidos nucleicos amplificados que se analizan en los sistemas de ensayo de la invención, y tales métodos se usan preferiblemente para aumentar el número de copias de locus seleccionados en una muestra mixta de una manera que permite la conservación de información sobre el contenido inicial de los locus seleccionados en la muestra mixta. Aunque no se describen todas las combinaciones de amplificación y análisis en el presente documento en detalle, está bien dentro de la habilidad de los expertos en la técnica usar diferentes métodos de amplificación y/o herramientas analíticas para aislar y/o analizar los ácidos nucleicos de región acorde con esta memoria descriptiva, y tales variaciones serán evidentes para un experto en la técnica tras leer la presente divulgación.

Tales métodos de amplificación incluyen pero no se limitan a, reacción en cadena de la polimerasa (PCR) (patentes estadounidenses n.ºs 4.683.195; y 4.683.202; PCR Technology: Principles and Applications for DNA Amplification, ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), reacción en cadena de ligasa (LCR) (Wu y Wallace, Genomics 4:560, 1989; Landegren *et al.*, Science 241:1077, 1988), amplificación de desplazamiento de la hebra (SDA) (patentes estadounidenses n.ºs 5.270.184; y 5.422.252), amplificación mediada por transcripción (TMA) (patente estadounidense n.º 5.399.491), amplificación lineal unida (LLA) (patente estadounidense n.º 6.027.923), y similares, replicación de secuencia automantenida (Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87, 1874 (1990) y documento WO90/06995), amplificación selectiva de secuencias de polinucleótido diana (patente estadounidense n.º 6.410.276), reacción en cadena de la polimerasa cebada de secuencia de consenso (CP-PCR) (patente estadounidense n.º 4.437.975), reacción en cadena de la polimerasa cebada de manera arbitraria (AP-PCR) (patentes estadounidenses n.ºs 5.413.909, 5.861.245) y secuencia amplificación basada en ácido nucleicos (NASBA). (Véanse las patentes estadounidenses n.ºs 5.409.818, 5.554.517 y 6.063.603. Otros métodos de amplificación que pueden usarse incluyen: Q $\beta$  replicasa, descrita en la solicitud de patente PCT n.º PCT/US87/00880, métodos de amplificación isotérmicos tales como SDA, descritos en Walquer *et al.*, Nucleic Acids Res. 20(7):1691-6 (1992), y amplificación de círculo rodante, descrita en la patente estadounidense n.º 5.648.245. Otros métodos de amplificación que pueden usarse se describen en, las patentes estadounidenses n.ºs 5.242.794, 5.494.810, 4.988.617 y en la solicitud de patente estadounidense con n.º de serie 09/854.317 y la publicación estadounidense n.º 2003014359 9. En algunos aspectos, se amplifica ADN mediante PCR específica del locus múltiple. En un aspecto preferido, el ADN se amplifica usando ligación de adaptador y PCR de cebador único. Otros métodos de amplificación disponibles, tal como PCR equilibrada (Makrigiorgos, *et al.*, Nature Biotechnol, 20:936-9 (2002)) y métodos de amplificación isotérmicos tales como amplificación basada en secuencia de ácidos nucleicos (NASBA) y replicación de secuencia automantenida (Guatelli *et al.*, PNAS USA 87:1874 (1990)). Basándose en tales metodologías, un experto en la técnica puede diseñar fácilmente cebadores en cualquier región adecuada 5' y 3' para un locus de interés. Tales cebadores pueden usarse para amplificar ADN de cualquier longitud siempre que el ADN comprenda los locus seleccionados de interés.

La longitud de un locus amplificado seleccionado de una región genómica de interés es lo suficientemente larga para proporcionar suficiente información de secuencia para distinguir el locus amplificado de otros locus que se amplifican y/o se seleccionan. Generalmente, un ácido nucleico amplificado correspondiente a un locus seleccionado es de al menos aproximadamente 16 nucleótidos de longitud, y más normalmente, un ácido nucleico amplificado correspondiente a un locus seleccionado es de al menos aproximadamente 20 nucleótidos de longitud. En un aspecto preferido de la invención, un ácido nucleico amplificado correspondiente a un locus seleccionado es de al menos aproximadamente 30 nucleótidos de longitud. En un aspecto más preferido de la invención, un ácido nucleico amplificado correspondiente a un locus seleccionado es de al menos aproximadamente 32, 40, 45, 50 ó 60 nucleótidos de longitud. En otros aspectos de la invención, un ácido nucleico amplificado correspondiente a un locus seleccionado puede ser de aproximadamente 100, 150 o hasta 200 de longitud.

En determinados aspectos, amplificación selectiva comprende una etapa de amplificación lineal inicial, que puede ser particularmente útil si la cantidad de partida de ADN de la muestra mixta es bastante limitada, por ejemplo, cuando el ADN libre de células en una muestra está disponible en cantidades limitadas. Este mecanismo aumenta la cantidad de moléculas de ADN que son representativas del contenido de ADN original, y ayuda a reducir el error de muestreo cuando se necesita cuantificación precisa del ADN o una fracción del ADN (por ejemplo, contribución de ADN fetal en una muestra materna).

Por tanto, en un aspecto, se realiza un número limitado de ciclos de amplificación lineal específico de secuencia en

la muestra mixta de partida que comprende ADNlc. El número de ciclos es generalmente menos que el usado para una amplificación de PCR típico, por ejemplo, 5-30 ciclos o menos. Pueden diseñarse cebadores o sondas para amplificar segmentos genómicos específicos o regiones que comprenden locus seleccionados. Los cebadores o sondas pueden modificarse con una etiqueta final en el extremo 5' (por ejemplo, con biotina) o en otro lugar a lo largo del cebador o sonda de modo que los productos de amplificación pueden purificarse o unirse a un sustrato sólido (por ejemplo, perla o matriz) para aislamiento o análisis adicional. En un aspecto preferido, los cebadores se multiplexan de modo que una única reacción proporciona múltiples fragmentos de ADN de diferentes regiones. Entonces pudieron amplificarse adicionalmente productos de amplificación de la amplificación lineal con métodos de PCR convencionales o con amplificación lineal adicional.

Por ejemplo, puede aislarse ADNlc a partir de sangre, plasma o suero de una mujer embarazada, y se incuba con cebadores frente a un número establecido de locus seleccionados que corresponden a cromosomas de interés. Preferiblemente, el número de cebadores usados para amplificación lineal inicial será de 12 o más, más preferiblemente 24 o más, más preferiblemente 36 o más, incluso más preferiblemente 48 o más, e incluso más preferiblemente 96 o más. Cada uno de los cebadores corresponde a un único locus seleccionado, y se etiqueta opcionalmente para identificación y/o aislamiento. Un número limitado de ciclos, preferiblemente 10 o menos, se realizan con amplificación lineal. Los productos de amplificación se aíslan posteriormente, por ejemplo, cuando los cebadores se unen a una molécula de biotina los productos de amplificación pueden aislarse por medio de la unión a avidina o estreptavidina sobre un sustrato sólido. Los productos de amplificación se someten entonces a procesos bioquímicos adicionales tales como amplificación adicional con otros cebadores y/o técnicas de detección tales como determinación de secuencia e hibridación.

Las eficacias de amplificación lineal pueden variar entre sitios y entre ciclos de modo que en determinados sistemas puede usarse normalización para garantizar que los productos de la amplificación lineal son representativos en frecuencia y secuencia de los ácidos nucleicos en la muestra mixta. Un practicante del sistema de ensayo de la invención puede usar información de múltiples alícuotas de una muestra para determinar la variación en la cantidad de diferentes productos de amplificación que representan los locus seleccionados, incluyendo variación en diferentes locus seleccionados y/o entre locus seleccionados tras la amplificación lineal inicial limitada. Tal información puede usarse para determinar los niveles iniciales de locus seleccionados en el contenido de ADN de la muestra, lo que permite la normalización de la frecuencia de locus seleccionados.

### ***Amplificación universal***

En aspectos preferidos de la invención, los locus amplificados selectivamente se amplifican preferiblemente además mediante la amplificación universal de todos o sustancialmente todos los diversos locus seleccionados usando los sistemas de ensayo de la invención. Se añaden regiones de cebador universal a los oligonucleótidos de secuencia fija de modo que los locus amplificados selectivamente pueden amplificarse además una única reacción de amplificación universal. Estas secuencias de cebador universal pueden añadirse a las regiones de ácidos nucleicos durante el proceso de amplificación selectiva, es decir, los cebadores para la amplificación selectiva comprenden secuencias de cebador universal. Alternativamente, pueden añadirse adaptadores que comprenden secuencias de amplificación universal a los extremos de los locus amplificados selectivamente seleccionados como adaptadores tras amplificación inicial y después del aislamiento de los locus amplificados selectivamente seleccionados de la muestra mixta.

En un aspecto a modo de ejemplo, se amplifican inicialmente locus seleccionados de una muestra mixta usando cebadores complementarios a locus seleccionados de interés, seguido por una etapa de amplificación universal para aumentar el número de locus para el análisis. La introducción de regiones de cebador a los productos de amplificación iniciales de una muestra mixta permite amplificación universal controlada posterior de todos o una porción de ácidos nucleicos seleccionados antes de o durante el análisis, por ejemplo, determinación de secuencia.

Pueden introducirse sesgo y variabilidad durante la amplificación de ADN, tal como el observado durante la reacción en cadena de la polimerasa (PCR). En casos en los que se multiplexa una reacción de amplificación, existe la posibilidad de que se amplificarán diferentes locus seleccionados a tasas o eficacia diferentes. Esto puede deberse, en parte, a algunos cebadores en una reacción múltiple que tienen mejor eficacia (es decir, cinética de hibridación más favorable), debido a condiciones experimentales que favorecen algunos cebadores respecto a otros tales como contenido de secuencia del cebador y ADN de molde, condiciones de tampón y otras condiciones. La amplificación universal de ADN en un sistema de ensayo multiplexado introduce generalmente menos sesgo y variabilidad entre locus amplificados.

Por consiguiente, en un aspecto, se realiza un pequeño número (por ejemplo, 1-10, preferiblemente 3-5) de ciclos de amplificación seleccionada usando secuencias específicas de locus, seguido por amplificación universal usando cebadores universales. El número de ciclos usando cebadores universales variará, pero será preferiblemente de al menos 10 ciclos, más preferiblemente de al menos 5 ciclos, incluso más preferiblemente de 20 ciclos o más. Al pasar a amplificación universal tras un número menor de ciclos de amplificación, se reduce el sesgo de que determinados locus se amplifiquen a tasas mayores que otros.

Opcionalmente, el sistema de ensayo incluirá una etapa entre el proceso de amplificación seleccionada y el proceso de amplificación universal para retirar cualquier ácido nucleico que no se amplifique selectivamente en la reacción de amplificación selectiva.

5 Puede usarse todo el producto o una alícuota del producto a partir de la amplificación seleccionada para la reacción de amplificación universal. Pueden usarse las mismas o diferentes condiciones (por ejemplo, polimerasa, tampones y similares) en las etapas de amplificación, por ejemplo, para garantizar que el sesgo y la variabilidad no se introducen involuntariamente debido a las condiciones experimentales. Además, pueden usarse variaciones en las concentraciones de cebador para limitar diferencialmente el número de ciclos de amplificación específicos de  
10 secuencia para algunos locus seleccionados en comparación con otros locus seleccionados.

En determinados aspectos, las regiones de cebador universal de los cebadores o adaptadores usados en el sistema de ensayo se diseñan para ser compatibles con métodos de ensayo multiplexados convencionales que usan mecanismos de cebado generales para analizar grandes números de ácidos nucleicos simultáneamente en una  
15 reacción en un recipiente. Tales métodos de cebado "universales" permiten el análisis eficaz, de gran volumen de la cantidad de locus seleccionados presentes en una muestra mixta, y permiten una cuantificación comprensiva de la presencia de locus seleccionados dentro de una muestra mixta de este tipo para la determinación de aneuploidía.

Los ejemplos de tales método de ensayo incluyen, pero no se limitan a, métodos de multiplexación usados para  
20 amplificar y/o genotipar una variedad de muestras simultáneamente, tal como los descritos en Oliphant *et al.*, patente estadounidense n.º 7.582.420.

Algunos aspectos usan reacciones acopladas para la detección múltiple de secuencias de ácido nucleico donde oligonucleótidos de una fase temprana de cada proceso contienen secuencias que pueden usarse por  
25 oligonucleótidos de una fase tardía del proceso. Los procesos a modo de ejemplo para amplificar y/o detectar ácidos nucleicos en muestras pueden usarse solos o en combinación, incluyendo pero sin limitarse a los métodos descritos a continuación, cada uno de los cuales se incorporan como referencia en su totalidad.

En determinados aspectos, el sistema de ensayo de la invención usa una de las siguientes técnicas de amplificación  
30 universales y selectivas combinadas: (1) reacción de detección de ligasa ("LDR") acoplada con reacción en cadena de la polimerasa ("PCR"); (2) PCR primaria acoplada a PCR secundaria acoplada a LDR; y (3) PCR primaria acoplada a PCR secundaria. Cada uno de estos aspectos de la invención tiene una aplicabilidad particular en la detección de determinadas características de ácido nucleico. Sin embargo, cada uno requiere el uso de reacciones acopladas para la detección múltiple de diferencias de secuencias de ácidos nucleicos donde los oligonucleótidos de  
35 una fase temprana de cada proceso contienen secuencias que pueden usarse mediante oligonucleótidos de una fase tardía del proceso.

Barany *et al.*, patentes estadounidenses n.ºs 6.852.487, 6.797.470, 6.576.453, 6.534.293, 6.506.594, 6.312.892, 6.268.148, 6.054.564, 6.027.889, 5.830.711, 5.494.810, describen el uso del ensayo de reacción en cadena de  
40 ligasa (LCR) para la detección de secuencias de nucleótidos específicas en una variedad de muestras de ácido nucleico.

Barany *et al.*, patentes estadounidenses n.ºs 7.807.431, 7.455.965, 7.429.453, 7.364.858, 7.358.048, 7.332.285, 7.320.865, 7.312.039, 7.244.831, 7.198.894, 7.166.434, 7.097.980, 7.083.917, 7.014.994, 6.949.370, 6.852.487, 6.797.470, 6.576.453, 6.534.293, 6.506.594, 6.312.892 y 6.268.148 describen PCR acoplada a LDR para la  
45 detección de ácidos nucleicos.

Barany *et al.*, patentes estadounidenses n.ºs 7.556.924 y 6.858.412, describen el uso de sondas precírculo (también denominadas "sondas candado" o "sondas de inversión múltiple") con LDR acoplada y reacción en cadena de la  
50 polimerasa ("PCR") para la detección de ácidos nucleicos.

Barany *et al.*, patentes estadounidenses n.ºs 7.807.431, 7.709.201 y 7.198.814 describen el uso de escisión de endonucleasas combinada y reacciones de ligación para la detección de secuencias de ácido nucleico.

55 Willis *et al.*, patentes estadounidenses n.ºs 7.700.323 y 6.858.412, describen el uso de sondas precírculo en la amplificación, detección y genotipado de ácidos nucleicos multiplexados.

Ronaghi *et al.*, patente estadounidense n.º 7.622.281 describen técnicas de amplificación para marcar y amplificar un ácido nucleico usando un adaptador que comprende un cebador único y un código de barras.

60 En un aspecto preferido, los productos de amplificación se multiplexan, tal como se describió anteriormente. En un aspecto preferido, los productos de amplificación múltiples se cuantifican mediante análisis de los productos de amplificación. En un aspecto preferido, se aísla una muestra de representación de moléculas individuales de los procesos de amplificación del resto de la muestra para análisis adicional. Para obtener una muestra de  
65 representación de moléculas individuales, el número promedio de moléculas por locus debe superar el ruido de muestreo creado por la reacción multiplexada. En un aspecto, el número promedio por locus es superior a 100. En

otro aspecto, el número promedio por locus es superior a 500. En otro aspecto el número promedio por locus es superior a 1000.

Se aíslan preferiblemente de manera física moléculas individuales del producto de amplificación de las otras moléculas de una manera que permite que los diferentes productos de amplificación se distingan entre sí en un análisis. En un aspecto preferido, este aislamiento se produce sobre un sustrato sólido. Cada molécula aislada puede asociarse con una dirección identificable o física particular o bien antes del análisis, o bien la dirección puede hacerse conocida para los productos de amplificación particulares basándose en el resultado de los análisis. El sustrato puede ser una superficie plana o una superficie tridimensional tal como una perla.

Una vez aislado, el producto de amplificación individual puede amplificarse adicionalmente para hacer múltiples copias idénticas de esa molécula en la misma ubicación conocida o identificable. La amplificación puede producirse antes o después de que la ubicación se vuelva una dirección identificable o física. El producto de amplificación y/o sus copias (que pueden ser idénticas o complementarias al producto de amplificación) se analizan entonces basándose en la secuencia del producto de amplificación o sus copias para identificar el locus particular y/o alelo que representa.

En un aspecto preferido, toda la longitud del producto de amplificación o una porción del producto de amplificación puede analizarse usando determinación de secuencia. El número de bases que han de determinarse debe ser suficiente para identificar de manera única el producto de amplificación como que pertenece a un locus específico y/o alelo. En un aspecto preferido, el locus se analiza mediante determinación de secuencia del producto de amplificación.

Numerosos métodos de determinación de secuencia son compatibles con los sistemas de ensayo de las invenciones. Los métodos a modo de ejemplo para la determinación de secuencia incluyen, pero no se limitan a, métodos basados en hibridación, tales como los dados a conocer en Drmanac, patentes estadounidenses n.ºs 6.864.052; 6.309.824; y 6.401.267; y Drmanac *et al.*, publicación de patente estadounidense 2005/0191656, secuenciación mediante métodos de síntesis, por ejemplo, Nyren *et al.*, patente estadounidense n.º 7.648.824, 7.459.311 y 6.210.891; Balasubramanian, patentes estadounidenses n.ºs 7.232.656 y 6.833.246; Quake, patente estadounidense n.º 6.911.345; Li *et al.*, Proc. Natl. Acad. Sci., 100: 414-419 (2003); secuenciación de pirofosfato tal como se describe en Ronaghi *et al.*, patentes estadounidenses n.ºs 7.648.824, 7.459.0311, 6.828.100 y 6.210.891; y métodos de determinación de secuenciación basados en ligación, por ejemplo, Drmanac *et al.*, solicitud de patente estadounidense n.º 20100105052, y Church *et al.*, solicitudes de patente estadounidense n.ºs 20070207482 y 20090018024.

Puede determinarse información de secuencia usando métodos que determinan muchas (normalmente de cientos a miles de millones) de secuencias de ácido nucleico de manera intrínsecamente paralela, donde muchas secuencias se leen preferiblemente en paralelo usando un procedimiento en serie de alto rendimiento. Tales métodos incluyen pero no se limitan a pirosecuenciación (por ejemplo, tal como se comercializa por 454 Life Sciences, Inc., Branford, CT); secuenciación mediante ligación (por ejemplo, tal como se comercializa en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, CA); secuenciación mediante síntesis usando nucleótidos modificados (tales como los comercializados en la tecnología TruSeq™ y HiSeq™ por Illumina, Inc., San Diego, CA, HeliScope™ por Helicos Biosciences Corporation, Cambridge, MA, y PacBio RS por Pacific Biosciences of California, Inc., Menlo Park, CA); secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, CA); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain Vista, CA); tecnologías de secuenciación basadas en nanoporos (por ejemplo, tal como la desarrollada por Oxford Nanopore Technologies, LTD, Oxford, R.U.), y métodos de secuenciación altamente paralelizados similares.

Alternativamente, en otro aspecto, puede analizarse toda la longitud del producto de amplificación o una porción del producto de amplificación usando técnicas de hibridación. Métodos para llevar a cabo ensayos de hibridación de polinucleótidos para la detección de se han desarrollado bien en la técnica. Los procedimientos y condiciones del ensayo de hibridación variarán según la aplicación y se seleccionan según los métodos de unión generales conocidos incluyendo aquellos a los que se hace referencia en: Maniatis *et al.* Molecular Cloning: A Laboratory Manual (2ª ed. Cold Spring Harbor, N.Y., 1989); Berger y Kimmel Methods in Enzymology, vol. 152, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young y Davis, P.N.A.S, 80: 1194 (1983). Se han descrito Métodos y aparatos para llevar a cabo reacciones de hibridación repetidas y controladas en las patentes estadounidenses n.ºs 5.871.928, 5.874.219, 6.045.996 y 6.386.749, 6.391.623.

La presente invención también contempla la detección de señales de hibridación entre ligandos en determinados aspectos preferidos. Véanse las patentes estadounidenses n.ºs 5.143.854, 5.578.832; 5.631.734; 5.834.758; 5.936.324; 5.981.956; 6.025.601; 6.141.096; 6.185.030; 6.201.639; 6.218.803; y 6.225.625, en la solicitud de patente estadounidense 60/364.731 y en la solicitud PCT PCT/US99/06097 (publicada como el documento WO99/47964).

Se dan a conocer métodos y aparatos para la detección de señales y procesamiento de datos de intensidad en, por ejemplo, las patentes estadounidenses n.ºs 5.143.854, 5.547.839, 5.578.832, 5.631.734, 5.800.992, 5.834.758; 5.856.092, 5.902.723, 5.936.324, 5.981.956, 6.025.601, 6.090.555, 6.141.096, 6.185.030, 6.201.639; 6.218.803; y

6.225.625, en solicitud de patente estadounidense 60/364.731 y en la solicitud PCT PCT/US99/06097 (publicada como el documento WO99/47964).

