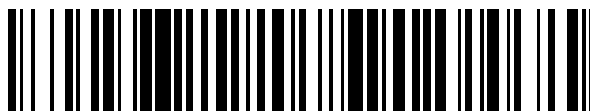


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 718 111**

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/US2011/046963**

87 Fecha y número de publicación internacional: **09.02.2012 WO12019193**

96 Fecha de presentación y número de la solicitud europea: **08.08.2011 E 11745881 (0)**

97 Fecha y número de publicación de la concesión europea: **02.01.2019 EP 2601310**

54 Título: **Sistemas de ensayo para análisis genético**

30 Prioridad:

06.08.2010 US 371605 P
25.01.2011 US 201113013732

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.06.2019

73 Titular/es:

ARIOSA DIAGNOSTICS, INC. (100.0%)
5945 Optical Court
San Jose, CA 95138, US

72 Inventor/es:

SPARKS, ANDREW;
OLIPHANT, ARNOLD;
ZAHN, JACOB;
SONG, KEN y
STUELPNAGEL, JOHN

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 718 111 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 análisis genético

5 Campo de la invención

Esta invención se refiere a la detección y cuantificación de variación genética en muestras mixtas en un ensayo único.

10 Antecedentes de la invención

En la siguiente discusión se describirán determinados artículos y métodos con propósitos introductorios y de antecedentes. Nada contenido en el presente documento se interpretará como una "aceptación" de la técnica anterior. El solicitante se reserva de manera expresa el derecho de demostrar, cuando sea apropiado, que los artículos y métodos referidos en el presente documento no forman parte de la técnica anterior en las disposiciones legales aplicables.

Las alteraciones genéticas justifican un amplio número de patologías, incluyendo las patologías producidas por aneuploidía cromosómica (por ejemplo, síndrome de Down), mutaciones de la línea germinal en genes específicos (por ejemplo, anemia drepanocítica) y patologías producidas por mutaciones somáticas (por ejemplo, cáncer). Los métodos de diagnóstico para determinar tales anomalías genéticas se han convertido en técnicas habituales para identificar enfermedades y trastornos específicos, así como para proporcionar información valiosa sobre la fuente de la enfermedad y opciones de tratamiento.

Aunque la tecnología convencional proporciona métodos de detección para estas diferentes alteraciones genéticas, en la actualidad se necesitan diferentes técnicas para interrogar diferentes clases de mutaciones. Un ensayo que proporcione la detección de una variación en el número de copias con detección simultánea de polimorfismos en genes individuales puede ser una potente herramienta en el tratamiento. La presente invención aborda esta necesidad.

Existe por tanto una necesidad de métodos no invasivos de detección para alteraciones genéticas, incluyendo variaciones en el número de copias, en muestras mixtas que comprenden ADN normal y posiblemente anómalo. La presente invención aborda esta necesidad.

35 Sumario de la invención

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

La presente invención proporciona un método de ensayo único para determinar la variación genética en muestras mixtas que comprenden material genómico (por ejemplo, ADN libre de células) a partir de una fuente principal y una o más fuentes secundarias de un sujeto. Más particularmente, la presente invención proporciona un método de ensayo único para la detección de variación en el número de copias de uno o más loci y la detección de polimorfismos en uno o más loci en una muestra que comprende ácidos nucleicos a partir de una fuente principal y una fuente secundaria dentro de un individuo. La presente invención utiliza un sistema de ensayo único que permite la identificación tanto de variaciones genéticas en el número de copias (VNC) como polimorfismos de un sólo gen, y la capacidad de distinguir entre estas VNC y polimorfismos en las fuentes principales y secundarias.

En un aspecto, el método de ensayo único utiliza amplificación y detección de loci seleccionados en una muestra mixta a partir de un individuo para identificar el número de copias y la presencia de polimorfismos en loci seleccionados a partir de una principal y una o más fuentes secundarias.

En un aspecto específico, la invención proporciona un método de ensayo único con la capacidad para determinar 1) la presencia o ausencia de VNC en uno o más loci y; 2) la presencia o ausencia de polimorfismos en uno o más loci en una muestra mixta a partir de un individuo. El método de ensayo único puede detectar específicamente el número de copias de los loci seleccionados y de polimorfismos presentes en loci seleccionados a partir de una fuente secundaria dentro de la muestra mixta, y para distinguir esta forma del número de copias y los polimorfismos presentes en loci seleccionados a partir de una fuente principal en la muestra mixta. Preferiblemente, los loci se analizan a través del uso de ácidos nucleicos libres de células, y preferiblemente los ácidos nucleicos libres de células analizados en el sistema de ensayo es ADN libre de células (ADNlc).