### Minimización de variaciones dentro de y entre muestras

5 Un desafío con la detección de anomalías cromosómicas mediante detección en una muestra mixta es que los ácidos nucleicos de la fuente secundaria pueden estar presentes en una abundancia mucho menor que los ácidos nucleicos de la fuente principal normal. En el caso de una muestra materna que contiene ADN fetal y materno, el ADN fetal libre circulante como porcentaje del ADN total puede variar entre menos del uno y el cuarenta por ciento, y más comúnmente está presente a menos del veinte por ciento y frecuentemente a o menos del diez por ciento. En la detección de una aneuploidía tal como trisomía 21 (síndrome de Down) en el ADN fetal de tal muestra materna, el aumento relativo en el cromosoma 21 es del 50% en el ADN fetal y, por tanto, como porcentaje del ADN total en una muestra materna donde, como ejemplo, el ADN fetal es del 5% del total, el aumento en el cromosoma 21 como porcentaje del total es del 2,5%. Si va a detectarse esta diferencia de manera robusta mediante los métodos descritos en el presente documento, la variación en la medición del cromosoma 21 ha de ser mucho menor que el aumento de porcentaje del cromosoma 21 contribuido por un tercer cromosoma 21 del ADN fetal.

20 La variación entre niveles encontrados entre muestras y/o para los locus dentro de una muestra puede minimizarse usando una combinación de métodos analíticos, muchos de los cuales se describen en esta solicitud. Por ejemplo, la variación se reduce usando una referencia interna en el ensayo. Un ejemplo de una referencia interna es el uso de un cromosoma presente en una abundancia "normal" (por ejemplo, disomía para un autosoma) para comparar frente a un cromosoma presente en abundancia supuestamente anómala, tal como aneuploidía, en la misma muestra. Aunque el uso de un cromosoma "normal" de este tipo como cromosoma de referencia puede ser suficiente, también es posible usar de dos a muchos cromosomas normales como cromosomas de referencia interna para aumentar el poder estadístico de la cuantificación.

30 Un método de uso de una referencia interna es calcular una razón de abundancia de los cromosomas supuestamente anómalos con respecto a la abundancia de los cromosomas normales en una muestra, denominada razón cromosómica. En el cálculo de la razón cromosómica, la abundancia o recuentos de cada uno de los locus seleccionados para cada cromosoma se suman juntos para calcular los recuentos totales para cada cromosoma. Los recuentos totales para un cromosoma se dividen entonces entre los recuentos totales para un cromosoma diferente para crear una razón cromosómica para esos dos cromosomas.

35 Alternativamente, puede calcularse una razón cromosómica para cada cromosoma sumando en primer lugar los recuentos de cada uno de los locus seleccionados para cada cromosoma, y después dividiendo la suma para un cromosoma entre la suma total para dos o más cromosomas. Una vez calculada, la razón cromosómica se compara entonces con la razón cromosómica promedio de una población normal.

40 El promedio puede ser la media, la mediana, la moda u otro promedio, con o sin normalización y exclusión de datos atípicos. En un aspecto preferido, se usa la media. En el desarrollo de los datos para la razón cromosómica de la población normal, se calcula la variación normal de los cromosomas medidos. Esta variación puede expresarse de varias formas, más normalmente como coeficiente de variación, o CV. Cuando la razón cromosómica de la muestra se compara con la razón cromosómica promedio de una población normal, si la razón cromosómica para la muestra se sitúa estadísticamente fuera de la razón cromosómica promedio para la población normal, la muestra contiene una aneuploidía. Los criterios para establecer el umbral estadístico para declarar una aneuploidía dependen de la variación en la medición de la razón cromosómica y las tasas de falsos positivos y falsos negativos aceptables para el ensayo. En general, este umbral puede ser un múltiplo de la variación observada en la razón cromosómica. En un ejemplo, este umbral es tres o más veces la variación de la razón cromosómica. En otro ejemplo, es cuatro o más veces la variación de la razón cromosómica. En otro ejemplo, es cinco o más veces la variación de la razón cromosómica. En el ejemplo anterior, la razón cromosómica se determina mediante la suma de los recuentos de locus seleccionados por el cromosoma. Normalmente, se usa el mismo número de locus seleccionados para cada cromosoma. Un método alternativo para generar la razón cromosómica sería calcular los recuentos promedio para los locus para cada cromosoma. El promedio puede ser cualquier estimación de la media, la mediana o la moda, aunque normalmente se usa un promedio. El promedio puede ser la media de todos los recuentos o alguna variación tal como un promedio recortado o ponderado. Una vez se han calculado los recuentos promedio para cada cromosoma, los recuentos promedio para cada cromosoma pueden dividirse entre el otro para obtener una razón cromosómica entre dos cromosomas, los recuentos promedio para cada cromosoma pueden dividirse entre la suma de los promedios para todos los cromosomas medidos para obtener una razón cromosómica para cada cromosoma tal como se describió anteriormente. Tal como se destacó anteriormente, la capacidad de detectar una aneuploidía en una muestra materna en la que el ADN supuesto está en una abundancia relativa baja depende enormemente de la variación en las mediciones de diferentes locus seleccionados en el ensayo. Pueden usarse numerosos métodos analíticos que reducen esta variación y, por tanto, mejoran la sensibilidad de este método para detectar una aneuploidía. Un método para reducir la variabilidad del ensayo es aumentar el número de locus seleccionados usados para calcular la abundancia de los cromosomas. En general, si la variación medida de un único locus seleccionado de un cromosoma es el X% e Y locus diferentes seleccionados se miden en el mismo cromosoma, la variación de la

medición de la abundancia cromosómica calculada sumando o promediando la abundancia de cada locus seleccionado en ese cromosoma será aproximadamente el  $X\%$  dividido entre  $Y^{1/2}$ . Dicho de otra manera, la variación de la medición de la abundancia de los cromosomas sería aproximadamente la variación promedio de la medición de la abundancia de cada locus seleccionado dividida entre la raíz cuadrada del número de locus.

5 En un aspecto preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de al menos 24. En otro aspecto preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de al menos 48. En otro aspecto preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de al menos 100. En otro aspecto preferido de esta invención el número de locus seleccionados medidos para cada cromosoma es de al menos 200. Existe un coste progresivo para la medición de cada locus y, por tanto, es importante minimizar el número de locus seleccionados. En un aspecto preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de menos de 2000. En un aspecto preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de menos de 1000. En un aspecto más preferido de esta invención, el número de locus seleccionados medidos para cada cromosoma es de al menos 48 y menos de 1000. En un aspecto, tras la medición de la abundancia para cada locus seleccionado, puede usarse un subconjunto de los locus seleccionados para determinar la presencia o ausencia de aneuploidía. Existen muchos métodos convencionales para elegir el subconjunto de locus seleccionados. Estos métodos incluyen exclusión de resultados atípicos, donde los locus seleccionados con niveles detectados por debajo y/o por encima de un determinado percentil se descartan del análisis. En un aspecto, el percentil puede ser el 5% más bajo y el más alto tal como se mide mediante la abundancia. En otro aspecto, el percentil puede ser el 10% más bajo y el más alto tal como se mide mediante la abundancia. En otro aspecto, el percentil puede ser el 25% más bajo y el más alto tal como se mide mediante la abundancia.

25 Otro método para elegir el subconjunto de locus seleccionados incluye la eliminación de regiones que se sitúan fuera de algún límite estadístico. Por ejemplo, pueden retirarse del análisis locus seleccionados que se sitúan fuera de una o más desviaciones estándar de la abundancia media. Otro método para elegir el subconjunto de locus seleccionados puede ser comparar la abundancia relativa de un locus seleccionado con la abundancia esperada del mismo locus seleccionado en una población sana y descartar cualquier locus seleccionado que no pase la prueba de expectativas. Para minimizar adicionalmente la variación en el ensayo, puede aumentarse el número de veces que se mide cada locus seleccionado. Tal como se comentó, a diferencia de los métodos aleatorios de detección de aneuploidía en los que el genoma se mide en promedio menos de una vez, los sistemas de ensayo de la presente invención miden de manera intencionada cada locus seleccionado múltiples veces. En general, cuando se cuentan acontecimientos, la variación en el recuento se determina mediante estadística de Poisson, y la variación de recuento es normalmente igual a una dividida entre la raíz cuadrada del número de recuentos. En un aspecto preferido de la invención, los locus seleccionados se miden cada uno en promedio al menos 100 veces. En un aspecto preferido de la invención, los locus seleccionados se miden cada uno en promedio al menos 500 veces. En un aspecto preferido de la invención, los locus seleccionados se miden cada uno en promedio al menos 1000 veces. En un aspecto preferido de la invención, los locus seleccionados se miden cada uno en promedio al menos 2000 veces. En un aspecto preferido de la invención, los locus seleccionados se miden cada uno en promedio al menos 5000 veces.

45 En otro aspecto, los subconjuntos de locus pueden elegirse aleatoriamente pero con números suficientes de locus seleccionados para dar un resultado estadísticamente significativo en la determinación de si existe una anomalía cromosómica. Pueden realizarse múltiples análisis de diferentes subconjuntos de locus dentro de una muestra mixta para dar más poder estadístico. En este ejemplo, puede o no ser necesario retirar o eliminar cualquier locus seleccionado antes del análisis aleatorio. Por ejemplo, si hay 100 locus seleccionados para el cromosoma 21 y 100 locus seleccionados para el cromosoma 18, podría realizarse una serie de análisis que evalúen menos de 100 locus por cada uno de los cromosomas.

50 Además de los métodos anteriores para reducir la variación en el ensayo, pueden usarse en combinación otras técnicas analíticas, muchas de las cuales se describen anteriormente en esta solicitud. En general, la variación en el ensayo puede reducirse cuando todos los locus seleccionados para cada muestra se interrogan en una única reacción en un único recipiente. De manera similar, la variación en el ensayo puede reducirse cuando se usa un sistema de amplificación universal. Además, la variación del ensayo puede reducirse cuando se limita el número de ciclos de amplificación.

*Uso de sistemas de ensayo para la detección en muestras mixtas de pacientes con cáncer*

60 El sistema de ensayo permite la detección de alteraciones específicas de tumor cuantitativas y cualitativas de ADN, tales como integridad de las hebras de ADN, frecuencia de mutaciones, anomalías de microsatélites y metilación de genes, como marcadores de diagnóstico, pronóstico y monitorización en pacientes con cáncer. La capacidad para combinar tal detección de alteraciones de genes individuales (incluyendo mutaciones puntuales, indels y variación en el número de copias) con detección de VNC proporciona un método poderoso para ayudar con el diagnóstico clínico, tratamientos, predicción del desenlace y monitorización de la progresión en pacientes con o que se sospecha que tienen un tumor maligno.

- En algunos aspectos, el sistema de ensayo de la invención se usa para fines de diagnóstico por ejemplo, para detectar la presencia y/o naturaleza de un tumor maligno en un paciente o para proporcionar una estimación cuantitativa de la carga tumoral en un paciente. Se han asociado microARN y ADN tumorales circulantes con determinados cánceres, tales como cáncer de pulmón (Roth C *et al.*, Mol Oncol. Junio de 2011; 5(3):281-91. Publicación electrónica 24 de febrero de 2011). También se han detectado variaciones en el número de copias en determinados cánceres, tales como receptor de estrógenos y HER2 amplificado en el ADNlc en pacientes con cáncer de mama. (Page K., Br J Cancer. 12 de abril de 2011; 104(8):1342-8. Publicación electrónica 22 de marzo de 2011).
- En otros aspectos de la invención, el sistema de ensayo se usa en pacientes con cáncer para monitorizar una respuesta al tratamiento y/o seguir la progresión de la enfermedad, por ejemplo, para medir alteraciones génicas individuales y ADNlc en pacientes que reciben quimioterapia (CRT). Para determinados cánceres, se ha mostrado que el índice de integridad del ADNlc puede estar asociado significativa e independientemente con la respuesta del tumor al tratamiento. Agostini M *et al.*, Ann Surg Oncol. 17 de marzo de 2011. Además, la presencia o ausencia de determinadas alteraciones genéticas y/o diferencias en la variación en el número de copias se ha asociado con la respuesta a la quimioterapia y/o el pronóstico de una enfermedad. Véase, por ejemplo, Savas S., Acta Oncol. Noviembre de 2010; 49(8):1217-26. Publicación electrónica 29 de julio de 2010, que describen variaciones genéticas útiles para la determinación de la respuesta al tratamiento y supervivencia en cáncer. Por ejemplo, la detección de los niveles de ADNlc combinada con la detección de mutaciones en el gen de K-RAS y/o el gen de p53 proporcionan una herramienta poderosa, relativamente no invasiva en la medición del pronóstico de diversos cánceres, incluyendo cáncer de ovario, cáncer endometrial y linfomas. Dobrzycka B *et al.*, Ann Oncol. Mayo de 2011; 22(5):1133-40. Publicación electrónica 23 de noviembre de 2010; Dobrzycka B *et al.*, Int J Cancer. 1 de agosto de 2010; 127(3):612-21; Hosny G *et al.*, Cancer Lett. 18 de marzo de 2009; 275(2):234-9. Publicación electrónica 28 de noviembre de 2008. Tal análisis puede ayudarse adicionalmente usando herramientas tales como Varietas, un portal de de base de datos funcional para la identificación de variación genética y la asociación con desenlaces del tratamiento y pronóstico. Paananen J *et al.*, Database (Oxford). 29 de julio de 2010; 2010:baq016.

*Uso de sistemas de ensayo para la detección de muestras mixtas a partir de pacientes de trasplante*

- Los sistemas de ensayo de la invención pueden usarse para monitorizar la salud de los órganos en un paciente de trasplante usando una combinación de detección de ADNlc y detección de SNP o mutaciones en uno o más genes individuales. Los órganos trasplantados tienen genomas que son distintos del genoma de un paciente receptor, y la salud de los órganos puede detectarse usando un sistema de ensayo. Por ejemplo, se ha mostrado que el rechazo celular agudo está asociado con niveles significativamente aumentados de ADN libre circulante del genoma del donante en receptores de trasplante cardiaco. Snyder TM *et al.*, Proc Natl Acad Sci U S A. 12 de abril de 2011; 108(15):6229-34. Publicación electrónica 28 de marzo de 2011. Además, quimiocinas y moléculas de adhesión median en el rechazo de aloinjertos reclutando leucocitos en el aloinjerto, y se ha mostrado que los SNP ubicados en interleucina (IL)-8, CXCR1, CXCR2, se correlacionan con los desenlaces del aloinjerto. Ro H. *et al.*, Transplantation. 15 de junio de 2011; 91(1):57-64. Por tanto, los sistemas de ensayo de la invención pueden proporcionar pruebas no invasivas para monitorizar a los receptores de trasplantes de órganos sólidos, y pueden ayudar en la identificación de signos tempranos de rechazo sin la necesidad de biopsias de órganos u otras técnicas de pronóstico o diagnóstico más onerosas.

*Uso de sistemas de ensayo para la detección en muestras maternas*

- En determinados aspectos específicos, la determinación del porcentaje relativo de ADN fetal en una muestra materna puede ser beneficiosa en el análisis de los productos de amplificación, ya que el porcentaje de ADN fetal en la muestra proporciona información importante sobre la presencia estadística esperada de cromosomas, y la variación con respecto a esa expectativa puede ser indicativa de aneuploidía fetal. Esto puede ser especialmente útil en circunstancias en las que el nivel de ADN fetal en una muestra materna es bajo, ya que el porcentaje de contribución fetal puede usarse en la determinación de la significación estadística cuantitativa en las variaciones de los niveles de locus identificados seleccionados en una muestra materna. En otros aspectos, la determinación del porcentaje relativo de ADNlc fetal en una muestra materna puede ser beneficiosa en la estimación del nivel de certidumbre o potencia en la detección de una aneuploidía fetal.
- En algunos aspectos específicos, la contribución relativa fetal del ADN materno al alelo de interés puede compararse con la contribución paterna a ese alelo para determinar la concentración de ADN fetal aproximada en la muestra. En otros aspectos específicos, la cantidad relativa de secuencias derivadas únicamente por vía paterna (por ejemplo, secuencias de cromosoma Y o polimorfismos específicos de vía paterna) puede usarse para determinar la concentración relativa de ADN fetal en una muestra materna.

- Otro enfoque a modo de ejemplo para determinar el porcentaje de contribución fetal en una muestra materna es a través del análisis de fragmentos de ADN con diferentes patrones de metilación del ADN entre ADN fetal y materno. En un aspecto preferido, el ADN amplificado a partir de ADN libre circulante es mediante reacción en cadena de la polimerasa (PCR). Otros mecanismos para la amplificación pueden usarse también, incluyendo los descritos en más detalle en el presente documento, tal como resultará evidente para un experto en la técnica tras la lectura de la

presente divulgación.

5 En circunstancias en las que el feto es masculino, el porcentaje de ADN fetal en una muestra puede determinarse a través de la detección de locus específicos de Y y la comparación con el contenido en ADN materno calculado. Las cantidades de un locus específico de Y amplificado, tal como una región a partir del gen de la región Y determinante del sexo (SRY), que está ubicada en el cromosoma Y y es por tanto representativa del ADN fetal, puede determinarse a partir de la muestra y compararse con uno o más locus seleccionados amplificados que están presentes en tanto ADN materno como ADN fetal y que no son preferiblemente de un cromosoma que se cree que es posiblemente aneuploide en el feto, por ejemplo, una región autosómica que no está en el cromosoma 21 ó 18. 10 Preferiblemente, esta etapa de amplificación se realiza en paralelo con la esta de amplificación selectiva, aunque puede realizarse o bien antes o bien después de la amplificación selectiva según la naturaleza del ensayo multiplexado.

15 En aspectos particulares, el porcentaje de ADN fetal libre de células en una muestra materna puede determinarse mediante PCR usando ADN diluido en serie aislado de la muestra materna, que puede cuantificar de manera exacta el número de genomas que comprenden los genes amplificados. La PCR usando ADN diluido en serie aislado de la muestra materna puede preferirse cuando se determina el porcentaje de ADN fetal con un feto macho. Por ejemplo, si la muestra de sangre contiene el 100% de ADN fetal masculino, y se realizan diluciones en serie 1:2, entonces en promedio la señal ligada a Y desaparecerá 1 dilución antes de la señal autosómica, puesto que hay 1 copia del gen 20 ligado a Y y 2 del gen autosómico.

25 En un aspecto específico, el porcentaje de ADN fetal libre en plasma materno se calcula para un feto macho usando la siguiente fórmula: porcentaje de ADN fetal libre =  $(n.^{\circ} \text{ de copias del gen ligado a Y} \times 2 \times 100) / (n.^{\circ} \text{ de copias del gen autosómico})$ , en donde el número de copias de cada gen se determina observando la dilución en serie más alta en la que se detectó el gen. La fórmula contiene un factor de multiplicación de 2, que se usa para normalizar para el hecho de que hay sólo 1 copia del gen ligado a Y en comparación con dos copias del gen autosómico en cada genoma, fetal o materno.

30 *Determinación del contenido de ADN de fuente secundaria en una muestra mixta*

35 En determinados aspectos de la invención, la determinación de la contribución de ADN a partir de una fuente secundaria puede ser útil en la determinación de la variación en el número de copias de locus en esas muestras. Por ejemplo, en cada muestra derivada por vía materna, el ADN de un feto heredará aproximadamente el 50% de los locus genéticos de la madre y el 50% de los locus genéticos del padre. La determinación de los locus que contribuyen al feto a partir de fuentes no maternas permite la estimación de ADN fetal en una muestra materna, y por tanto proporciona información usada para calcular las diferencias estadísticamente significativas en las frecuencias cromosómicas para cromosomas de interés.

40 En determinados aspectos, la determinación de polimorfismos de fuente secundaria requiere análisis de mutación y/o SNP seleccionados como diana para identificar la presencia del ADN de fuente secundaria en una muestra mixta. La información necesaria para este cálculo puede proporcionarse usando el ensayo de la invención. En algunos aspectos, el uso de genotipado previo es útil, por ejemplo, genotipado del donante de un trasplante, genotipado del padre y la madre en una muestra materna. Pero generalmente esta información referente al genotipado previo no es necesaria para realizar el ensayo, y el genotipado se realiza simultáneamente con la 45 determinación del número de copias de locus seleccionados dentro de una muestra mixta.

50 En un aspecto preferido, el porcentaje de ácidos nucleicos de fuente secundaria en una muestra mixta puede cuantificarse usando detección de SNP multiplexada sin usar el conocimiento genotípico previo. En este aspecto, se usan dos o más locus polimórficos seleccionados con un SNP conocido en cada región. En un aspecto preferido, los locus polimórficos seleccionados son locus amplificados. En un aspecto preferido, la amplificación es universal. En una realización preferida, los locus polimórficos seleccionados se amplifican en una reacción en un recipiente. Cada alelo de los locus polimórficos seleccionados en la muestra materna se determina y se cuantifica. En un aspecto preferido, se usa secuenciación de alto rendimiento para tal determinación y cuantificación. Los locus se identifican cuando los genotipos de fuente principal y secundaria son diferentes, por ejemplo, el genotipo del donante es 55 homocigoto y el genotipo del receptor es heterocigoto. Esta identificación se realiza observando una frecuencia relativa alta de un alelo (>80%) y una frecuencia relativa baja (<20% y >0,15%) del otro alelo para un locus seleccionado particular. El uso de múltiples locus es particularmente ventajoso ya que reduce la cantidad de variación en la medición de la abundancia de los alelos. Todos o un subconjunto de los locus que cumplen este requisito se usan para determinar la concentración de ácido nucleico de fuente secundaria a través de análisis estadístico. En un aspecto, la concentración se determina sumando los alelos de frecuencia baja a partir de dos o 60 más locus juntos, dividiendo entre la suma de los alelos de frecuencia alta y multiplicando por dos. En otro aspecto, el porcentaje de ácido nucleico de fuente secundaria se determina calculando el promedio de los alelos de frecuencia baja a partir de dos o más locus, dividiendo entre el promedio de los alelos de frecuencia alta y multiplicando por dos.