Por tanto, en un aspecto específico, la invención proporciona un método de ensayo único para la detección de la

presencia o ausencia de una variación en el número de copias (VNC) de una región genómica y la presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, 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 uno o más loci en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los loci que corresponden a una región genómica y/o los loci con un posible polimorfismo; amplificar los productos de ligamiento contiguos 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 una o más regiones genómicas y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

En determinados aspectos, los oligonucleótidos de secuencia fija se hibridan con regiones adyacentes en los loci, y los oligonucleótidos de secuencia fija se ligan directamente durante la etapa de ligamiento del ensayo. En otros aspectos, los oligonucleótidos de secuencia fija se hibridan con regiones no adyacentes en los loci que se separan por uno, unos pocos o varios nucleótidos, y la región entre los oligonucleótidos de secuencia fija se usa como un molde para la extensión del cebador para cerrar el hueco entre los oligonucleótidos de secuencia fija antes de la etapa de ligamiento.

Por tanto, en un segundo aspecto, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de VNC de una región genómica y presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija a una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija a la muestra mixta en condiciones que permitan que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; extender la región entre el primer y el segundo oligonucleótido hibridado de al menos un conjunto de oligonucleótidos de secuencia fija con una polimerasa y dNTP para crear dos oligonucleótidos de secuencia fija hibridados de manera adyacente a partir de este conjunto; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los loci que corresponden a una región genómica y/o los loci con un posible polimorfismo; amplificar los productos de ligamiento contiguos usando las regiones de cebadores universales 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 una o más regiones genómicas y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

En un aspecto preferido, los oligonucleótidos de secuencia fija usados para interrogar los loci específicos comprenden regiones de cebadores universales que son comunes sustancialmente a todos los oligonucleótidos de secuencia fija usados en un ensayo particular. Este permite que sustancialmente todos los productos de ligamiento producidos en un ensayo único se amplifiquen, se aislen y/o se analicen usando mecanismos similares (por ejemplo, hibridación o determinación de secuencia). El uso de tales regiones de cebadores universales evita la necesidad de restos detectables distinguibles de manera individual asociados con loci o alelos particulares, y proporciona mecanismos más eficaces y rentables para interrogación y/o análisis de multiplexado de múltiples loci de una o múltiples muestras. En un aspecto preferido, las regiones de cebadores universales se usan en la determinación de secuencia de los productos de amplificación. En otro aspecto preferido, las mismas regiones de cebadores universales se usan 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.

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

La región genómica puede ser un locus único, o puede ser una región más grande, hasta e incluyendo un cromosoma. Para la determinación de regiones genómicas más grandes, pueden usarse conjuntos de loci para determinar el tamaño y en determinados aspectos los límites de tales regiones genómicas para las que se han detectado VNC.

En determinados aspectos, los productos de amplificación del ensayo se aíslan antes de la detección. Preferiblemente, los productos de amplificación se aíslan como moléculas individuales antes de la detección. Estos productos de amplificación aislados pueden amplificarse además opcionalmente para crear copias idénticas de todos o una parte de los productos de amplificación individuales antes de la detección, o amplificarse además para

crear copias idénticas de moléculas complementarias a todos o una parte de los productos de amplificación individuales antes de la detección.

5 Los ensayos multiplexados de la invención permiten el análisis de más de 70, preferiblemente 80 o más, preferiblemente 90 o más, y más preferiblemente 96 o más loci seleccionados simultáneamente. Estos loci seleccionados pueden ser diferentes loci de una muestra única, o pueden ser loci de dos o más individuos. En el último caso, al menos uno de los dos oligonucleótidos de secuencia fija usados para el análisis de un locus seleccionado puede comprender un identificador de muestra (por ejemplo, un "índice de muestra") que permitirá al locus asociarse con una muestra particular. Alternativamente, puede añadirse un índice de muestra durante la
10 amplificación del producto de ligamiento usando un cebador que comprende el índice de muestra.

Preferiblemente, al menos un locus interrogado para VNC en una muestra mixta es diferente de todos los loci interrogados para polimorfismos en la muestra mixta. En aspectos específicos, varios loci interrogados para VNC en una muestra mixta son diferentes de los loci interrogados para polimorfismos en la muestra mixta. En aspectos más
15 específicos, la mayoría de loci interrogados para VNC en una muestra mixta son diferentes de los loci interrogados para polimorfismos en la muestra mixta.