65 Para muchos alelos, las secuencias de ácido nucleico de fuente principal y secundaria pueden ser homocigotas e

idénticas, y ya que esta información no se distingue, no es útil en la determinación de ácido nucleico de fuente secundaria en una muestra mixta. La presente invención usa información alélica en la que hay una diferencia distinguible entre las fuentes de células (por ejemplo, un alelo fetal que contiene al menos un alelo que difiere del alelo materno) en cálculos de los porcentajes de ácido nucleico de fuente secundaria. Datos referentes a regiones alélicas que son iguales para la fuente principal y secundaria no se seleccionan por tanto para el análisis, o se retiran de los datos pertinentes antes de la determinación de porcentajes saturar los datos útiles.

Pueden encontrarse métodos a modo de ejemplo para cuantificar ADN fetal en plasma materno, por ejemplo, en Chu *et al.*, Prenat Diagn 2010; 30:1226-1229.

En un aspecto, pueden excluirse locus seleccionados si la cantidad o frecuencia de la región parece ser un resultado atípico debido a error experimental, o a partir de sesgo genético idiopático dentro de una muestra particular. En otro aspecto, los ácidos nucleicos seleccionados pueden experimentar ajuste estadístico o matemático tal como normalización, estandarización, agrupamiento o transformación antes de la suma o el cálculo del promedio. En otro aspecto, los ácidos nucleicos seleccionados pueden someterse a tanto normalización como exclusión de errores experimentales de los datos antes de la suma o el cálculo del promedio.

En un aspecto preferido, se usan 12 o más locus para el análisis. En otro aspecto preferido, se usan 24 o más locus para el análisis. En otro aspecto preferido, se usan 48 o más locus para el análisis. En otro aspecto, se usan uno o más índices para identificar la muestra.

En un aspecto específico, la contribución de fuente secundaria puede cuantificarse usando detección de SNP en tándem. Se da a conocer técnicas para identificar SNP en tándem en ADN extraído de, por ejemplo, una muestra materna en Mitchell *et al*, patente estadounidense n.º 7.799.531 y las solicitudes de patente estadounidense n.ºs 12/581.070, 12/581.083, 12/689.924 y 12/850.588. Estas describen la diferenciación de locus fetales y maternos a través de la detección de al menos un polimorfismo de un solo nucleótido (SNP) en tándem en una muestra materna que tiene un haplotipo diferente entre el genoma fetal y materno. La identificación y cuantificación de estos haplotipos puede realizarse directamente sobre la muestra materna, tal como se describe en las divulgaciones de Mitchell *et al.*, y usarse para determinar el porcentaje de contribución fetal en la muestra materna.

Puesto que una muestra que no se ha sometido a prueba previamente para determinar el sexo del feto tiene una probabilidad del 50% aproximada de que el feto sea masculino, sólo podrán aplicarse secuencias específicas de Y en la mitad de tales muestras. En un aspecto específico, los locus polimórficos usados para la determinación de la contribución de fuente fetal no están en el cromosoma Y.

Una vez que se ha calculado el porcentaje de ADNlc para la fuente secundaria, estos datos pueden combinarse con métodos para la detección de aneuploidía para determinar la probabilidad de que una muestra mixta pueda contener una aneuploidía. En un aspecto, se usa un método de detección de aneuploidía que usa análisis de segmentos de ADN al azar, tal como el descrito en, por ejemplo, Quake, solicitud de patente estadounidense n.º 11/701.686; Shoemaker *et al.*, solicitud de patente estadounidense n.º 12/230.628. En un aspecto preferido, los métodos de detección de aneuploidía que utilizan el análisis de locus seleccionados en una muestra mixta incluyen tanto regiones para la determinación del contenido de ADN de fuente secundaria así como regiones no polimórficas a partir de dos o más cromosomas para detectar una anomalía cromosómica en una única reacción. La única reacción ayuda a minimizar el riesgo de contaminación o sesgo que pueden introducirse durante diversas etapas en el sistema de ensayo, lo que por lo demás desvía los resultados cuando se utiliza el contenido de ADN de fuente secundaria para ayudar a determinar la presencia o ausencia de una anomalía cromosómica. En otros aspectos, pueden utilizarse regiones o locus seleccionados tanto para la determinación del contenido de ADN de fuente secundaria así como la detección de anomalías cromosómicas de fuente secundaria. La utilización de las mismas regiones para tanto el contenido en ADN como la detección de anomalías cromosómicas puede ayudar adicionalmente a minimizar cualquier sesgo debido a error experimental o contaminación.

### Ejemplos

Los siguientes ejemplos se exponen para proporcionar a los expertos habituales en la técnica una divulgación y descripción completas de cómo producir y usar la presente invención, y no pretenden limitar el alcance de lo que los inventores consideran su invención, ni pretenden representar o implicar que los experimentos a continuación son todos o los únicos experimentos realizados. Los expertos en la técnica apreciarán que pueden hacerse numerosas variaciones y/o modificaciones en la invención tal como se muestra en los aspectos específicos sin apartarse del alcance de la invención tal como se describe y se define ampliamente en las reivindicaciones adjuntas. Los presentes aspectos, por tanto, han de considerarse en todos los sentidos ilustrativos y no restrictivos.

Se han hecho esfuerzos para garantizar la exactitud con respecto a los números usados (por ejemplo, cantidades, temperatura, etc.) pero deben considerarse algunos errores y desviaciones experimentales. A menos que se indique otra cosa, las partes son partes en peso, el peso molecular es el peso molecular promedio en peso, la temperatura es en grados centígrados y la presión está en o cerca de la atmosférica.

Ejemplo 1: Aspectos generales de los sistemas de ensayo de la invención

5 Se sometieron a prueba varios formatos de ensayo para demostrar la capacidad para realizar amplificación y detección selectivas de locus independientes para demostrar la detección multiplexada, basada en ligación de un gran número (por ejemplo, 96 o más) de locus de interés. Estos locus incluían locus que eran indicativos de la presencia de un cromosoma particular o la presencia o ausencia de una mutación o polimorfismo en un alelo particular.

10 Estos ensayos se diseñaron basándose en secuencias genómicas humanas, y cada interrogación consistía en dos oligonucleótidos de secuencia fija por locus seleccionado interrogado en el ensayo. El primer oligonucleótido, complementario a la región en 3' de una región genómica, comprendía los siguientes elementos de oligonucleótidos secuenciales (de 5' a 3'): una secuencia de cebado de PCR universal común para todos los ensayos: TACACCGGCGTTATGCGTCGAGAC (SEQ ID NO: 1); un índice de identificación de nueve nucleótidos específico para el locus seleccionado; una secuencia específica de locus o locus/alelo de 9 bases que actúa como índice de locus el primer conjunto independiente de SNP y un índice de locus/alelo en el segundo conjunto específico de polimorfismo; un nucleótido de rotura de hibridación que es diferente de la base correspondiente en el locus genómico; y una secuencia de 20-24 pb complementaria al locus genómico seleccionado. En casos en los que se detectó un SNP o una mutación en esta porción del locus genómico seleccionado, el conjunto de interrogación específico de alelo consistía en dos primeros cebadores de ligación en tándem de secuencia fija, cada uno con un índice de locus/alelo diferente y una base específica de alelo diferente en la posición de SNP. Estos primeros oligonucleótidos se diseñaron para cada ácido nucleico seleccionado para proporcionar una  $T_f$  uniforme predicha con una variación de dos grados a través de todas las interrogaciones en el conjunto de ensayo.

25 El segundo oligonucleótido de secuencia fija, complementario a la región en 5' de los locus genómicos, comprendía los siguientes elementos secuenciales (de 5' a 3'): una secuencia de 20-24 b complementaria a la región en 5' en el locus genómico; un nucleótido de rotura de hibridación diferente de la base correspondiente en el locus genómico; y una secuencia de cebado de PCR universal que era común para los tres oligonucleótidos en el conjunto de ensayo: ATTGCGGGACCGATGATCGCGTC (SEQ ID NO: 2).

30 En casos en los que se detectó un SNP o una mutación en el locus genómico seleccionado, el conjunto de interrogación específico de alelo consistía en dos cebadores de ligación en tándem, cada uno con un índice de locus/alelo diferente y una base específica de alelo diferente en la posición de mutación/SNP. Este segundo oligonucleótido de secuencia fija se diseñó para cada ácido nucleico seleccionado para proporcionar una  $T_f$  uniforme predicha con una variación de dos grados a través de todas las interrogaciones en el conjunto de ensayo que era sustancialmente el mismo intervalo de  $T_f$  que para el primer conjunto de oligonucleótidos.

35 En determinados aspectos sometidos a prueba, se usaron uno o más oligonucleótidos en puente que eran complementarios a la secuencia del locus genómico entre la región complementaria a los oligonucleótidos de secuencia fija primero y segundo usados para cada locus seleccionado. En aspectos específicos sometidos a prueba, se usó más de un más oligonucleótido en puente para abarcar el hueco entre los oligonucleótidos de secuencia fija, y el uno o más oligonucleótidos en puente pueden diseñarse opcionalmente para identificar una o más mutaciones o SNP en la secuencia. La longitud de los oligonucleótidos en puente usados en los sistemas de ensayo variaba desde 5 hasta 36 pares de bases.

45 Todos los oligonucleótidos usados en los formatos de ligación en tándem se sintetizaron usando química en fase sólida convencional. Los segundos oligonucleótidos de secuencia fija y los oligonucleótidos en puente se sintetizaron con restos fosfato en 5' para permitir la ligación a extremos terminales hidroxilo 3' de oligonucleótidos adyacentes.

Ejemplo 2: Preparación de ADN para su uso en procedimientos de ligación en tándem

50 Se obtuvo ADN genómico de un varón de raza blanca (NA12801) o una mujer de raza blanca (NA11995) de Coriell Cell Repositories (Camden, Nueva Jersey) y se fragmentó mediante fragmentación acústica (Covaris, Woburn, MA) hasta un tamaño de fragmento medio de aproximadamente 200 pb.

55 Se biotiniló el ADN de Coriell usando procedimientos convencionales. En resumen, se repararon los extremos del ADN fragmentado por Covaris generando la siguiente reacción en un microtubo de 1,5 ml: 5 µg de ADN, 12 µl de tampón de ligasa de T4 10X (Enzymatics, Beverly MA), 50 U de polinucleótido cinasa de T4 (Enzymatics, Beverly MA) y H<sub>2</sub>O hasta 120 µl. Se incubó esto a 37°C durante 30 minutos. Se diluyó el ADN usando Tris 10 mM, EDTA 1 mM pH 8,5 hasta una concentración final deseada de ~2 ng/µl.

60 Se colocaron 5 µl de ADN en cada pocillo de una placa de 96 pocillos, y se selló la placa con un sellador de placas adhesivo y se centrifugó durante 10 segundos a 250 x g. Entonces se incubó la placa a 95°C durante 3 minutos, se enfrió hasta 25°C y se centrifugó de nuevo durante 10 segundos a 250 x g. Se preparó una mezcla maestra de biotinilación en un microtubo de 1,5 ml hasta una concentración final de: tampón TdT IX (Enzymatics, Beverly, MA), 8 U de TdT (Enzymatics, Beverly, MA), CoCl<sub>2</sub> 250 µM, biotina-16-dUTP 0,01 nmol/µl (Roche, Nutley NJ) y H<sub>2</sub>O hasta

1,5 ml. Se tomaron alícuotas de 15  $\mu$ l de la mezcla maestra en cada pocillo de una placa de 96 pocillos, y se selló la placa con sellador de placas adhesivo. Se centrifugó la placa durante 10 segundos a 250 x g y se incubó durante 37°C durante 60 minutos. Tras la incubación, se centrifugó la placa de nuevo durante 10 segundos a 250 x g, y se añadieron a cada pocillo 7,5  $\mu$ l de mezcla de precipitación (azul dextrano 1  $\mu$ g/ $\mu$ l, NaOAC 3 mM).

5 Se selló la placa con un sellador de placas adhesivo y se mezcló usando un agitador de vórtex de placas IKA durante 2 minutos a 3000 rpm. Se añadieron 27,5  $\mu$ l de isopropanol en capa pocillo, se selló la placa con sellador de placas adhesivo y se agitó con vórtex durante 5 minutos a 3000 rpm. Se centrifugó la placa durante 20 minutos a 3000 x g, se decantó el sobrenadante y se invirtió la placa y se centrifugó a 10 x g durante 1 minuto sobre una  
10 toallita absorbente. Se secó al aire la placa durante 5 minutos y se resuspendió el sedimento en 30  $\mu$ l de Tris 10 mM pH 8,0, EDTA 1 mM.

### Ejemplo 3: Formatos de ensayo a modo de ejemplo usando ligación en tándem

15 Se sometieron a prueba numerosos formatos de ensayo de ligación en tándem usando el ADN biotinilado para ilustrar la prueba de concepto para los sistemas de ensayo de la invención, y se demostró la capacidad para realizar una detección dirigida, altamente multiplexada de un gran número de locus independientes usando la series de diferentes formatos de ensayo. Los sistemas de ensayo a modo de ejemplo de la invención se diseñaron para que comprendieran 96 o más interrogaciones por locus en una muestra genética, y en casos en los que se detectaron  
20 SNP los formatos de ensayo utilizaron 192 o más interrogaciones diferenciadas, utilizando cada una la detección de diferentes alelos por 96 locus en muestras genéticas. Los ejemplos descritos para cada formato de ensayo utilizaron dos conjuntos diferentes de oligonucleótidos de secuencia fija y/o oligonucleótidos en puente (tal como se describe en el ejemplo 1), que comprendían un total 96 ó 192 reacciones de interrogación para los locus seleccionados según si se identificaron o no SNP.

25 Un primer formato de ensayo a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de locus y oligonucleótidos en puente, en donde había un hueco de una base entre el primer oligonucleótido de secuencia fija y los oligonucleótidos en puente, y un segundo hueco de una base entre los oligonucleótidos en puente y el segundo oligonucleótido de secuencia fija. Cada uno de los dos huecos abarcaba dos SNP diferentes. En este formato, se  
30 usó una ADN polimerasa para incorporar cada una de las bases de SNP, y se usó una ligasa para sellar las mellas formadas de ese modo. La discriminación de bases de SNP se derivaba de la fidelidad de incorporación de bases por la polimerasa, y en el caso de incorporación incorrecta, la tendencia de la ligasa a no sellar las mellas adyacentes a las bases desapareadas.

35 El segundo formato de ensayo a modo de ejemplo usaba dos oligonucleótidos de secuencia fija específicos de locus sin un oligonucleótido en puente, en donde había un hueco de -15-35 bases entre los oligonucleótidos de secuencia fija, y en donde el hueco abarcaba uno o más SNP. En este formato, se usó una polimerasa para incorporar las bases que faltaban del hueco, y se usó una ligasa para sellar la mella formada de ese modo. La discriminación de bases de SNP se derivaba de la fidelidad de incorporación de bases por la polimerasa, y en el caso de incorporación  
40 incorrecta, la tendencia de la ligasa a no sellar mellas adyacentes a las bases desapareadas.

Un tercer formato de ensayo a modo de ejemplo usaba oligonucleótidos de secuencia fija primero y segundo específicos de alelo sin un oligonucleótido en puente, en donde había un hueco de -15-35 bases gap entre los oligonucleótidos de secuencia fija primero y segundo, y en donde el hueco abarcaba uno o más SNP. Se usaron dos  
45 primeros oligonucleótidos de secuencia fija específicos de alelo diferenciados y dos segundos oligonucleótidos de secuencia fija específicos de alelo diferenciados. Se usó una polimerasa para incorporar las bases que faltaban, y se usó una ligasa para sellar la mella formada de ese modo. La discriminación de bases de SNP se derivaba de la especificidad de hibridación, la tendencia de la polimerasa sin corrección de errores a no extender cebadores apareados con apareamientos erróneos cerca del extremo 3', y la tendencia de la ligasa a no sellar mellas adyacentes a bases desapareadas.

Un cuarto formato a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de alelo y un oligonucleótido en puente específico de locus. En este formato, dos oligonucleótidos de secuencia fija diferenciados complementarios al extremo 3' de los locus de interés, el primer con una base en 3' específica para un alelo del SNP  
55 seleccionado como diana, y the segundo con una base en 3' específica para el otro alelo del SNP seleccionado como diana. De manera similar, se usaron dos segundos oligonucleótidos de secuencia fija diferenciados, el primero con una base en 5' específica para un alelo de un segundo SNP seleccionado como diana, y el segundo con una base en 5' específica para el otro alelo del segundo SNP seleccionado como diana. Los oligonucleótidos en puente eran complementarios a la región directamente adyacente a las regiones de locus complementarias a los  
60 oligonucleótidos de secuencia fija primero y segundo, y por tanto no se necesitaba polimerasa antes de la ligación. Se usó una ligasa para sellar las mellas entre los oligonucleótidos de secuencia fija y el oligonucleótido en puente. La discriminación de bases de SNP en este formato de ensayo se derivaba de la especificidad de hibridación y la tendencia de la ligasa a no sellar mellas adyacentes a bases desapareadas. Este formato a modo de ejemplo se sometió a prueba usando o bien ligasa de T4 o bien Taq ligasa para la creación del molde contiguo, y ambas  
65 demostraron ser eficaces en la reacción tal como se describe más adelante.

Un quinto formato a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de locus que eran complementarios a regiones adyacentes en el ácido nucleico de interés, y por tanto no se creó un hueco mediante la hibridación de estos oligonucleótidos. En este formato, no se requería polimerasa, y se usó una ligasa para sellar la mella individual entre los oligonucleótidos.

5 Un sexto formato a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de alelo y oligonucleótidos en puente específicos de locus, en donde había un hueco de bases corto de cinco bases entre la región de locus complementaria a los oligonucleótidos de secuencia fija. El oligonucleótido en puente específico de locus en este ejemplo tenía 5 meros complementario a las regiones directamente adyacentes a las regiones complementarias a los oligonucleótidos de secuencia fija primero y segundo. En este formato, no se requería polimerasa, y se usó una ligasa para sellar las dos mellas entre los oligonucleótidos.

15 Un séptimo formato a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de locus y un oligonucleótido en puente específico de locus, en donde había un hueco de bases más corto de cinco bases que contenía un SNP en la región complementaria al oligonucleótido en puente. Se incluyeron oligonucleótidos en puente específicos de alelo correspondientes a los posibles SNP en la reacción de hibridación y ligación. En este formato, no se requería polimerasa, y se usó una ligasa para sellar las dos mellas entre los oligonucleótidos. La discriminación de bases de SNP en este formato de ensayo se derivaba de la especificidad de hibridación y la tendencia de la ligasa a no sellar mellas adyacentes a bases desapareadas.

20 Un octavo formato a modo de ejemplo usaba oligonucleótidos de secuencia fija específicos de locus y dos oligonucleótidos en puente específicos de locus adyacentes, en donde había un hueco de 10 bases entre las regiones complementarias a los oligonucleótidos de secuencia fija primero y segundo. Se incluyeron oligonucleótidos en puente específicos de locus en la reacción de ligación, requiriendo el hueco dos 5 meros contiguos para formar el puente en el hueco. En este formato, no se requería polimerasa, y se usó una ligasa para sellar las tres mellas entre los oligonucleótidos.

30 Para cada uno de los formatos de ensayo descritos anteriormente, se creó una agrupación equimolar (40 nM cada una) de conjuntos de oligonucleótidos de secuencia fija específicos de locus o alelo primero y segundo a partir de los oligonucleótidos preparados tal como se expuso en el ejemplo 2. Se creó asimismo una agrupación equimolar diferenciada (20  $\mu$ M cada una) de oligonucleótidos en puente para los procedimientos de ensayo basados en las secuencias de los locus genómicos seleccionados.

35 Se transfirieron 100  $\mu$ g de perlas de estreptavidina a los pocillos de una placa de 96 pocillos, y se retiró el sobrenadante. Se añadieron a las perlas 60  $\mu$ l de tampón BB2 (Tris 100 mM pH 8,0, EDTA 10 mM, NaCl<sub>2</sub> 500 mM, formamida al 58%, Tween-80 al 0,17%), 10  $\mu$ l de agrupación de oligonucleótidos de secuencia fija 40 nM y 30  $\mu$ l del ADN de molde biotinilado preparado en el ejemplo 2. Se selló la placa con un sellador de placas adhesivo y se agitó con vórtex a 3000 rpm hasta que las perlas se resuspendieron. Se aparearon los oligonucleótidos con el ADN de molde mediante incubación a 70°C durante 5 minutos, seguido por enfriamiento lento hasta temperatura ambiente.

40 Se colocó la placa en una placa magnética de barra elevada durante 2 minutos para arrastrar las perlas magnéticas y el ADN asociado al lateral de los pocillos. Se retiró el sobrenadante pipeteando, y se reemplazó por 50  $\mu$ l de BB2 al 60% (v/v en agua). Se resuspendieron las perlas agitando con vórtex, se colocaron en el imán de nuevo y se retiró el sobrenadante. Se repitió este procedimiento de lavado de perlas una vez usando 50  $\mu$ l de BB2 al 60%, y se repitió dos veces más usando 50  $\mu$ l de tampón de lavado (Tris 10 mM pH 8,0, EDTA 1 mM, NaCl<sub>2</sub> 50 mM).

45 Se resuspendieron las perlas en 37  $\mu$ l de mezcla de reacción de ligación que consistía en tampón de Taq ligasa IX (Enzymatics, Beverly, MA), 1 U de Taq ligasa y agrupación de oligonucleótidos en puente 2  $\mu$ M (según el formato de ensayo), y se incubaron a 37°C durante una hora. Cuando era apropiado, y según el formato de ensayo, se incluyó en esta mezcla polimerasa termoestable sin corrección de errores más 200 nM cada dNTP. Se colocó la placa sobre una placa magnética de barra elevada durante 2 minutos para arrastrar las perlas magnéticas y el ADN asociado al lateral de los pocillos. Se retiró el sobrenadante pipeteando, y se reemplazó por 50  $\mu$ l de tampón de lavado. Se resuspendieron las perlas agitando con vórtex, se colocaron sobre el imán de nuevo y se retiró el sobrenadante. Se repitió el procedimiento de lavado una vez.

55 Para eluir los productos de las perlas de estreptavidina, se añadieron 30  $\mu$ l de Tris 10 mM, EDTA 1 mM, pH 8,0 a cada pocillo de una placa de 96 pocillos. Se selló la placa y se mezcló de nuevo usando un agitador de vórtex de IKA durante 2 minutos a 3000 rpm para resuspender las perlas. Se incubó la placa a 95°C durante 1 minuto, y se aspiró el sobrenadante usando un pipeteador de 8 canales. Se transfirieron 25  $\mu$ l de sobrenadante de cada pocillo a una placa de 96 pocillos nueva para la amplificación universal.

#### Ejemplo 4: Amplificación universal de productos ligados en tándem

65 Se amplificaron ácidos nucleicos polimerizados y/o ligados usando cebadores de PCR universales complementarios a las secuencias universales presentes en los oligonucleótidos de secuencia fija primero y segundo hibridados con

los locus de interés. Se usaron 25 µl de cada una de las mezclas de reacción del ejemplo 3 en cada reacción de amplificación. Una reacción de PCR universal de 50 µl consistía en 25 µl de productos de ligación eluidos más tampón Pfusion IX (Finnzymes, Finlandia), betaína 1 M, 400 nM cada dNTP, 1 U de ADN polimerasa termoestable con corrección de errores Pfusion y los siguientes pares de cebadores:

5 TAATGATACGGCGACCACCGAGATCTACACCGGCGTTATGCGTCGAGA (SEQ ID NO: 3) y

10 TCAAGCAGAAGACGGCATACGAGATXAAACGACGCGATCATCGGTCC CCGCAA (SEQ ID NO: 4), en donde X representa uno de 96 índices de muestras diferentes usados para identificar de manera única muestras individuales antes del agrupamiento y la secuenciación. Se llevó a cabo la PCR en condiciones rigurosas usando un termociclador Tetrad™ de BioRad.

15 Se agruparon 10 µl de producto de PCR universal de cada una de las muestras y se purificó el producto de PCR agrupado usando perlas AMPureXP™ SPRI (Beckman-Coulter, Danvers, MA), y se cuantificaron usando el instrumento Quant-iT™ PicoGreen (Invitrogen, Carlsbad, CA).

Ejemplo 5: Detección y análisis de locus seleccionados

20 Se secuenciaron los productos de PCR purificados de cada formato de ensayo en un único carril de un portaobjetos en un instrumento Illumina HiSeq™ 2000 (Illumina, San Diego, CA). Las rondas de secuenciación normalmente dan lugar a ~100M lecturas sin procesar, de las cuales ~85M (85%) se mapean en estructuras de ensayo esperadas. Esto se traduce en un promedio de ~885K lecturas/muestra a través del experimento, y (en el caso de un experimento que usa 96 locus) 9,2K lecturas/réplica/locus a través de 96 locus. Se analizaron las lecturas mapeadas en recuentos de réplicas/locus/alelos, y se calcularon diversas métricas para cada condición, incluyendo:

25 Rendimiento: una métrica de la proporción de ADN de entrada que se consultó en la secuenciación, calculado como el número promedio de lecturas únicas por locus (contando sólo lecturas de índice identificación únicas por réplica/locus) dividido entre el número total de equivalentes genómicos contenidos en el ADN de entrada.

30 Intervalo de frecuencia de locus de percentil 80: una métrica de la variabilidad de frecuencia de locus en los datos de secuenciación, interpretados como el intervalo en veces que abarca el 80% de los locus. Se calculó sobre la distribución de lecturas totales por locus, a través de todos los locus, como el percentil 90 de lecturas totales por locus dividido entre el percentil 10 de las lecturas totales por locus.

35 Tasa de error de SNP: una métrica de la tasa de error de la posición de SNP, y calculada como la proporción de lecturas que contienen una base discordante en la posición de SNP.

Estos resultados se resumen en la tabla 1:

40 Tabla 1: Resumen de resultados de formatos de ensayo de ligación en tándem

FORMATO DE ENSAYO	OLIGONUCLEÓTIDO DE SECUENCIA FIJA (1° y/o 2°)	OLIGONUCLEÓTIDO EN PUENTE USADO	ENZIMA USADA	RENDIMIENTO	INTERVALO DE FREC. DE LOC. DEL 80%	TASA DE ERROR DE SNP
1	ESPECÍFICO DE LOCUS	Específico de locus	pol+lig	9,5%	5,3	0,18%
2	ESPECÍFICO DE LOCUS	No	pol+lig	1,4%	58,3	0,19%
3	ESPECÍFICO DE ALELO	No	pol+lig	0,4%	61,7	1,00%
4	ESPECÍFICO DE ALELO	Específico de locus	Taq lig	5,0%	5,9	0,92%
4	ESPECÍFICO DE ALELO	Específico de locus	T4 lig	5,3%	4,4	0,95%
5	ESPECÍFICO DE LOCUS	No	Taq lig	22,5%	1,7	N/A
6	ESPECÍFICO DE LOCUS	Específico de locus	Taq lig	12,5	2,9	N/A

7	ESPECÍFICO DE LOCUS	Específico de locus	Taq lig	14,3	2,8	0,20%
8	ESPECÍFICO DE LOCUS	2 específicos de locus	Taq lig	18,5%	2,8	N/A

La tabla 1 indica que el ensayo de ligación en tándem específico de locus usando un oligonucleótido en puente convirtió ADN de molde en producto seleccionado como diana con alto rendimiento (~10%), con una alta proporción de producto derivado de locus seleccionados como diana (el 15% de las lecturas no contenían estructuras de ensayo esperadas), con sesgo de locus limitado (el 80% de los locus se encuentran dentro de un intervalo de concentración de -5 veces), y con alta exactitud de SNP (tasa de error de SNP del 0,2%). El ensayo de ligación en tándem específico de locus sin el uso de un oligonucleótido en puente produjo rendimientos reducidos y sesgo de locus sustancial, pero todavía producía datos de genotipado de SNP de alta exactitud. El ensayo de ligación en tándem específico de alelo con un oligonucleótido en puente produjo rendimientos intermedios en comparación con el ensayo específico de locus usando tanto Taq ligasa como ligasa de T4, pero todavía producía un sesgo de locus limitado y datos de genotipado de SNP de alta exactitud. El ensayo de ligación en tándem específico de alelo sin un puente produjo rendimientos reducidos y sesgo de locus sustancial, pero todavía producía datos de genotipado de SNP de alta exactitud.

Los formatos de ensayo seis a ocho mostraron que puede convertirse ADN de molde en producto seleccionado como diana con alto rendimiento (12-18%), con una alta proporción de producto derivado de locus seleccionados como diana (~76% de las lecturas contenían estructuras de ensayo esperadas), y con sesgo de locus limitado (el 80% de los locus se encuentran dentro de un intervalo de concentración de 2-3 veces). La figura 5 ilustra el rendimiento de genotipado que se obtuvo usando el formato de ensayo siete, que compara los recuentos de secuencia para los dos alelos de todos los ensayos polimórficos observados en una única muestra. Obsérvese la separación clara de los agrupamientos homocigótico y heterocigótico, así como los recuentos de fondo bajos observados entre los agrupamientos homocigotos.

Ejemplo 6: Determinación del porcentaje de ADN fetal usando ligación en tándem

Se diseñó un sistema de ensayo a modo de ejemplo de la invención para determinar el porcentaje de concentración de ADN fetal en una muestra genética así como para proporcionar recuentos para locus seleccionados dentro de la muestra. Este ensayo a modo de ejemplo comprendía 480 interrogaciones diferenciadas, utilizando cada una la detección de diferentes locus en una muestra materna. El ejemplo inicial utilizó una determinación del porcentaje de ADN fetal en sujetos que portan un feto macho, y de ese modo se utilizaron locus en el cromosoma Y así como locus que contenían un SNP fetal heredado por vía paterna que es diferente de la secuencia materna.

Específicamente, se interrogaron 480 ácidos nucleicos seleccionados usando el sistema de ensayo. Los 480 ácidos nucleicos seleccionados comprendían 48 interrogaciones específicas de secuencia de ácidos nucleicos correspondientes a locus en el cromosoma Y, 192 interrogaciones específicas de secuencia de ácidos nucleicos correspondientes a locus en el cromosoma 21, 192 interrogaciones específicas de secuencia de ácidos nucleicos seleccionados correspondientes a locus en el cromosoma 18 y 144 interrogaciones específicas de secuencia de ácidos nucleicos seleccionados correspondientes a locus polimórficos en los cromosomas 1-16. Se diseñaron estos ensayos basándose en secuencias genómicas humanas, y cada interrogación usaba tres oligonucleótidos por ácido nucleico seleccionado interrogado en el ensayo.

El primer oligonucleótido usado para cada interrogación era complementaria a la región en 3' de la región genómica seleccionada, y comprendía los siguientes elementos de oligonucleótido secuenciales (de 5' a 3'): una secuencia de cebado de PCR universal común para todos los: TACACGGCGTTATGCGTCGAGAC (SEQ ID NO: 1); un índice de identificación específico para los locus seleccionados que comprende nueve nucleótidos; y una secuencia de 20-24 pb complementaria al locus genómico seleccionado. El primer oligonucleótido se diseñó para cada ácido nucleico seleccionado para proporcionar una T<sub>r</sub> uniforme predicha con una variación de dos grados a través de todas las interrogaciones en el conjunto de 480 ensayos.

El segundo oligonucleótido usado para cada interrogación era un oligonucleótido en puente complementario a la secuencia de locus genómico directamente adyacente a la región genómica complementaria al primer oligonucleótido. Basándose en los ácidos nucleicos seleccionados de interés, los oligonucleótidos en puente se diseñaron para permitir la utilización de un total de 12 secuencias de oligonucleótido que podrían servir como oligonucleótidos en puente para todas las 480 interrogaciones en el conjunto de ensayo.

El tercer oligonucleótido usado para cada interrogación era complementario a la región en 5' del locus genómico seleccionado, comprendía los siguientes elementos secuenciales (de 5' a 3') elementos: una secuencia de 20-24 b complementaria a la región en 5' en el locus genómico; un nucleótido de rotura de hibridación que era diferente de la base correspondiente en el locus genómico; y una secuencia de cebado de PCR universal que era común a todos los terceros oligonucleótidos en el conjunto de ensayo: ATTGCGGGACCGATGATCGCGTC (SEQ ID NO: 2). Este

tercer oligonucleótido si diseñó para cada ácido nucleico seleccionado para proporcionar una  $T_f$  uniforme predicha con una variación de dos grados a través de todas las interrogaciones en el conjunto de 480 ensayos, y el intervalo de  $T_f$  era sustancialmente el mismo que el intervalo de  $T_f$  que el primer conjunto de oligonucleótidos.

5 Todos los oligonucleótidos se sintetizaron usando química en fase sólida convencional. Los primeros oligonucleótidos y los oligonucleótidos en puente se sintetizaron con restos fosfato en 5' para permitir la ligación a extremos terminales hidroxilo en 3' de oligonucleótidos adyacentes. Se creó una agrupación equimolar de conjuntos de los oligonucleótidos primero y tercero para todas las interrogaciones en el ensayo multiplexado, y se creó una agrupación equimolar diferenciada de todos los oligonucleótidos en puente para permitir reacciones de hibridación diferenciadas.

15 Se aisló ADN a partir de 5 ml de plasma usando el kit Dynal Silane viral NA (Invitrogen, Carlsbad, CA). Se procesaron aproximadamente 12 ng de ADN de cada una de 37 mujeres, incluyendo 7 sujetos femeninos no embarazadas, 10 sujetos femeninos embarazadas de varones y 22 sujetos femeninos embarazadas de mujeres. Se biotiniló el ADN usando procedimientos convencionales, y se inmovilizó el ADN biotinilado sobre una superficie sólida recubierta con estreptavidina para permitir la retención del ADN genómico en etapas de ensayo posteriores.

20 Se hibridó el ADN inmovilizado con la primera agrupación que comprendía los oligonucleótidos primero y tercero para cada secuencia interrogada en condiciones de hibridación rigurosas. Los oligonucleótidos no hibridados en la agrupación se lavaron entonces de la superficie del soporte sólido, y se hibridó el ADN inmovilizado con la agrupación que comprendía los oligonucleótidos en puente en condiciones de hibridación rigurosas. Una vez que se permitió que los oligonucleótidos en puente se hibridaran con el ADN inmovilizado, se lavaron los oligonucleótidos no unidos restantes de la superficie y se ligaron los tres oligonucleótidos hibridados unidos a los locus seleccionados usando ligasa de T4 para proporcionar un molde de ADN contiguo para la amplificación.

25 Se amplificó el ADN ligado del sustrato sólido usando una ADN polimerasa termoestable con corrección de errores, un primer cebador de PCR universal TAATGATACGGCGACCACCGAGATCTACACCGCGTTATGCGTCGA GA (SEQ ID NO: 3) y un segundo cebador de PCR universal TCAAGCAGAAGACGGCATAACGATXAAACGACGCGATCATCGGTC CCCGCAA (SEQ ID NO: 4), en donde X representa uno de 96 índices de muestras diferentes usados para identificar de manera única muestras individuales antes del agrupamiento y la secuenciación. Se agruparon 10  $\mu$ l de producto de PCR universal de cada una de las 37 muestras descritas anteriormente y se purificó el producto de PCR agrupado usando perlas AMPure™ SPRI (Beckman-Coulter, Danvers, MA), y se cuantificó usando el instrumento Quant-iT™ PicoGreen (Invitrogen, Carlsbad, CA).

35 Se secuenció el producto de PCR purificado sobre 6 carriles de un portaobjetos individual sobre un instrumento Illumina HiSeq™ 2000. La ronda de secuenciación dio lugar a 384M lecturas sin procesar, de las cuales 343M (89%) se mapearon en locus genómicos esperados, dando como resultado un promedio de 3,8M lecturas por muestra a través de las 37 muestras, y 8K lecturas por muestra por locus a través de los 480 locus. Se analizaron las lecturas mapeadas en recuentos de muestras y locus, y se calcularon dos métricas diferenciadas de porcentaje de ADN fetal tal como sigue.

45 El porcentaje ADN de varón detectado por locus de cromosoma Y corresponde a la proporción relativa de lecturas derivadas de interrogaciones de locus de cromosoma Y frente a la proporción relativa de lecturas derivadas de interrogaciones de locus autosómicos y se calculó como (número de lecturas de cromosoma Y en un sujeto de prueba/número de lecturas autosómicas en el sujeto de prueba)/(número de lecturas en sujeto de control masculino/número de lecturas autosómicas en el sujeto de control masculino). Se usó esta métrica como medida del porcentaje de ADN fetal en el caso de un feto macho usando las lecturas relativas de cromosoma Y.

50 El porcentaje de ADN fetal detectado por locus polimórficos corresponde a la proporción de lecturas derivadas de alelos no maternos frente a maternos en locus en donde puede hacerse una distinción de este tipo. En primer lugar, para cada locus identificado, se dividió el número de lecturas para el alelo con los recuentos más bajos (el alelo de frecuencia baja) entre el número total de lecturas para proporcionar una frecuencia de alelo minoritario (MAF) para cada locus. Entonces, se identificaron locus con una MAF de entre el 0,075% y el 15% como locus informativos. Se calculó el porcentaje estimado de ADN fetal para la muestra como la media de la frecuencia de alelo minoritario de los locus informativos multiplicada por dos, es decir calculada como 2X la aparición promedio (MAF) en donde  $0,075\% < \text{MAF} < 15\%$ .

60 La figura 6 demuestra los resultados de estos dos cálculos. Tal como se muestra en la figura 6, el porcentaje de locus macho determinado usando la métrica de cromosoma Y descrita anteriormente (círculos grises) puede separar embarazos que implican fetos macho de embarazos que implican fetos hembra (rombos grises) y muestras de no embarazadas (círculos negros). Además, el cálculo del porcentaje cantidad fetal en una muestra por métrica de locus polimórficos puede distinguir muestras de embarazadas de muestras de no embarazadas. Finalmente, había una correlación entre las estimaciones de porcentaje ADN fetal para una muestra obtenida del cromosoma Y y locus polimórficos en embarazos que implican fetos macho. Esta correlación persiste hasta valores de porcentaje fetal bastante bajos.

Ejemplo 7: Detección de aneuploidía en una muestra materna

- 5 Se usaron los sistemas de ensayo de la invención en la detección de polimorfismos y anomalías cromosómicas en dos cohortes diferenciadas de mujeres embarazadas. Se sometieron a prueba una primera cohorte de 190 embarazos normales, 36 T21 y 8 T18 y una segunda cohorte de 126 embarazos normales, 36 T21 y 8 T18 para detectar aneuploidía fetal. Se detectaron las aneuploidías cromosómicas usando 576 ensayos de cromosoma 21 y 576 ensayos de cromosoma 18, se agruparon juntos y se sometieron a ensayo en una única reacción, tal como se expone a continuación.
- 10 Los elementos usados en los ensayos de detección de aneuploidía se ilustran en la figura 7. Se usó el ADNlc 701 aislado de muestras maternas se usó como plantilla para la hibridación, ligación y amplificación de múltiples locus seleccionados de tanto el cromosoma 21 como el cromosoma 18 en cada muestra materna. Se hibridaron tres oligonucleótidos con cada locus seleccionado para crear productos de ligación para la amplificación y detección. El oligonucleótido de secuencia fija izquierdo (o primero) comprendía una región complementaria a un locus 709 seleccionado y una primera región 711 de cebador universal. El oligonucleótido 705 de secuencia fija derecho (o segundo) comprendía una segunda región complementaria al locus 713 seleccionado y una segunda región 715 de cebador universal. Los oligonucleótidos 707 en puente usados se diseñaron para que se hibridaran cada uno con regiones en puente de dos o más locus seleccionados usados en el ensayo de detección de aneuploidía. Cuando los oligonucleótidos 703, 705 de secuencia fija y the oligonucleótido 707 en puente se hibridaron con la región complementaria en el ADNlc 701, sus extremos terminales formaron dos mellas. Tras la ligación de los oligonucleótidos hibridados con el ADNlc, se creó un producto de ligación para cada locus seleccionado que comprendía 703, 705 y 707 que se usó como molde para los cebadores de amplificación 719, 721.
- 15 20
- 25 Se usaron entonces dos cebadores 719, 721 de amplificación que comprendían regiones complementarias a las regiones de cebador universales primera y segunda, respectivamente, para amplificar el producto de ligación. Este producto de amplificación comprendía la secuencia del locus seleccionado. El cebador de amplificación derecho también comprendía un índice 717 de muestra para identificar la muestra particular a partir de la que se obtuvo el locus en el ensayo multiplexado. La amplificación con 96 cebadores 729 de amplificación derechos distintos permitió el agrupamiento y la secuenciación simultánea de 96 productos de amplificación diferentes en un único carril.
- 30 Los cebadores 719, 721 de amplificación también contenían una secuencia 723 de agrupamiento izquierda (TAATGATACGGCGACCACCGA) (SEQ ID NO: 7) y una secuencia 725 de agrupamiento derecha (ATCTCGTATGCCGTCTTCTGCTTGA) (SEQ ID NO: 8) que soportaban la amplificación de agrupamientos para la secuenciación usando el sistema Illumina HiSeq™ 2000 (Illumina, San Diego, CA). Se usó un cebador 727 de secuenciación que comprendía la primera secuencia de cebador universal para determinar la secuencia del producto de amplificación, y se usó un segundo cebador 729 de secuenciación para determinar el índice 717 de muestra del producto de amplificación.
- 35 40
- 45 En resumen, se recogieron aproximadamente 10 ml de sangre periférica de cada paciente en un tubo de BCT (Streck, Omaha, NE), que se envió por medio de correo nocturno a Tandem Diagnostics. Se aisló plasma de los tubos BCT en el plazo de 72 h de la recogida de sangre mediante centrifugación a 1600 g durante 10 m. Se transfirió el plasma a un segundo tubo y se centrifugó a 16000 g durante 10 m para eliminar cualquier célula restante. Se aisló ADNlc de 4-5 ml de plasma por paciente. Se aislaron aproximadamente 15 ng de ADNlc de cada muestra de paciente y se dispusieron en pocillos individuales de una placa de 96 pocillos. Todo el procesamiento posterior se produjo sobre lotes multiplexados de hasta 96 muestras de pacientes de ADNlc por método de sistema de matriz.
- 50 55
- Se biotinó el ADNlc aislado de las muestras maternas en cada pocillo, se precipitó y se resuspendió en 30 µl de TE como en el ejemplo 3 anteriormente. Se mezcló el ADN de molde biotinilado con 100 µg de perlas magnéticas recubiertas con estreptavidina MyOneC1 (Life Technologies, Carlsbad, CA), 60 µl de tampón BB2 (Tris 100 mM pH 8,0, EDTA 10 mM, NaCl<sub>2</sub> 500 mM, formamida al 58%, Tween-80 al 0,17%), y 10 µl de oligonucleótidos de secuencia fija izquierdo 703 y derecho 705 agrupados 40 nM. Se calentó la mezcla hasta 70°C, y se enfrió 2 horas. Entonces se inmovilizaron magnéticamente las perlas en el lateral del pocillo, se lavaron dos veces con 50 µl de BB2 al 60% (v/v con H<sub>2</sub>O), se lavaron dos veces más con 50 µl de tampón de lavado (Tris 10 mM pH 8,0, EDTA 1 mM, NaCl<sub>2</sub> 50 mM), y luego se resuspendieron en una reacción de 50 µl que contenía 1 U de Taq ligasa (Enzymatics, Beverly MA), tampón de Taq ligasa IX (Enzymatics) y 10 µM de un oligonucleótido 707 en puente de 5' meros fosforilado en 5'. Se incubó la mezcla a 37°C durante 1 hora. Se inmovilizaron magnéticamente de nuevo las perlas en el lateral del pocillo, se lavaron dos veces con 50 µl de tampón de lavado y luego se resuspendieron en 30 µl de TE.
- 60 65
- Se eluyeron los productos de ligación de las perlas inmovilizadas mediante incubación a 95°C durante 3 minutos. Se amplificaron los productos de ligación eluidos mediante 26 ciclos de PCR en una reacción de 50 µl que contenía 1 U de polimerasa Pfusion (Thermo Fisher, Waltham MA), betaína 1 M, tapón Pfusion IX y cebadores de amplificación izquierdo y derecho 400 nM (719, 721 respectivamente). El cebador derecho contenía un índice de muestra de 7 bases (717) que permitía la secuenciación multiplexada de 96 muestras en el sistema HiSeq2000 (Illumina, San Diego, CA). La secuencia del oligonucleótido de secuencia fija izquierdo era:

TAATGATACGGCGACCACCGAGATCTACACCGCGTTATGCGTCGAGA

C

(SEQ ID NO: 5)

Y la secuencia del oligonucleótido de secuencia derecho era:

5 TCAAGCAGAAGACGGCATAACGAGATNNNNNNNAAACGACGCGATCATCG

GTCCCCGCAAT (SEQ ID NO:6)

10 Se agruparon los productos de amplificación de una única placa de 96 pocillos en un volumen igual, y se purificaron los productos de amplificación agrupados con perlas AMPureXP™ SPRI (Beckman-Coulter, Danvers, MA) según las instrucciones del fabricante. Se usó cada biblioteca agrupada purificada como molde para la amplificación de agrupamientos en una célula de flujo de kit de agrupamiento TruSeq v2 SR de Illumina (Illumina, San Diego, CA) según los protocolos del fabricante. Se procesó el portaobjetos en un instrumento HiSeq™ 2000 de Illumina (Illumina, San Diego, CA) para producir 56 bases de secuencia específica de locus a partir de un cebador 723 de secuencia izquierdo y se obtuvo una lectura diferenciada de 8 bases de secuencia específica de muestra del  
15 segundo cebador 725 de secuencia. Se recogieron un promedio de 903K lecturas sin procesar por muestra. Un promedio de 876K (97%) de las lecturas se asignaron a estructuras de ensayo esperadas.