En algunos aspectos de la invención, los oligonucleótidos de secuencia fija se hibridan con regiones adyacentes en un locus, y el ligamiento de estos oligonucleótidos da como resultado un producto de ligamiento que une los dos oligonucleótidos de secuencia fija. En otros aspectos, los oligonucleótidos de secuencia fija se hibridan con regiones no adyacentes en un locus, y las regiones entre los oligonucleótidos de secuencia fija se extienden usando una polimerasa y dNTP. En aspectos preferidos, los oligonucleótidos de secuencia fija se hibridan con regiones no adyacentes en un locus, y uno o más oligonucleótidos de puente se hibridan con la región entre y adyacente al conjunto de oligonucleótidos de secuencia fija. En determinados aspectos preferidos, los oligonucleótidos de puente
20 usados pueden proporcionar información del contenido en polimorfismos en los loci así como información sobre la frecuencia de loci para el análisis de la VNC.
25

En otros aspectos preferidos, la interrogación de estos loci utiliza técnicas de amplificación universales que permiten la amplificación de múltiples loci en una reacción de amplificación única. Los ácidos nucleicos seleccionados para la detección tanto de la VNC como de los polimorfismos que usa el sistema de ensayo de la invención pueden amplificarse usando métodos de amplificación universales después de la amplificación selectiva inicial de la muestra mixta. El uso de amplificación universal permite que múltiples regiones de ácidos nucleicos de una o múltiples muestras se amplifiquen usando un número único o limitado de cebadores de amplificación, y es especialmente útil para amplificar múltiples regiones seleccionadas en una reacción única.
30
35

Por tanto, en un aspecto preferido de la invención, se introducen secuencias complementarias a los cebadores para su uso en amplificación universal en los loci seleccionados durante o después de la 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.
40

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 preferiblemente de manera conjunta con la detección de la VNC o alteración cromosómica y la detección de polimorfismo de manera que la VNC y el polimorfismo pueden asignarse de manera apropiada a la muestra de origen.
45

Las frecuencias de loci seleccionados pueden determinarse para una región genómica de interés y compararse con las frecuencias de loci 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 las frecuencias de loci en la muestra mixta.
50

En los sistemas 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 asistir en la detección posterior. Después del aislamiento, los productos de amplificación pueden amplificarse además para crear copias idénticas de todos o una parte de los productos de amplificación individuales antes de la detección. Alternativamente, los productos de amplificación aislados pueden amplificarse además para crear copias idénticas de moléculas complementarias a todos o una parte de los productos de amplificación individuales antes de la detección.
55
60

Pueden emplearse diversos métodos de detección de VNC junto con la detección de los polimorfismos en los sistemas de ensayo de la divulgación.

En un aspecto general, el sistema de ensayo emplea un método para la determinación de una VNC en uno o más loci 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
65

de un segundo locus de interés en la muestra mixta, determinar la frecuencia relativa de los loci seleccionados, comparar la frecuencia relativa de los loci seleccionados e identificar la presencia o ausencia de una VNC basándose en las frecuencias relativas comparadas de los ácidos nucleicos seleccionados del primer y el segundo loci. Preferiblemente, el método de ensayo amplifica dos o más loci seleccionados de diferentes regiones genómicas, aunque los loci pueden ubicarse en la misma región genómica general para confirmación de las VNC que surgen de alteraciones cromosómicas en lugar de las VNC de un locus único.

En otros aspectos, los oligos de puente son oligos degenerados. En otros aspectos, los oligos de puente se proporcionan como reservas de oligos con secuencias aleatorias que comprenden sustancialmente cada combinación para el tamaño particular del oligo de puente usado con los oligonucleótidos de secuencia fija.

Aunque los aspectos de la invención que usan oligonucleótidos de puente se describen principalmente usando un oligo de puente único, se prevé que pueden usarse múltiples oligos de puente que se hibridan con regiones complementarias, adyacentes entre oligonucleótidos de secuencia fija en los métodos descritos.