20 La figura 8 muestra datos a modo de ejemplo para un subconjunto de las muestras de pacientes a partir de la segunda cohorte, que se analizaron todas en un ensayo multiplexado en un único carril de una ronda de secuenciación. Inicialmente se ejecutaron 96 muestras diferentes en esta ronda particular, pero posteriormente se excluyeron todas menos seis muestras de este conjunto analítico ya que no cumplían los umbrales de control de calidad de las muestras. Se calculó una media recortada para cada cromosoma 18 y cromosoma 21 para las muestras basándose en las lecturas producidas en el ensayo. Esta media recortada se calculó eliminando el 10% de recuentos altos y bajos para cada cromosoma por muestra. Se usaron los productos de amplificación detectados  
25 correspondientes a los diversos locus seleccionados para calcular una métrica de proporción de cromosoma 21 y una métrica de proporción de cromosoma 18 para cada muestra. Para la proporción de cromosoma 21, se calculó esto como la media recortada de recuentos en los 384 locus de cromosoma 21 seleccionados dividida entre la suma de medias recortadas de recuentos para los 576 locus de cromosoma 21 y 576 locus de cromosoma 18 para cada  
30 muestra.

35 En promedio, se observaron 834 recuentos de lectura por locus seleccionado en las muestras maternas de la primera cohorte, y se observaron 664 recuentos de lectura por locus seleccionado a partir de la segunda cohorte. Se usaron estos recuentos para calcular las puntuaciones z de proporción de cromosoma para el cromosoma 21 y cromosoma 18.

40 En resumen, se calcularon las puntuaciones z escalando la mediana por recuento de locus hasta un valor común (por ejemplo, 1000) para cada muestra, y se transformaron los recuentos escalados mediante log en base 2. Se realizaron un pulido de medianas y modelado lineal logarítmico de RMA (Bolstad, B.M *et al.* (2003) *Bioinformatics* 19(2):185-193; Rafael. A. (2003) *Nucleic Acids Research* 31(4):e15; Irizarry, RA *et al.* (2003) *Biostatistics* 4(2):249-64) para estimar los efectos de cromosomas, los efectos de locus, los efectos de muestras y los residuos. Se fijaron los efectos de cromosomas estimados a un valor común, por ejemplo, 0, y se calculó  $2^{(\text{efecto de cromosomas} + \text{efecto de muestras} + \text{residuo})}$  para cada locus para crear recuentos normalizados. Se escalaron las puntuaciones Z usando censura iterativa de modo que tuviesen una media de 0 y una desviación estándar de 1.

45 Se usaron los datos obtenidos de la primera cohorte de muestras para determinar las puntuaciones z de la primera cohorte para el cromosoma 21 y cromosoma 18 y se ilustran en las figuras 9 y 10, respectivamente. Las muestras normales se muestran como rombos gris oscuro, y las muestras con una trisomía se muestran como rombos gris claro. 179/180 (99,4%) muestras normales (rombos gris oscuro) tenían puntuaciones z <3; una muestra normal tenía una puntuación z de cromosoma 21 de 3,4 y una puntuación z de cromosoma 18 de 3,0. 35/35 (100%) muestras de T21 y 7/7 (100%) de T18 tenían puntuaciones z de proporción de cromosomas >3. La puntuación z de T18 media era de 8,5, y el intervalo era de 5,8-10,9. La puntuación z de T21 media era de 11,5, y el intervalo era de 6,1-19,8.

55 Los datos proporcionados en la figura 8 se combinaron con datos de las muestras restantes de la segunda cohorte para determinar puntuaciones z para el cromosoma 21 y cromosoma 18 y se ilustran en las figuras 11 y 12, respectivamente. Las muestras normales se muestran como rombos gris oscuro, y las muestras con una trisomía se muestran como rombos gris claro. 125/125 muestras normales tenían puntuaciones z <3, 36/36 (100%) muestras de T21 y 8/8 (100%) muestras de T18 tenían puntuaciones z >3. La puntuación z de T18 media era de 9,5 y el intervalo era de 5,1-19,8. La puntuación z de T21 media era de 11,4 y el intervalo era de 3,4-21,8.

Además de la detección de aneuploidía en estas cohortes, también se usaron polimorfismos específicos para determinar el porcentaje de contribución fetal a las muestras maternas. La metodología general usada para la determinación de estos porcentajes de contribución fetal se describe en el documento US con n.º de serie 61/509.188 presentado el 19 de julio de 2011, que se incorpora como referencia en su totalidad.

5 En resumen, la secuenciación de determinados locus que tenían polimorfismos detectables identificó estos locus como informativos. Se usaron los recuentos de los locus identificados que tenían regiones polimórficas fetales diferentes de las regiones polimórficas maternas para calcular la contribución fetal aproximada para la muestra materna. Cada uno de los locus usados en el cálculo del porcentaje de contribución fetal a la muestra materna tenía un mínimo de 256 recuentos. Se ilustran conjuntos de datos de SNP a modo de ejemplo para este cálculo a 10 continuación en las tablas 2 y 3. Los datos correspondientes a los locus informativos identificados a partir de estos conjuntos se usaron en el cálculo del porcentaje de contribución. Los locus informativos se muestran en cada tabla en negrita.

15 Tabla 2: Detección de SNP y porcentaje fetal calculado para la muestra materna 1

Cromosoma/locus	Recuentos de SNP_1	Recuentos de SNP_2	Recuentos totales	Porcentaje fetal calculado
<b>Ch01_Lc067487</b>	<b>294</b>	<b>26</b>	<b>320</b>	0,210138
Ch01_Lc067489	187	167	354	
Ch01_Lc067490	389	1	390	
Ch01_Lc067491	233	113	346	
Ch01_Lc067492	0	267	267	
Ch01_Lc067493	145	132	277	
Ch01_Lc067495	106	172	278	
<b>Ch01_Lc067496</b>	<b>308</b>	<b>28</b>	<b>336</b>	
Ch01_Lc067497	298	0	298	
Ch01_Lc067498	310	1	311	
Ch01_Lc067499	256	1	257	
<b>Ch01_Lc067501</b>	<b>26</b>	<b>273</b>	<b>299</b>	
Ch01_Lc067503	296	0	296	
Ch01_Lc067504	134	149	283	
Ch02_Lc067508	0	337	337	
<b>Ch02_Lc067510</b>	<b>37</b>	<b>324</b>	<b>361</b>	
Ch02_Lc067511	138	147	285	
Ch02_Lc067512	180	251	431	
Ch02_Lc067514	0	383	383	
<b>Ch02_Lc067515</b>	<b>316</b>	<b>31</b>	<b>347</b>	
Ch02_Lc067516	276	2	278	
<b>Ch02_Lc067519</b>	<b>42</b>	<b>276</b>	<b>318</b>	
<b>Ch02_Lc067521</b>	<b>312</b>	<b>47</b>	<b>359</b>	
Ch02_Lc067522	158	170	328	
<b>Ch02_Lc067523</b>	<b>38</b>	<b>328</b>	<b>366</b>	
Ch02_Lc067524	177	127	304	
Ch02_Lc067525	292	0	292	
Ch02_Lc067526	361	0	361	
<b>Ch02_Lc067527</b>	<b>261</b>	<b>26</b>	<b>287</b>	
Ch02_Lc067529	140	146	286	
Ch03_Lc067530	0	268	268	
Ch03_Lc067531	217	178	395	
Ch03_Lc067532	245	153	398	

ES 2 685 465 T3

Ch03_Lc067533	1	286	287
<b>Ch03_Lc067534</b>	<b>384</b>	<b>38</b>	<b>422</b>
Ch03_Lc067535	192	114	306
<b>Ch03_Lc067537</b>	<b>32</b>	<b>276</b>	<b>308</b>
<b>Ch03_Lc067538</b>	<b>243</b>	<b>15</b>	<b>258</b>
Ch03_Lc067539	132	247	379
Ch03_Lc067540	162	105	267
<b>Ch03_Lc067541</b>	<b>239</b>	<b>35</b>	<b>274</b>
Ch03_Lc067542	3	406	409
<b>Ch03_Lc067544</b>	<b>2</b>	<b>271</b>	<b>273</b>
Ch03_Lc067545	373	0	373
Ch03_Lc067546	354	0	354
Ch03_Lc067547	1	256	257
Ch03_Lc067548	365	0	365
Ch03_Lc067549	187	111	298
<b>Ch04_Lc067550</b>	<b>33</b>	<b>312</b>	<b>345</b>
Ch04_Lc067552	323	1	324
Ch04_Lc067553	217	119	336
<b>Ch04_Lc067557</b>	<b>35</b>	<b>236</b>	<b>271</b>
Ch04_Lc067558	184	166	350
<b>Ch04_Lc067559</b>	<b>295</b>	<b>32</b>	<b>327</b>
Ch04_Lc067560	140	141	281
Ch04_Lc067561	160	123	283
Ch04_Lc067562	313	2	315
Ch04_Lc067566	142	191	333
Ch04_Lc067569	117	206	323
Ch05_Lc067570	0	403	403
Ch05_Lc067571	229	219	448
Ch05_Lc067572	185	134	319
<b>Ch05_Lc067573</b>	<b>271</b>	<b>22</b>	<b>293</b>
Ch05_Lc067575	261	142	403
Ch05_Lc067578	0	399	399
<b>Ch05_Lc067579</b>	<b>307</b>	<b>46</b>	<b>353</b>
Ch05_Lc067581	189	109	298
Ch05_Lc067582	0	268	268
Ch05_Lc067583	167	203	370
Ch05_Lc067585	209	119	328
<b>Ch05_Lc067586</b>	<b>3</b>	<b>327</b>	<b>330</b>
Ch05_Lc067587	321	0	321
Ch06_Lc067589	286	0	286
Ch06_Lc067590	2	344	346
Ch06_Lc067591	124	179	303
Ch06_Lc067592	0	330	330
Ch06_Lc067593	0	286	286
Ch06_Lc067594	396	2	398
Ch06_Lc067595	349	0	349

ES 2 685 465 T3

Ch06_Lc067597	340	1	341
Ch06_Lc067598	0	412	412
Ch06_Lc067599	182	93	275
<b>Ch06_Lc067600</b>	<b>44</b>	<b>307</b>	<b>351</b>
<b>Ch06_Lc067601</b>	<b>43</b>	<b>324</b>	<b>367</b>
Ch06_Lc067602	358	1	359
Ch07_Lc067603	160	141	301
Ch07_Lc067604	302	0	302
<b>Ch07_Lc067605</b>	<b>37</b>	<b>414</b>	<b>451</b>
Ch07_Lc067606	269	290	559
Ch07_Lc067607	166	159	325
Ch07_Lc067609	1	396	397
Ch07_Lc067610	225	134	359
<b>Ch07_Lc067611</b>	<b>48</b>	<b>391</b>	<b>439</b>
Ch07_Lc067612	2	333	335
Ch07_Lc067614	200	246	446
Ch07_Lc067615	188	184	372
Ch07_Lc067616	167	116	283
Ch07_Lc067617	204	186	390
Ch07_Lc067618	<b>281</b>	<b>28</b>	<b>309</b>
<b>Ch07_Lc067619</b>	<b>44</b>	<b>297</b>	<b>341</b>
Ch07_Lc067620	336	0	336
<b>Ch07_Lc067621</b>	<b>48</b>	<b>342</b>	<b>390</b>
Ch08_Lc067622	313	1	314
Ch08_Lc067623	414	0	414
Ch08_Lc067624	230	142	372
Ch08_Lc067625	0	377	377
<b>Ch08_Lc067626</b>	<b>41</b>	<b>357</b>	<b>398</b>
Ch08_Lc067627	133	258	391
Ch08_Lc067628	388	1	389
Ch08_Lc067629	348	0	348
<b>Ch08_Lc067630</b>	<b>37</b>	<b>314</b>	<b>351</b>
Ch08_Lc067631	185	129	314
<b>Ch08_Lc067632</b>	<b>49</b>	<b>308</b>	<b>357</b>
Ch08_Lc067633	186	195	381
Ch08_Lc067634	174	217	391
Ch08_Lc067635	161	152	313
Ch08_Lc067637	0	284	284
<b>Ch08_Lc067638</b>	<b>343</b>	<b>50</b>	<b>393</b>
Ch09_Lc067639	164	99	263
Ch09_Lc067640	185	186	371
Ch09_Lc067641	344	0	344
Ch09_Lc067642	294	0	294
<b>Ch09_Lc067643</b>	<b>36</b>	<b>336</b>	<b>372</b>
Ch09_Lc067644	221	144	365
Ch09_Lc067645	315	36	351

ES 2 685 465 T3

Ch09_Lc067646	141	143	284
<b>Ch09_Lc067647</b>	<b>33</b>	<b>270</b>	<b>303</b>
<b>Ch09_Lc067648</b>	<b>43</b>	<b>349</b>	<b>392</b>
Ch09_Lc067649	147	152	299
Ch09_Lc067650	201	187	388
Ch09_Lc067651	176	151	327
<b>Ch10_Lc067652</b>	<b>29</b>	<b>277</b>	<b>306</b>
Ch10_Lc067653	134	157	291
Ch10_Lc067654	174	196	370
Ch10_Lc067655	189	181	370
Ch10_Lc067656	125	174	299
Ch10_Lc067657	375	2	377
Ch10_Lc067658	0	345	345
Ch10_Lc067659	204	174	378
Ch10_Lc067661	236	271	507
<b>Ch10_Lc067662</b>	<b>39</b>	<b>325</b>	<b>364</b>
Ch11_Lc067663	2	378	380
<b>Ch11_Lc067664</b>	<b>42</b>	<b>298</b>	<b>340</b>
Ch11_Lc067666	196	200	396
Ch11_Lc067667	220	164	384
<b>Ch11_Lc067668</b>	<b>20</b>	<b>290</b>	<b>310</b>
Ch11_Lc067670	1	356	357
Ch12_Lc067671	212	195	407
Ch12_Lc067673	1	298	299
Ch12_Lc067674	242	30	272
Ch12_Lc067675	2	292	294
Ch12_Lc067676	1	381	382
Ch12_Lc067677	139	179	318

Tabla 3: Detección de SNP de la muestra 2 y porcentaje fetal calculado

Cromosoma/locus	Recuentos de SNP_1	Recuentos de SNP_2	Recuentos totales	Porcentaje fetal calculado
Ch01_Lc067487	181	134	315	0,096075
<b>Ch01_Lc067489</b>	<b>18</b>	<b>337</b>	<b>355</b>	
<b>Ch01_Lc067490</b>	<b>17</b>	<b>333</b>	<b>350</b>	
Ch01_Lc067491	356	0	356	
<b>Ch01_Lc067492</b>	<b>12</b>	<b>264</b>	<b>276</b>	
<b>Ch01_Lc067493</b>	<b>7</b>	<b>385</b>	<b>392</b>	
Ch01_Lc067494	140	118	258	
Ch01_Lc067495	163	118	281	
Ch01_Lc067496	198	172	370	
<b>Ch01_Lc067498</b>	<b>15</b>	<b>301</b>	<b>316</b>	
Ch01_Lc067499	162	170	332	
<b>Ch01_Lc067501</b>	<b>11</b>	<b>252</b>	<b>263</b>	
Ch01_Lc067502	129	130	259	
Ch01_Lc067503	148	172	320	

ES 2 685 465 T3

Ch01_Lc067504	157	146	303
Ch02_Lc067508	188	196	384
Ch02_Lc067510	1	356	357
Ch02_Lc067511	3	308	311
Ch02_Lc067512	262	193	455
Ch02_Lc067513	318	0	318
Ch02_Lc067514	1	440	441
Ch02_Lc067515	169	166	335
Ch02_Lc067516	189	149	338
Ch02_Lc067519	218	133	351
Ch02_Lc067521	153	180	333
<b>Ch02_Lc067522</b>	<b>14</b>	<b>326</b>	<b>340</b>
Ch02_Lc067523	180	178	358
Ch02_Lc067524	330	1	331
Ch02_Lc067526	202	185	387
Ch02_Lc067527	149	192	341
Ch02_Lc067529	140	160	300
Ch03_Lc067530	132	128	260
<b>Ch03_Lc067531</b>	<b>20</b>	<b>392</b>	<b>412</b>
Ch03_Lc067532	202	270	472
Ch03_Lc067533	328	0	328
Ch03_Lc067534	224	223	447
Ch03_Lc067535	188	142	330
Ch03_Lc067537	315	1	316
Ch03_Lc067538	265	0	265
Ch03_Lc067539	191	214	405
Ch03_Lc067540	166	124	290
Ch03_Lc067542	256	208	464
Ch03_Lc067543	139	123	262
Ch03_Lc067544	190	180	370
<b>Ch03_Lc067545</b>	<b>378</b>	<b>12</b>	<b>390</b>
<b>Ch03_Lc067546</b>	<b>22</b>	<b>352</b>	<b>374</b>
Ch03_Lc067547	184	162	346
Ch03_Lc067548	363	0	363
Ch03_Lc067549	0	312	312
Ch04_Lc067550	162	156	318
Ch04_Lc067552	331	0	331
Ch04_Lc067553	225	155	380
Ch04_Lc067555	121	146	267
<b>Ch04_Lc067557</b>	<b>284</b>	<b>13</b>	<b>297</b>
Ch04_Lc067558	229	208	437
<b>Ch04_Lc067559</b>	<b>7</b>	<b>311</b>	<b>318</b>
Ch04_Lc067560	154	136	290
<b>Ch04_Lc067561</b>	<b>12</b>	<b>258</b>	<b>270</b>
Ch04_Lc067562	410	1	411
<b>Ch04_Lc067566</b>	<b>320</b>	<b>18</b>	<b>338</b>

ES 2 685 465 T3

Ch04_Lc067569	0	289	289
<b>Ch05_Lc067570</b>	<b>19</b>	<b>444</b>	<b>463</b>
Ch05_Lc067571	498	0	498
Ch05_Lc067572	169	182	351
Ch05_Lc067573	294	0	294
Ch05_Lc067575	422	0	422
<b>Ch05_Lc067578</b>	<b>18</b>	<b>388</b>	<b>406</b>
<b>Ch05_Lc067579</b>	<b>17</b>	<b>304</b>	<b>321</b>
Ch05_Lc067580	156	149	305
Ch05_Lc067581	303	19	322
<b>Ch05_Lc067583</b>	<b>23</b>	<b>347</b>	<b>370</b>
<b>Ch05_Lc067585</b>	<b>22</b>	<b>293</b>	<b>315</b>
Ch05_Lc067586	391	0	391
Ch05_Lc067587	434	1	435
Ch05_Lc067588	157	129	286
Ch06_Lc067589	274	0	274
<b>Ch06_Lc067590</b>	<b>23</b>	<b>320</b>	<b>343</b>
<b>Ch06_Lc067591</b>	<b>10</b>	<b>342</b>	<b>352</b>
Ch06_Lc067592	181	177	358
Ch06_Lc067593	0	296	296
Ch06_Lc067594	267	200	467
Ch06_Lc067595	212	201	413
<b>Ch06_Lc067596</b>	<b>329</b>	<b>12</b>	<b>341</b>
Ch06_Lc067597	319	1	320
Ch06_Lc067598	243	186	429
Ch06_Lc067600	0	341	341
Ch06_Lc067601	417	0	417
Ch06_Lc067602	1	340	341
Ch07_Lc067603	168	185	353
<b>Ch07_Lc067604</b>	<b>333</b>	<b>11</b>	<b>344</b>
Ch07_Lc067605	211	287	498
Ch07_Lc067606	2	542	544
Ch07_Lc067607	310	0	310
Ch07_Lc067609	178	189	367
Ch07_Lc067610	425	0	425
Ch07_Lc067611	1	449	450
Ch07_Lc067612	169	149	318
Ch07_Lc067614	0	446	446
Ch07_Lc067615	0	427	427
<b>Ch07_Lc067616</b>	<b>13</b>	<b>278</b>	<b>291</b>
Ch07_Lc067617	239	246	485
<b>Ch07_Lc067618</b>	<b>17</b>	<b>267</b>	<b>284</b>
Ch07_Lc067619	0	335	335
Ch07_Lc067620	319	0	319
Ch07_Lc067621	0	354	354
<b>Ch08_Lc067622</b>	<b>25</b>	<b>232</b>	<b>257</b>

ES 2 685 465 T3

Ch08_Lc067623	470	0	470
Ch08_Lc067624	0	376	376
Ch08_Lc067625	212	202	414
Ch08_Lc067626	377	0	377
<b>Ch08_Lc067627</b>	<b>379</b>	<b>16</b>	<b>395</b>
Ch08_Lc067628	189	210	399
<b>Ch08_Lc067629</b>	<b>16</b>	<b>338</b>	<b>354</b>
Ch08_Lc067630	152	153	305
Ch08_Lc067631	0	379	379
<b>Ch08_Lc067632</b>	<b>25</b>	<b>355</b>	<b>380</b>
Ch08_Lc067633	186	236	422
Ch08_Lc067634	2	375	377
Ch08_Lc067635	169	159	328
<b>Ch08_Lc067636</b>	<b>13</b>	<b>274</b>	<b>287</b>
<b>Ch08_Lc067637</b>	<b>373</b>	<b>11</b>	<b>384</b>
<b>Ch08_Lc067638</b>	<b>28</b>	<b>431</b>	<b>459</b>
Ch09_Lc067640	367	0	367
Ch09_Lc067641	359	0	359
Ch09_Lc067642	307	1	308
<b>Ch09_Lc067643</b>	<b>350</b>	<b>32</b>	<b>382</b>
Ch09_Lc067644	168	241	409
Ch09_Lc067645	174	179	353
Ch09_Lc067646	133	177	310
Ch09_Lc067647	199	188	387
<b>Ch09_Lc067648</b>	<b>24</b>	<b>450</b>	<b>474</b>
Ch09_Lc067649	340	0	340
<b>Ch09_Lc067650</b>	<b>22</b>	<b>348</b>	<b>370</b>
<b>Ch09_Lc067651</b>	<b>20</b>	<b>365</b>	<b>385</b>
Ch10_Lc067652	0	292	292
Ch10_Lc067653	0	279	279
Ch10_Lc067654	0	396	396
Ch10_Lc067655	160	170	330
Ch10_Lc067656	175	121	296
Ch10_Lc067657	212	188	400
Ch10_Lc067658	0	356	356
<b>Ch10_Lc067659</b>	<b>400</b>	<b>13</b>	<b>413</b>
Ch10_Lc067661	263	268	531
<b>Ch10_Lc067662</b>	<b>20</b>	<b>324</b>	<b>344</b>
<b>Ch11_Lc067663</b>	<b>12</b>	<b>357</b>	<b>369</b>
Ch11_Lc067664	142	179	321
Ch11_Lc067665	278	0	278
Ch11_Lc067666	180	232	412
<b>Ch11_Lc067667</b>	<b>13</b>	<b>438</b>	<b>451</b>
Ch11_Lc067668	0	365	365
<b>Ch11_Lc067669</b>	<b>15</b>	<b>263</b>	<b>278</b>
Ch11_Lc067670	0	374	374

Ch12_Lc067671	233	183	416
Ch12_Lc067672	0	269	269
<b>Ch12_Lc067673</b>	<b>412</b>	<b>11</b>	<b>423</b>
<b>Ch12_Lc067676</b>	<b>37</b>	<b>436</b>	<b>473</b>
Ch12_Lc067677	168	197	365

Los datos para estos polimorfismos se obtuvieron en el mismo conjunto de datos que los datos de aneuploidía ilustrados en las figuras 11 y 12. Por tanto, un único ensayo demostró la capacidad para identificar aneuploidía fetal, y las diferencias polimórficas entre locus fetales y maternos permitió la identificación de locus informativos y el cálculo del porcentaje fetal estimado de ADNc en la muestra basándose en los locus informativos.

**Lista de secuencias**

- 10 <110> Tandem Diagnostics, Inc.  
Sparks, Andrew  
Struble, Craig  
Wang, Eric  
Oliphant, Arnold
- 15 <120> SISTEMAS DE ENSAYO PARA ANÁLISIS GENÉTICO  
  
<130> TAND012PCT  
  
<150> 13/013.732
- 20 <151> 25-01-2011  
<150> 61/371605  
  
<151> 06-08-2010
- 25 <160> 8  
  
<170> PatentIn versión 3.5
- 30 <210> 1  
<211> 24  
<212> ADN  
<213> SECUENCIA ARTIFICIAL
- 35 <220>  
<223> Oligonucleótido sintético  
  
<400> 1  
tacaccggcg ttatgcgtcg agac 24
- 40 <210> 2  
<211> 24  
<212> ADN  
<213> ORGANISMO ARTIFICIAL
- 45 <220>  
<223> Oligonucleótido sintético  
  
<400> 2  
attgcgggga ccgatgatcg cgtc 24
- 50 <210> 3  
<211> 48  
<212> ADN  
<213> SECUENCIA ARTIFICIAL
- 55 <220>  
<223> Oligonucleótido sintético

<400> 3  
 taatgatacg gcgaccaccg agatctacac cggcgttatg cgtcgaga 48

<210> 4  
 5 <211> 53  
 <212> ADN  
 <213> ORGANISMO ARTIFICIAL

<220>  
 10 <223> Oligonucleótido sintético

<220>  
 <221> Índice  
 <222> (26)..(26)  
 15 <223> n es un índice de secuencia.

<220>  
 <221> misc\_feature  
 <222> (26)..(26)  
 20 <223> n es a, c, g o t

<400> 4  
 tcaagcagaa gacggcatac gagatnaaac gacgcatca tcggtccccg caa 53

25 <210> 5  
 <211> 21  
 <212> ADN  
 <213> SECUENCIA ARTIFICIAL

30 <220>  
 <223> Oligonucleótido sintético

<400> 5  
 taatgatacg gcgaccaccg a 21

35 <210> 6  
 <211> 25  
 <212> ADN  
 <213> SECUENCIA ARTIFICIAL

40 <220>  
 <223> ATCTCGTATGCCGTCTTCTGCTTGA

<400> 6  
 45 atctcgatg ccgtctctg cttga 25

<210> 7  
 <211> 48  
 <212> ADN  
 50 <213> SECUENCIA ARTIFICIAL

<220>  
 <223> Oligonucleótido sintético

55 <400> 7  
 aatgatacgg cgaccaccga gatctacacc ggcggtatgc gtcgagac 48

<210> 8  
 <211> 60  
 60 <212> ADN  
 <213> SECUENCIA ARTIFICIAL

<220>  
 <223> Oligonucleótido sintético

65 <220>

## ES 2 685 465 T3

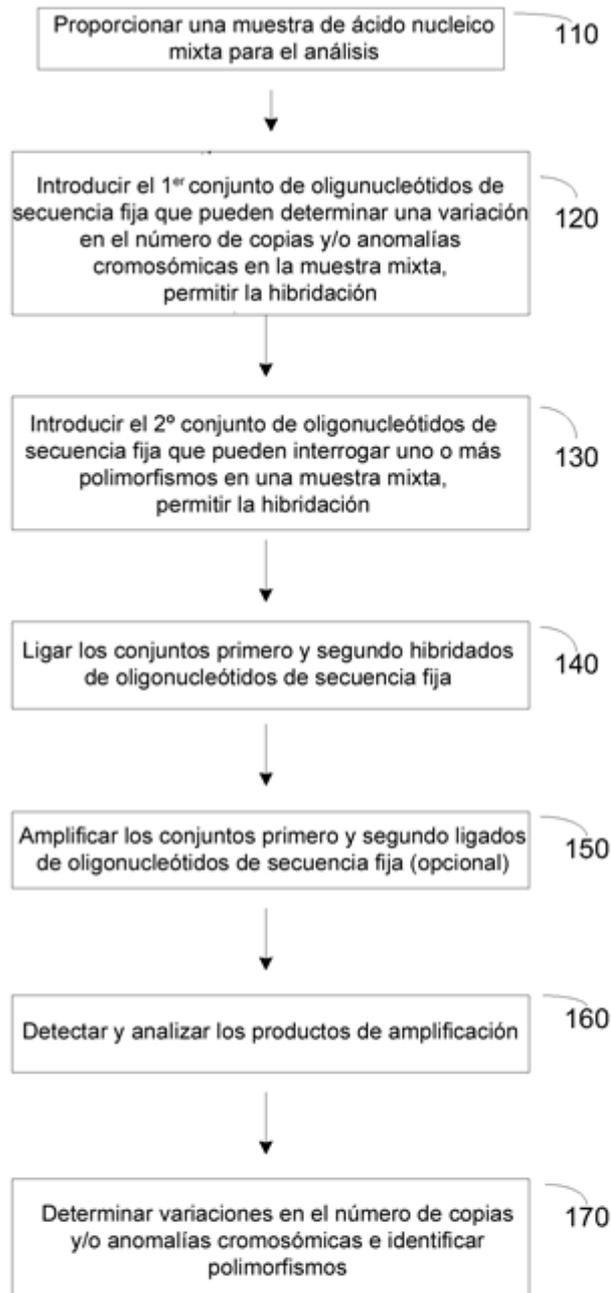
<221> n  
<222> (26)..(32)  
<223> n = G, A, T o C

5 <220>  
<221> misc\_feature  
<222> (26)..(32)  
<223> n es a, c, g o t

10 <400> 8  
tcaagcagaa gacggcatac gagatnnnn nnaaacgacg cgatcatcgg tccccgcaat 60

**REIVINDICACIONES**

1. Método de ensayo único para el cálculo de la contribución de fuente mediante una fuente principal y una secundaria y detección de la presencia o ausencia de variación en el número de copias (VNC) en una o más regiones genómicas en una muestra mixta, siendo la muestra mixta una muestra materna que comprende tanto ADN materno como fetal, comprendiendo el ensayo las etapas de:
  - 5 introducir un primer conjunto de oligonucleótidos de secuencia fija en la muestra mixta en un recipiente en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más locus en o asociados con una región genómica, en el que al menos uno de los oligonucleótidos de secuencia fija comprende una región de cebador universal;
  - 10 introducir un segundo conjunto de oligonucleótidos de secuencia fija en la muestra mixta en el recipiente en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más locus informativos, en el que al menos uno de los oligonucleótidos de secuencia fija comprende una región de cebador universal;
  - 15 en el que el número de locus medidos es menos de 2000 por cromosoma;
  - 20 ligar los oligonucleótidos hibridados para crear productos de ligación contiguos;
  - amplificar los productos de ligación contiguos usando la región de cebador universal para crear productos de amplificación;
  - 25 en el que oligonucleótidos de secuencia fija no hibridados se retiran antes de la amplificación de los productos de ligación contiguos;
  - 30 detectar la secuencia de los productos de amplificación usando secuenciación de alto rendimiento, en el que cada locus se mide en promedio al menos 100 veces;
  - 35 calcular la contribución de fuente determinando el porcentaje de ADN fetal en la muestra mixta, en el que el porcentaje de ADN fetal detectado mediante locus informativos corresponde a la proporción de lecturas derivadas de alelos no maternos frente a maternos en locus en donde puede hacerse una distinción de este tipo; y
  - usar las frecuencias relativas de los locus para determinar la presencia o ausencia de una variación en el número de copias.
2. Método de ensayo según la reivindicación 1, en el que las regiones de cebador universal se usan en la determinación de secuencia de los productos de amplificación.
3. Método de ensayo según la reivindicación 1, en el que uno o ambos conjuntos de oligonucleótidos de secuencia fija comprenden una sonda precírculo.
- 45 4. Método de ensayo según la reivindicación 1, en el que los productos de amplificación se aíslan antes de la detección.
5. Método de ensayo según la reivindicación 3, en el que los productos de amplificación se aíslan como moléculas individuales antes de la detección.
- 50 6. Método de ensayo según la reivindicación 5, en el que los productos de amplificación aislados individuales se amplifican adicionalmente para crear copias idénticas de todos o una porción de los productos de amplificación individuales antes de la detección.
- 55 7. Método de ensayo según la reivindicación 5, en el que los productos de amplificación aislados individuales se amplifican adicionalmente para crear copias idénticas de moléculas complementarias a todos o una porción de los productos de amplificación individuales antes de la detección.