En un aspecto preferido, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de variación en el número de copias (VNC) de una región genómica y la presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales a la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones no adyacentes, complementarias en uno o más loci con un posible polimorfismo; introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de puente se hibriden específicamente con regiones en los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija de los conjuntos; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los loci que corresponden a una región genómica y/o los loci con un posible polimorfismo; amplificar los productos de ligamiento contiguos usando las regiones de cebadores universales para crear productos de amplificación y detectar los productos de amplificación. La detección de los productos de amplificación correlaciona el número de copias de una o más regiones genómicas y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

En otro aspecto específico, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de VNC de una región genómica y la presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija a una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci en o asociados con una región genómica; introducir un segundo conjunto de oligonucleótidos de secuencia fija a la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; introducir uno o más oligonucleótidos de puente en condiciones que permite que los oligonucleótidos de puente se hibriden específicamente con regiones en los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija del conjunto; extender la región entre al menos un oligonucleótido de secuencia fija y un oligonucleótido de puente con una polimerasa y dNTP para crear oligonucleótidos de secuencia fija hibridados y oligonucleótidos de puente de manera adyacente; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los loci que corresponden a una región genómica y/o los loci con un posible polimorfismo; amplificar los productos de ligamiento contiguos usando las regiones de cebadores universales 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 una o más regiones genómicas y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

En aspectos preferidos de la invención que usan oligonucleótidos de puente, el primer y el segundo conjunto de oligonucleótidos de secuencia fija se introducen antes de la introducción de los oligonucleótidos de puente. Más preferiblemente, los oligonucleótidos de secuencia fija no hibridados se retiran antes de la introducción de los oligonucleótidos de puente. En algunos aspectos, los oligonucleótidos de puente se introducen simultáneamente con la mezcla de ligamiento. En otros aspectos, los productos de hibridación de los oligonucleótidos de secuencia fija y el locus se aíslan antes de la introducción de los oligonucleótidos de puente.

En un aspecto preferido, el sistema de ensayo proporciona interrogación de loci altamente multiplexada usando uno o más oligonucleótidos de puente comunes que son complementarios a regiones en dos o más loci interrogados. Por tanto, el número de oligonucleótidos de puente usados en el sistema de ensayo multiplexado será menor que el número de loci interrogados en el ensayo. En determinados aspectos específicos, el sistema de ensayo usa una reserva de oligonucleótidos de puente que se diseñan cada uno para ser compatibles con dos o más loci interrogados usando el sistema de ensayo de la invención. En estos aspectos, los oligonucleótidos de puente usados en el ensayo multiplexado se diseñan preferiblemente para tener una T_m en un intervalo de $\pm 5^\circ\text{C}$, más

preferiblemente en un intervalo de $\pm 2^{\circ}\text{C}$.

5 En determinados aspectos, el sistema de ensayo multiplexa la interrogación de loci usando uno o más oligonucleótidos de puente comunes que son complementarios a regiones en dos o más loci interrogados. Por tanto, el número de oligonucleótidos de puente usados en el sistema de ensayo multiplexado será menor que el número de loci interrogados en el ensayo. En determinados aspectos específicos, el sistema de ensayo usa una reserva de oligonucleótidos de puente que se diseñan cada uno para ser compatibles con dos o más loci interrogados usando el sistema de ensayo de la invención.

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

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

20 En tales aspectos, la invención proporciona sistemas de ensayo que comprenden un sistema de ensayo único con la capacidad para detectar dentro de una muestra mixta de un individuo 1) la presencia o ausencia de una alteración cromosómica asociada con una o más VNC y 2) la presencia o ausencia de uno o más polimorfismos en uno o más locus seleccionados. La presencia o ausencia de la alteración cromosómica se detecta preferiblemente a través del uso de loci informativos que permiten que el sistema de ensayo distinga entre ácidos nucleicos de una fuente principal y de una fuente secundaria.

25 Por tanto, en un aspecto específico, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; introducir un segundo conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más loci de un primer cromosoma; introducir un tercer conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más loci de un segundo cromosoma; introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de puente se hibriden específicamente con regiones en los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija del conjunto; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligamiento contiguos para crear productos de amplificación y detectar los productos de amplificación. La detección de los productos de amplificación se correlaciona con la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

45 En otro aspecto específico, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; introducir un segundo conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más loci de un primer cromosoma; introducir un tercer conjunto de oligonucleótidos de secuencia fija que comprende regiones de cebadores universales en la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en dos o más loci de un segundo cromosoma; introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de puente se hibriden específicamente con regiones en los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija del conjunto; extender la región entre el primer y el segundo oligonucleótido hibridado de al menos un conjunto de oligonucleótidos de secuencia fija con una polimerasa y dNTP para crear dos oligonucleótidos de secuencia fija hibridados de manera adyacente a partir de este conjunto; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligamiento contiguos para crear productos de amplificación y detectar los productos de amplificación. La detección de los productos de amplificación correlaciona la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.