100

FIGURA 1

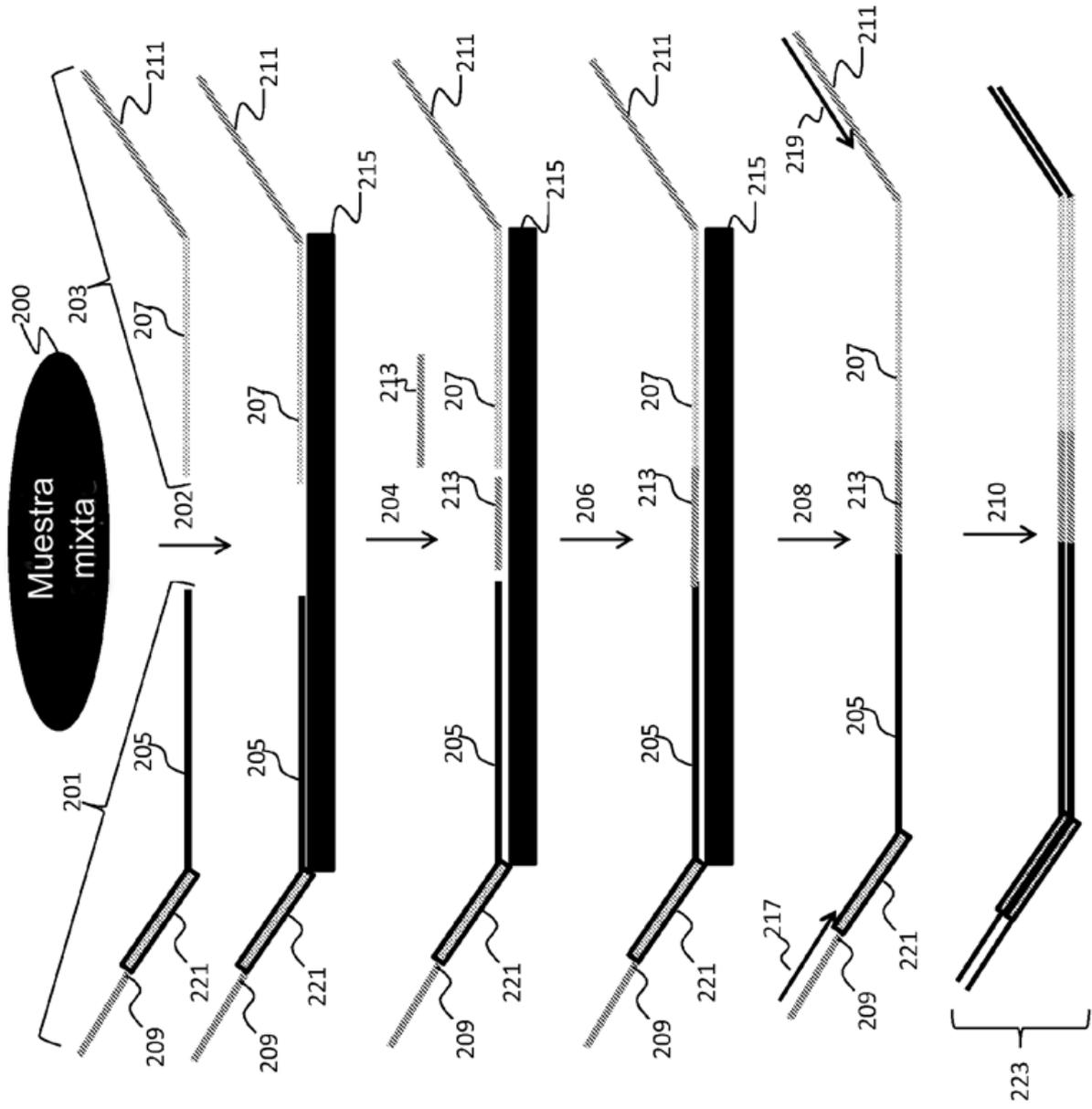


FIG. 2

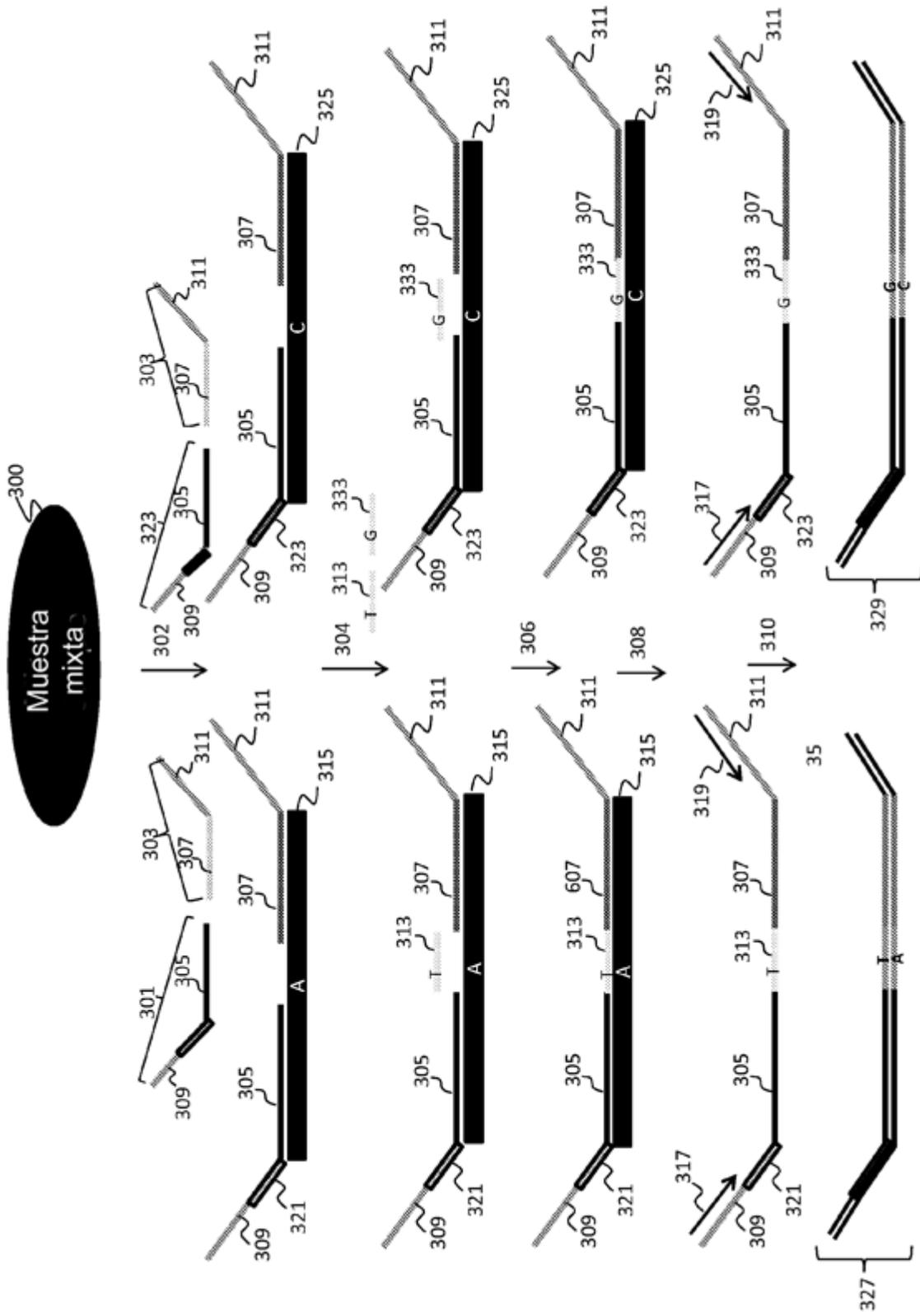


FIG. 3

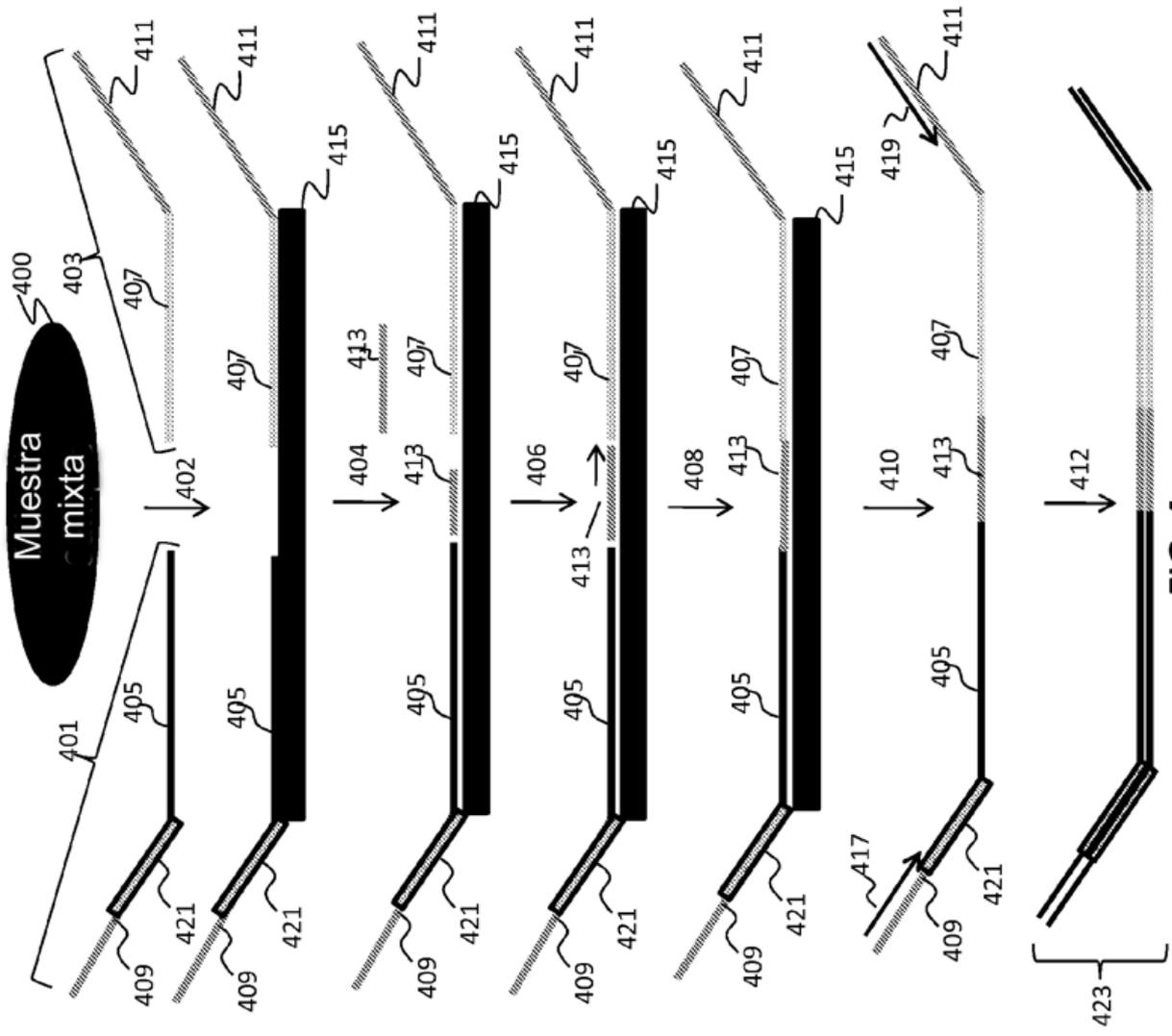


FIG. 4

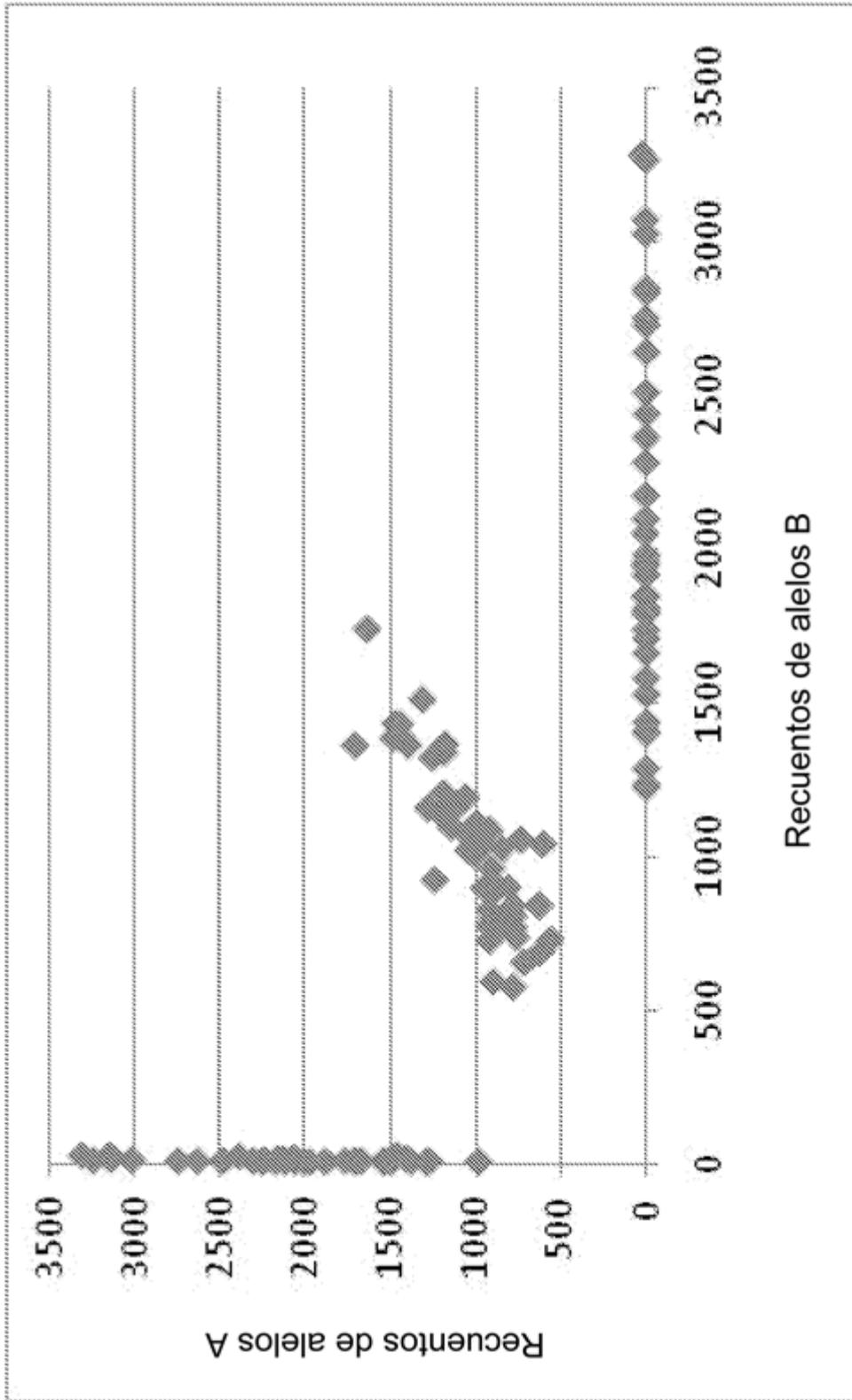


FIG. 5

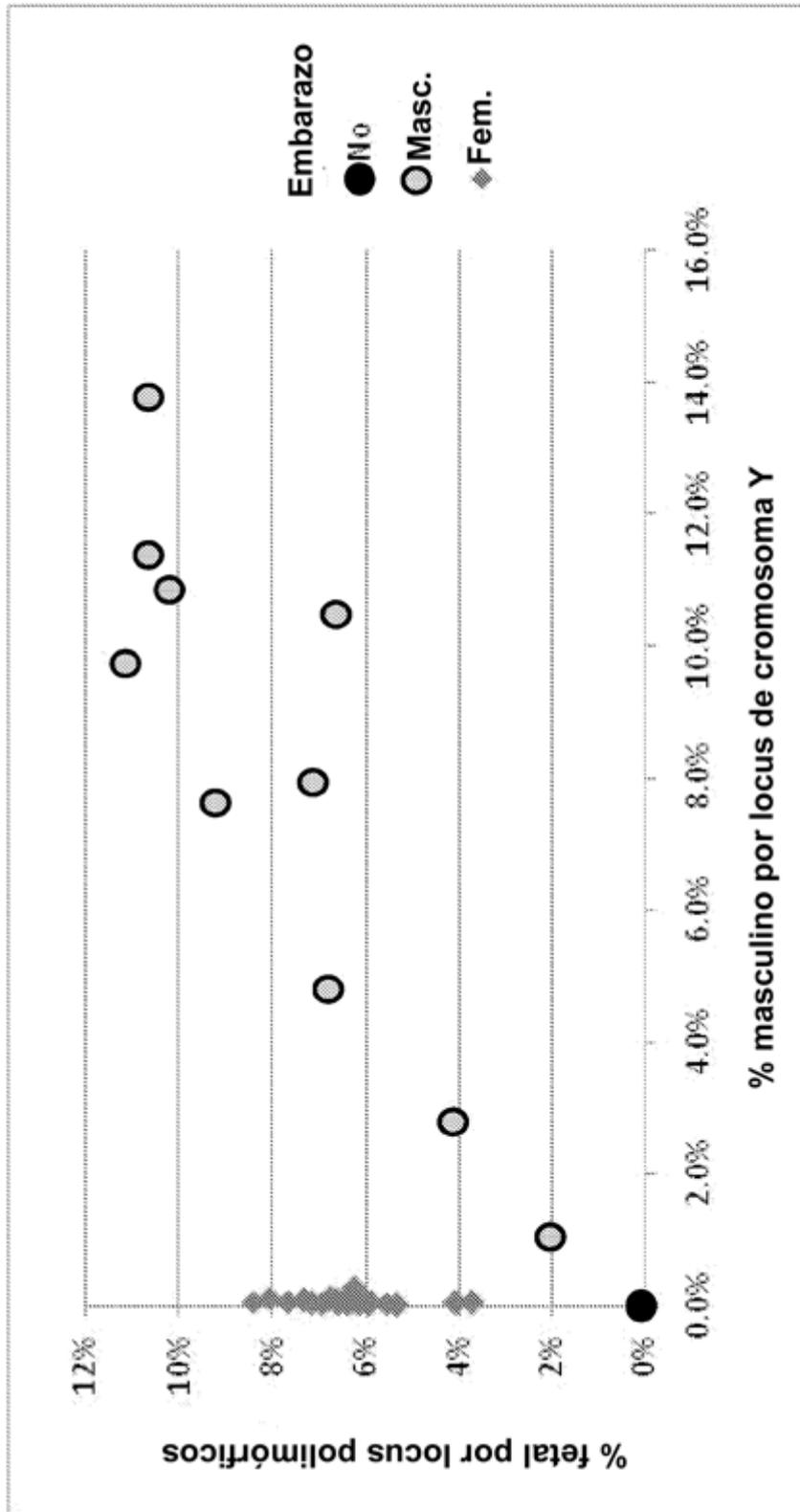


FIG. 6

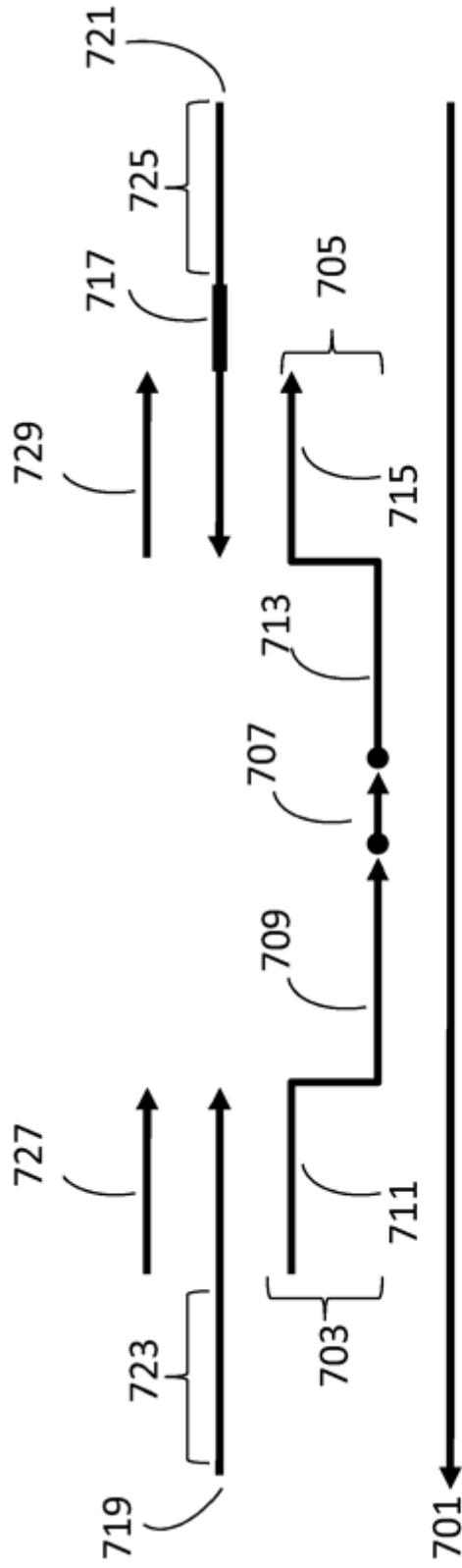


FIG. 7

Muestra	Etiqueta de muestra	Género fetal	Equiv. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad de gest. (día)	Dif. pol. de n.º de locus	% fetal	Media 18	Media 21	Prop18	Prop21	ZStat18	ZStat21
1	TGTCGCC	M	2293	46XX	46XY	45	110	38	0.210138	1005.426	999.4412	0.501493	0.498507	0.48436	-0.48436
2	ATCCTCC	F	3593	46XX	46XX	37	131	45	0.096075	996.6228	998.7823	0.499459	0.500541	-0.84491	0.84491
4	GCCAACC	F	1469	46XX	46XX	34	121	26	0.235051	991.0624	997.5193	0.498376	0.501624	-1.55236	1.552358
5	CCACGAC	F	1296	46XX	46XX	36	129	36	0.140539	1010.048	1004.468	0.501385	0.498615	0.413907	-0.41391
6	TCCTCGT	F	2110	46XX	46XX	33	136	24	0.049392	994.7681	1000.632	0.498531	0.501469	-1.45169	1.451689
7	CCTTGCT	F	2772	46XX	46XX	43	111	42	0.095395	997.3443	986.4745	0.50274	0.49726	1.299353	-1.29935
8	TCACICT	M	2506	46XX	46XY	34	126	32	0.158725	1004.807	996.2886	0.502128	0.497872	0.899842	-0.89984
9	CATCTTC	M	3996	46XX	46XY	36	121	34	0.110817	1005.714	1002.283	0.500854	0.499146	0.067201	-0.0672
10	TTGGCCG	F	4032	46XX	46XX	42	84	29	0.159776	999.7097	1000.399	0.499828	0.500172	-0.60393	0.60393
11	GTTGCC	F	4212	46XX	46XX	36	175	39	0.119147	1004.796	995.1465	0.502412	0.497588	1.085457	-1.08546
12	GCCCTTG	F	3816	46XX	46XX	36	111	27	0.085664	1002.909	999.3516	0.500888	0.499112	0.089291	-0.08929
13	CTTACAC	F	2160	46XX	46XX	41	123	40	0.117371	1002.012	995.0877	0.501734	0.498266	0.641893	-0.64189
14	TCGCAGC	F	2790	46XX	46XX	51	125	36	0.119314	1008.515	999.3439	0.502284	0.497716	1.001374	-1.00137
15	CGGCCAT	F	2783	46XX	46XX	51	125	50	0.123598	1005.525	998.867	0.501661	0.498339	0.594272	-0.59427
16	CCAGCTG	M	2272	46XX	46XY	34	110	40	0.119257	1010.173	1004.256	0.501468	0.498532	0.468561	-0.46856
17	AGCACCT	M	2401	46XX	46XY	34	140	43	0.111573	988.7338	986.5967	0.500541	0.499459	-0.13768	0.137679
18	CGATACC	F	3960	46XX	46XX	41	126	27	0.099872	996.8338	999.2182	0.499403	0.500597	-0.88162	0.88162
19	CCCGAAT	M	3816	46XX	46XY	37	133	30	0.095735	999.3982	991.0248	0.502103	0.497897	0.883556	-0.88356
20	CCCAGGG	M	3474	46XX	46XY	38	123	32	0.115184	1006.395	992.8395	0.50339	0.49661	1.724514	-1.72451
21	CAGCAGC	M	1080	46XX	46XY	38	97	28	0.13036	994.7498	1003.364	0.497844	0.502156	-1.90014	1.900139

FIG. 8

22	CTCTATG	M	2002	46XX	46XY	30	123	30	0.08322	994.7936	999.994	0.498697	0.501303	-1.3432	1.343199
	<b>Muestra de muestra</b>	<b>Etiqueta fetal</b>	<b>Género fetal</b>	<b>Equiv. de genoma materno</b>	<b>Cariotipo fetal</b>	<b>Edad de la madre</b>	<b>Edad de gest. (día)</b>	<b>Locus informativos</b>	<b>% fetal</b>	<b>Media 18</b>	<b>Media 21</b>	<b>Prop18</b>	<b>Prop21</b>	<b>ZStat18</b>	<b>ZStat21</b>
23	ACTTTAC	M	3042	46XX	46XY	35	130	33	0.138751	1004.97	1003.101	0.500465	0.499535	-0.18713	0.187128
24	CGTATCG	F	1922	46XX	46XX	23	159	29	0.254234	1003.218	1006.4	0.499208	0.500792	-1.00868	1.008678
25	AACCAIT	M	3708	46XX	46XY	38	148	31	0.145101	996.0673	995.7268	0.500085	0.499915	-0.43538	0.435383
26	CTGGGGC	M	4032	46XX	46XY	30	126	27	0.077438	1005.191	1003.628	0.500389	0.499611	-0.23694	0.23694
27	TGCCGAG	F	2383	46XX	46XX	35	138	32	0.131712	996.445	1002.698	0.498436	0.501564	-1.51345	1.513451
28	ATACCAG	F	4104	46XX	46XX	23	99	33	0.099894	991.6721	1001.903	0.497434	0.502566	-2.16839	2.168385
29	GAAACCG	F	2437	46XX	46XX	33	149	25	0.067606	1004.187	1006.778	0.499356	0.500644	-0.91231	0.912305
30	AAAGGCC	F	2765	46XX	46XX	23	126	29	0.144098	1001.308	995.2881	0.501508	0.498492	0.494093	-0.49409
31	TCTAAT	M	2416	46XX	46XY	23	119	28	0.105371	1006.535	1005.069	0.500364	0.499636	-0.25309	0.25309
32	GTTGATC	M	2653	46XX	46XY	34	148	29	0.164518	1003.738	998.0583	0.501419	0.498581	0.435957	-0.43596
33	GATGCT	M	7488	46XX	46XY	36	148	21	0.056835	1012.154	1002.257	0.502456	0.497544	1.114241	-1.11424
34	AGTCTGT	M	2653	46XX	46XY	26	140	39	0.21197	1004.959	996.9646	0.501997	0.498003	0.813799	-0.8138
35	AATCTG	F	3780	46XX	46XX	35	133	43	0.259642	997.0006	997.6736	0.499831	0.500169	-0.60149	0.601493
36	CTAAGIT	M	8136	46XX	46XY	26	151	33	0.082773	1007.02	1006.742	0.500069	0.499931	-0.44599	0.445992
37	GCGGTT	M	2228	46XX	46XY	37	114	37	0.15551	1004.912	1003.615	0.500323	0.499677	-0.28021	0.280209
38	TATAGGC	F	4248	46XX	46XX	27	165	48	0.064569	1014.599	995.8518	0.504662	0.495338	2.556064	-2.55606
39	TAGGTAC	M	3960	46XX	46XY	35	124	41	0.167184	997.8251	998.4044	0.499855	0.500145	-0.58608	0.586081
40	CAATTGG	F	4140	46XX	46XX	24	127	43	0.129398	997.6805	996.6003	0.500271	0.499729	-0.31423	0.314232
41	TCATAAG	M	4320	46XX	46XY	39	139	29	0.227198	1002.31	991.6654	0.502669	0.497331	1.253298	-1.2533
42	GAACGGT	F	3744	46XX	46XX	38	118	39	0.094147	1007.442	994.0998	0.503333	0.496667	1.687191	-1.68719

FIG. 8 (cont.)

43	AAGCGGT	F	6984	46XX	46XX	37	130	36	0.228387	998.8644	995.701	0.500793	0.499207	0.027059	-0.02706
	Muestra de muestra	Etiqueta fetal	Género fetal	Equiv. de genoma materno	Cariotipo fetal	Edad de la madre	Edad de gest. (día)	Locus informativos	% fetal	Media 18	Media 21	Prop18	Prop21	ZStat18	ZStat21
44	AGGAATC	F	5112	46XX	46XX	18	188	39	0.244261	1001.465	995.804	0.501417	0.498583	0.435058	-0.43506
45	AACATAG	F	3503	46XX	46XX	41	141	22	0.106969	999.5094	990.8389	0.502178	0.497822	0.932375	-0.93238
46	TTTGGAT	F	2606	46XX	46XX	39	143	32	0.196098	1001.959	1000.069	0.500472	0.499528	-0.18279	0.182795
47	TTGATGT	F	5292	46XX	46XX	36	111	30	0.109084	1004.708	1002.777	0.500481	0.499519	-0.17694	0.176939
48	ATGGTTG	M	5184	46XX	46XY	24	160	40	0.060517	1010.204	1006.291	0.500097	0.49903	0.143003	-0.143
49	GGTAGT	M	3996	46XX	46XY	23	153	30	0.151707	1003.952	998.371	0.501394	0.498606	0.419553	-0.41955
50	AGATGAT	F	5256	46XX	46XX	26	118	26	0.126345	998.1654	996.1826	0.500497	0.499503	-0.16635	0.166347
51	GTGAAAG	M	4068	46XX	46XY	33	140	29	0.214532	1004.428	996.9586	0.501866	0.498134	0.728436	-0.72844
52	AGTGAAG	M	11412	46XX	46XY	18	196	24	0.140258	1007.84	998.7992	0.502253	0.497747	0.9811	-0.9811
53	ACAACGC	M	3146	46XX	46XY	21	147	32	0.180869	1010.525	999.8358	0.502659	0.497341	1.24639	-1.24639
54	GACGAGG	F	5508	46XX	46XX	35	114	32	0.053622	994.8149	1005.452	0.497341	0.502659	-2.22916	2.229159
55	GGAATAC	F	6624	46XX	46XX	28	140	37	0.21238	999.7187	1006.639	0.498275	0.501725	-1.61846	1.618464
56	GTTGCG	M	6192	46XX	46XY	44	115	25	0.083131	1009.487	999.2224	0.502555	0.497445	1.178758	-1.17876
57	GTCCTAT	F	7056	46XX	46XX	25	115	37	0.135201	1005.611	996.4	0.5023	0.4977	1.01226	-1.01226
58	TAGTGT	F	5148	46XX	46XX	22	99	38	0.124127	996.363	999.6895	0.499167	0.500833	-1.03587	1.035871
59	TGCTTC	M	4104	46XX	46XY	30	197	28	0.310846	1004.509	1009.571	0.498743	0.501257	-1.31256	1.312564
60	TCTGTG	M	5364	46XX	46XY	27	122	26	0.118328	997.1746	992.0824	0.50128	0.49872	0.345307	-0.34531
61	TGGGCTC	F	3636	46XX	46XX	28	178	30	0.158146	989.5388	999.8624	0.497405	0.502595	-2.18709	2.187092
62	TGAGTAT	M	17856	46XX	46XY	NA	NA	45	0.064088	993.7189	996.5081	0.499299	0.500701	-0.94924	0.949237
63	TGCAAT	F	5004	46XX	46XX	35	109	39	0.076783	1003.939	992.7595	0.5028	0.4972	1.33858	-1.33858

FIG. 8 (cont.)

64	TATCAAT	M	8712	46XX	46XY	35	119	33	0.077373	1004.37	996.8203	0.501886	0.498114	0.741595	-0.74159
Muestra de muestra		Etiqueta fetal	Género fetal	Equiv. de genoma materno	Cariotipo fetal	Edad de la madre	Edad de gest. (día)	Locus informativos	% fetal	Media 18	Media 21	Prop18	Prop21	ZStat18	ZStat21
65	TAAGAGT	F	2552	46XX	46XX+13	35	90	41	0.144442	999.7136	988.3962	0.502846	0.497154	1.369064	-1.36906
66	TAAAGAT	F	7668	46XX	46XX+13	30	141	37	0.057703	998.6741	990.8426	0.501968	0.498032	0.795151	-0.79515
67	TCTGCAC	M	3996	46XX	46XY+13	35	161	26	0.21564	1001.317	998.1811	0.500784	0.499216	0.021337	-0.02134
68	GTTGGCT	F	7560	46XX	46XX+21	35	125	28	0.092562	983.1548	1024.755	0.489641	0.510359	-7.2619	7.2619
69	GTCGACTT	F	13248	46XX	46XX+21	19	132	30	0.085949	980.027	1020.997	0.489763	0.510237	-7.18225	7.182254
70	GGTCAGC	F	7452	46XX	46XX+21	40	91	31	0.056151	990.5328	1011.076	0.494868	0.505132	-3.84532	3.845324
71	GGCCGAC	M	5508	46XX	46XY+21	18	166	32	0.140102	965.1286	1037.339	0.48197	0.51803	-12.2758	12.27579
73	GGCTAAC	M	5184	46XX	46XY+18	29	147	29	0.157072	1044.147	960.7049	0.52081	0.47919	13.11012	-13.1101
75	GAGTGGC	F	2689	46XX	46XX+18	40	93	24	0.129256	1038.787	958.3378	0.520141	0.479859	12.67306	-12.6731
76	GAATGAC	F	7632	46XX	46XX+18	43	133	24	0.048391	1015.894	982.9916	0.50823	0.49177	4.887983	-4.88798
78	GCTTGTG	F	7128	46XX	46XX+18	25	128	30	0.06035	1018.978	985.6703	0.508308	0.491692	4.938613	-4.93861
79	GCAGAAC	M	4464	46XX	46XY+18	40	82	31	0.082749	1022.638	974.749	0.511988	0.488012	7.344068	-7.34407
80	ATGGAAT	F	2100	46XX	46XX+18	42	98	39	0.094043	1023.162	977.342	0.511452	0.488548	6.993832	-6.99383
81	AICAGAT	M	4176	46XX	46XY+21	35	147	33	0.148476	957.3759	1038.729	0.479622	0.520378	-13.8102	13.81016
82	AGTTACT	M	4500	46XX	46XY+21	39	124	19	0.205496	952.9748	1049.25	0.475958	0.524042	-16.205	16.20502
83	AGTAGAC	M	8640	46XX	46XY+21	37	125	30	0.130907	973.3572	1028.835	0.486146	0.513854	-9.54628	9.546281
84	AGGTCTT	M	16092	46XX	46XY+21	43	94	35	0.126545	976.6225	1029.797	0.486749	0.513251	-9.1521	9.152097
85	AGGATAT	M	4212	46XX	46XY+21	40	157	24	0.119821	969.1251	1025.932	0.485763	0.514237	-9.79643	9.796425
86	AGGACGG	M	4644	46XX	46XY+21	33	97	31	0.053737	992.2248	1011.59	0.495168	0.504832	-3.64041	3.640405
87	AGAGATT	M	4320	46XX	46XY+21	31	119	37	0.08022	978.7159	1014.132	0.491114	0.508886	-6.29894	6.298937

FIG. 8 (cont.)

89	AGCTGCC	F	6768	46XX	46XX+21	31	98	35	0.126696	963.2256	1024.818	0.484509	0.515491	-10.616	10.61595
	<b>Muestra</b>	<b>Etiqueta</b>	<b>Género</b>	<b>Equiv.</b>	<b>Cariotipo</b>	<b>Edad</b>	<b>Edad de gest.</b>	<b>% fetal</b>	<b>Media 18</b>	<b>Media 21</b>	<b>Prop18</b>	<b>Prop21</b>	<b>ZStat18</b>	<b>ZStat21</b>	
	<b>de muestra</b>	<b>fetal</b>	<b>de genoma</b>	<b>materno</b>	<b>fetal</b>	<b>de la</b>	<b>(día)</b>	<b>informativos</b>							
						<b>madre</b>									
90	AGCGCTG	M	1375	46XX	46XY+21	36	162	36	0.133202	961.1764	1032.034	0.482225	0.517775	-12.1088	12.10879
91	AGCATGC	M	5155	46XX	46XY+21	41	154	37	0.095349	967.2823	1017.613	0.487322	0.512678	-8.77777	8.777767
92	AATGGTT	F	2549	46XX	46XX+21	41	119	32	0.09505	965.3134	1025.717	0.484831	0.515169	-10.4056	10.4056
93	AAGTAAG	F	9216	46XX	46XX+21	42	253	32	0.209915	946.0902	1052.948	0.473273	0.526727	-17.96	17.96002
94	AAATAGC	F	2041	46XX	46XX+21	36	161	32	0.102554	973.7957	1025.545	0.487058	0.512942	-8.94982	8.949816
95	AAATCCT	M	1843	46XX	46XY+21	42	91	35	0.137238	965.4467	1031.552	0.483449	0.516551	-11.309	11.30904

FIG. 8 (cont.)

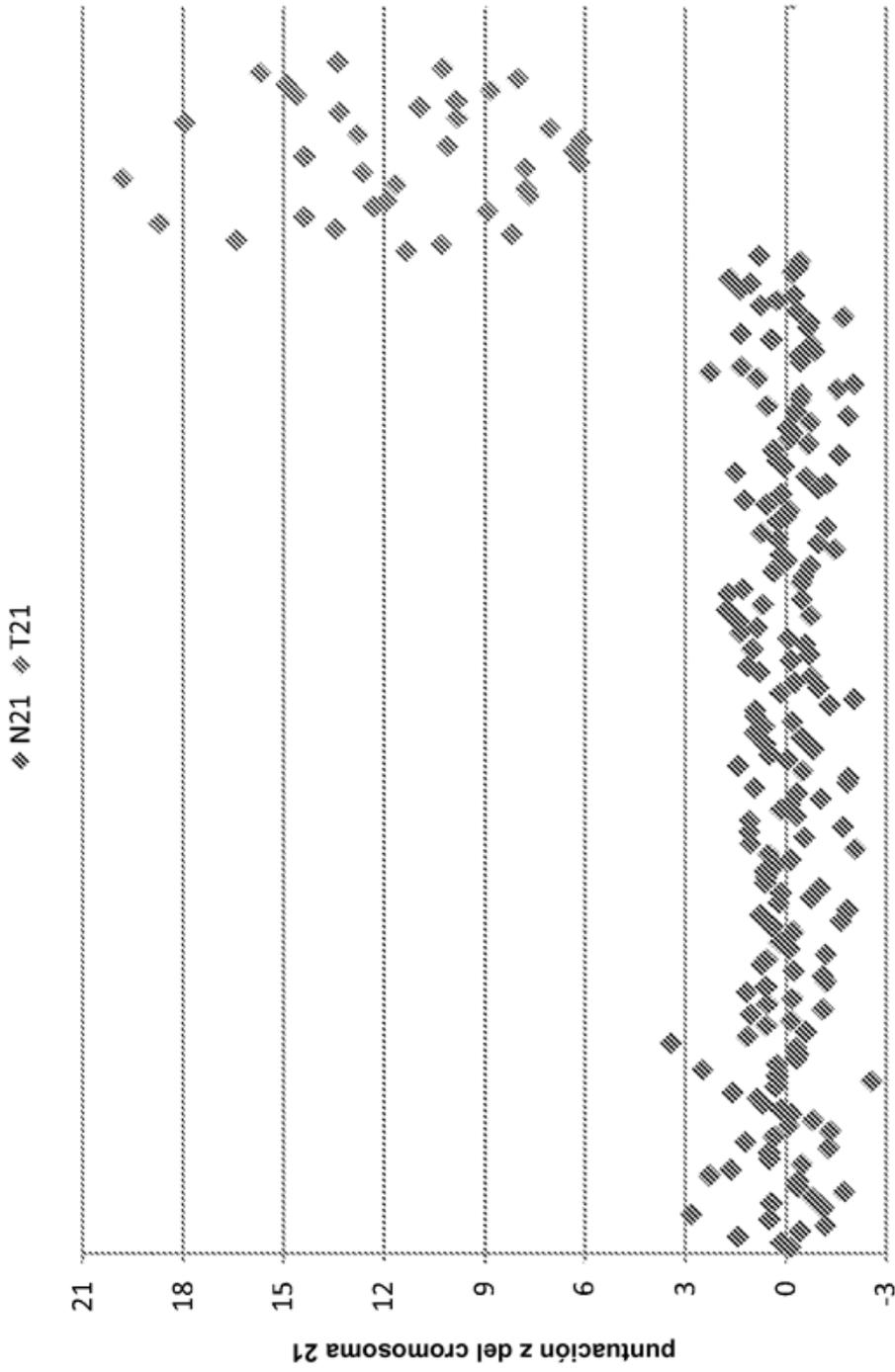


FIG. 9

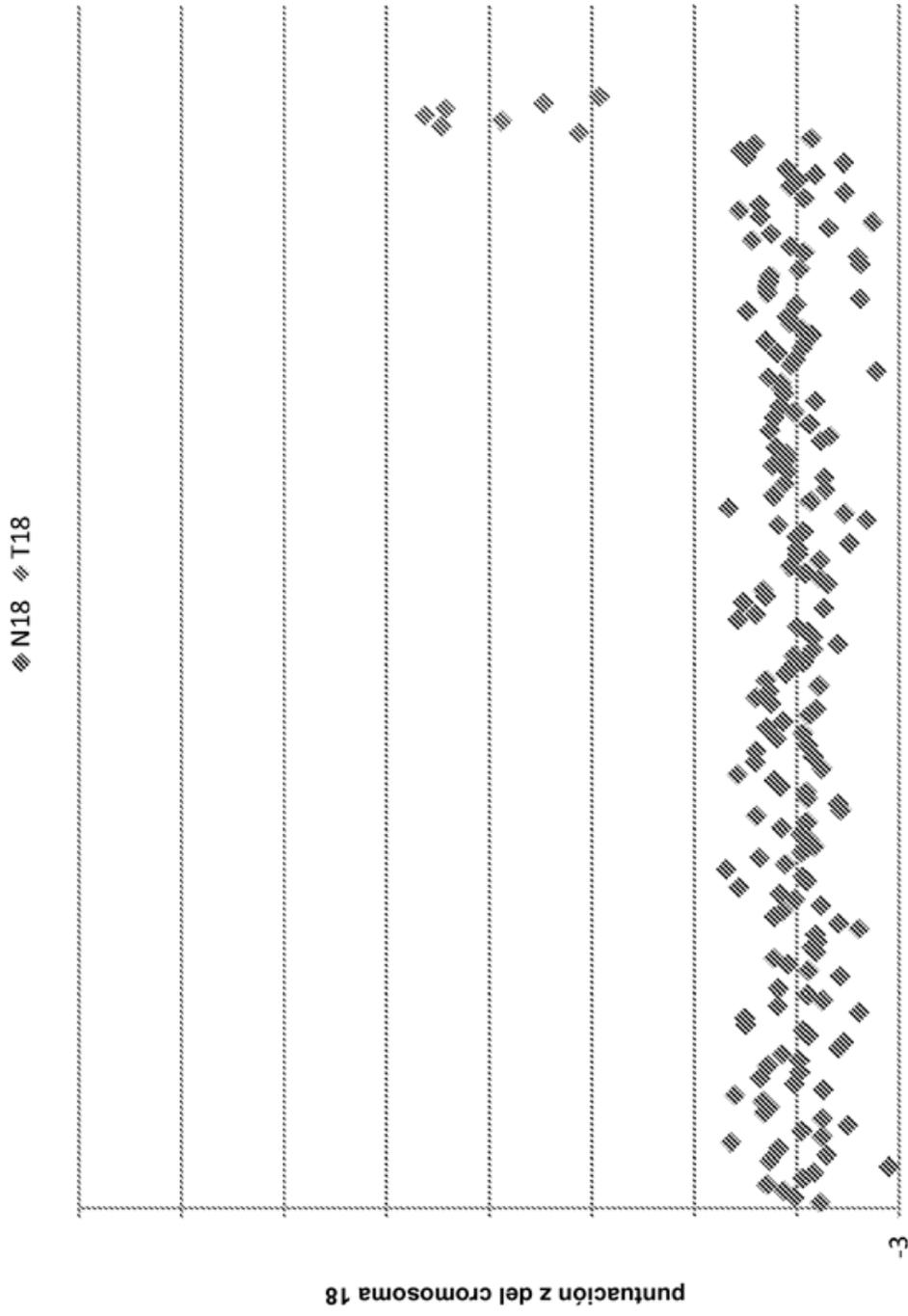


FIG. 10

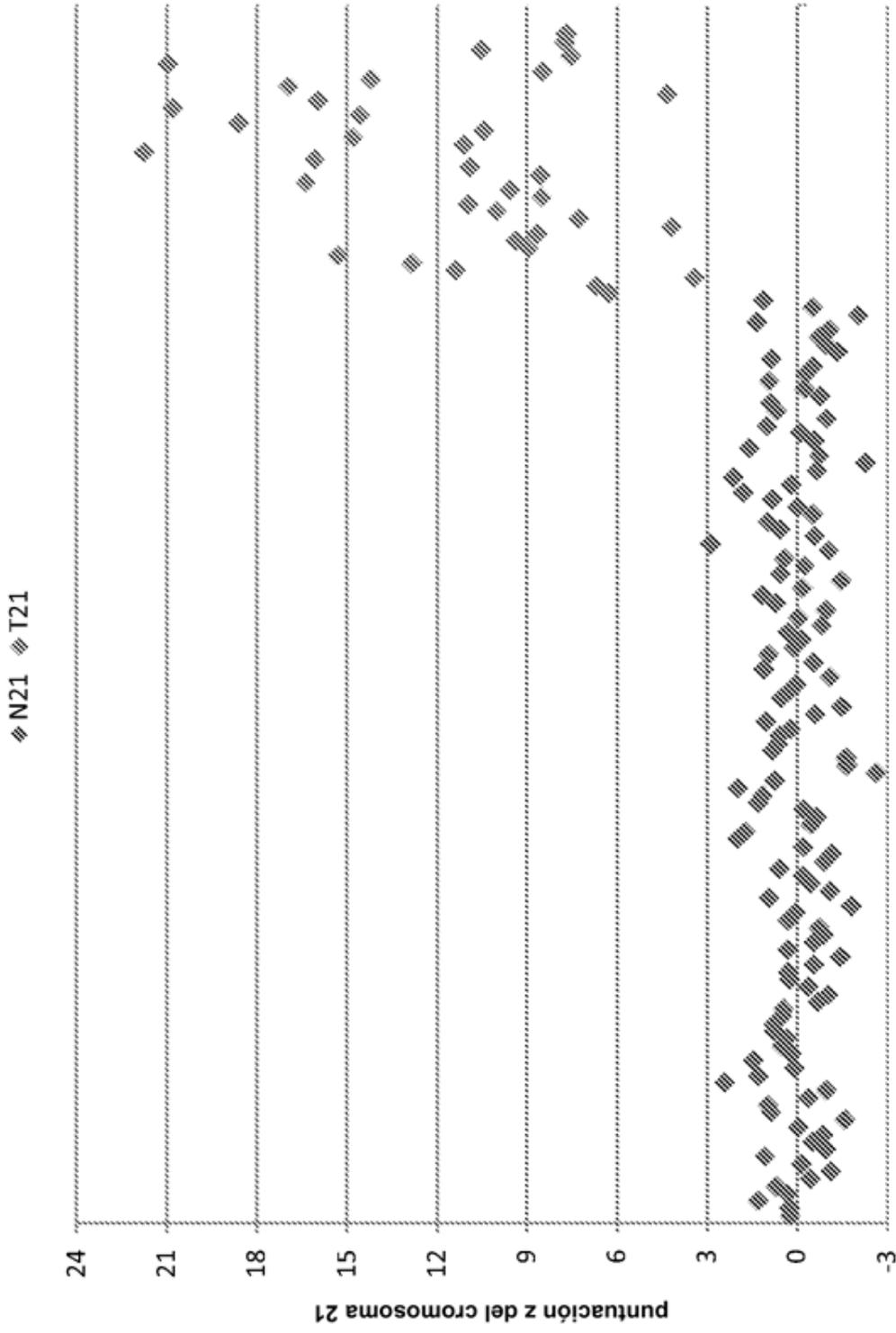


FIG. 11

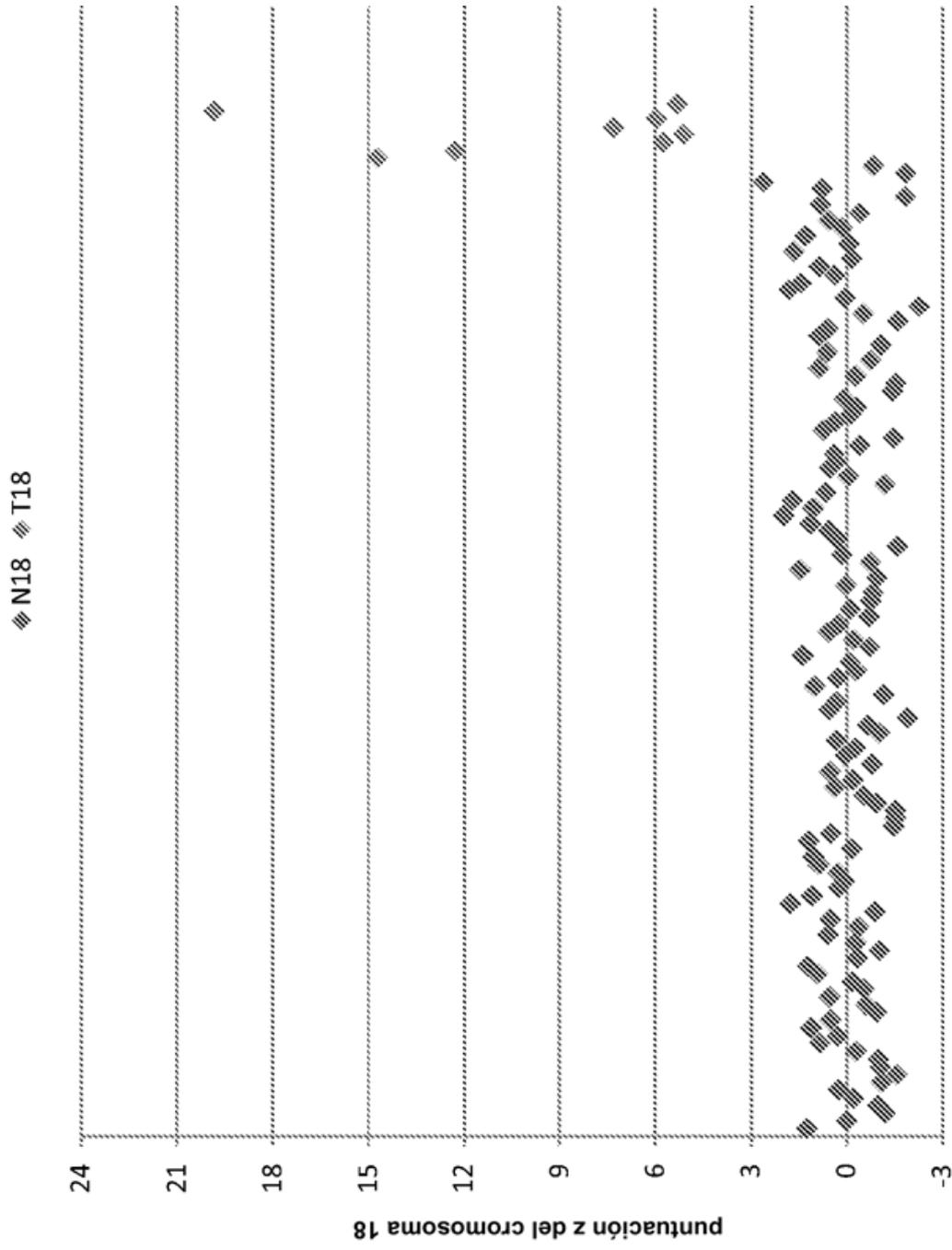


FIG. 12