65 En algunos casos, la alteración cromosómica se asocia con duplicación de genes o expansión de loci en un

cromosoma de interés. En otros casos, la alteración cromosómica se asocia con una translocación que da como resultado la presencia de un parte extra de un cromosoma en el genoma. En aún otros casos, la alteración cromosómica se asocia con aneuploidía de un cromosoma de interés.

5 Por tanto, en un aspecto, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de polimorfismos en uno o más loci usando un ensayo único, 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 ácidos nucleicos que corresponden a loci informativos en dos o más cromosomas; introducir un segundo 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 ácidos nucleicos indicativos de la presencia o ausencia de polimorfismos en uno o más loci; introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en los ácidos nucleicos, en el que uno o más oligonucleótidos de puente son complementarios a una región de los ácidos nucleicos entre e inmediatamente adyacente a las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligamiento contiguos para crear productos de amplificación y detectar los productos de amplificación. La detección del producto de amplificación se correlaciona a la detección de los loci en la muestra mixta y puede usarse para determinar la cantidad de loci asociados con la presencia o ausencia de la alteración cromosómica y la situación genética de uno o más loci en la muestra mixta. Los niveles detectados de estos ácidos nucleicos pueden usarse por tanto para determinar la presencia o ausencia de la alteración cromosómica y la presencia o ausencia de un polimorfismo en uno o más loci a partir de una fuente principal y/o una fuente secundaria dentro de una muestra mixta.

25 Por tanto, en un aspecto, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de una alteración cromosómica y la presencia o ausencia de polimorfismos en uno o más loci usando un ensayo único, 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 ácidos nucleicos que corresponden a loci informativos en dos o más cromosomas; introducir un segundo 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 ácidos nucleicos indicativos de la presencia o ausencia de polimorfismos en uno o más loci; introducir uno o más oligonucleótidos de puente en condiciones que permiten que oligonucleótidos de secuencia de puente se hibriden específicamente con regiones complementarias en los ácidos nucleicos, en el que uno o más oligonucleótidos de puente son complementarios a una región de los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto e inmediatamente adyacente al primero de los oligonucleótidos de secuencia fija; extender la región entre el oligonucleótido de puente y el segundo oligonucleótido de secuencia fija usando polimerasa y dNTP; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligamiento contiguos para crear productos de amplificación y detectar los productos de amplificación. La detección del producto de amplificación se correlaciona a la detección de los loci en la muestra mixta, y puede usarse para determinar la cantidad de loci asociados con la presencia o ausencia de la alteración cromosómica y la situación genética de uno o más loci en la muestra mixta. Los niveles detectados de estos ácidos nucleicos pueden usarse por tanto para determinar la presencia o ausencia de la alteración cromosómica y la presencia o ausencia de un polimorfismo en uno o más loci a partir de una fuente principal y/o una fuente secundaria dentro de una muestra mixta.

50 En otro aspecto, la presente invención utiliza técnicas que permiten la identificación tanto de VNC como de agentes infecciosos en una muestra mixta. Esto puede ser especialmente útil para monitorizar pacientes en los que el desenlace clínico puede comprometerse por la presencia de un agente infeccioso. Por ejemplo, un paciente que se ha sometido a un trasplante tomará probablemente medicamentos inmunodepresores, y por lo tanto será más propenso a la 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 adverso en la madre y/o el feto. Además, determinados tipos de cáncer se asocian con agentes infecciosos (por ejemplo, cáncer de hígado asociado con las infecciones por hepatitis B y C, cáncer de cuello uterino asociado con la infección por el virus del papiloma humano), y la identificación de los agentes infecciosos puede ser informativa para predecir el desenlace clínico o determinar el ciclo preferido de tratamiento médico para el paciente.

60 Por tanto, en determinados aspectos, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de variación genética en el número de copias (VNC) de una región genómica y la presencia o ausencia de un agente infeccioso en una muestra mixta usando un ensayo único, 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 uno o más loci 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 de secuencia fija se hibriden específicamente con regiones complementarias en ácidos nucleicos indicativos de un agente infeccioso;

ligar los oligonucleótidos hibridados para crear un producto de ligamiento contiguo complementario a los ácidos nucleicos; amplificar el producto de ligamiento contiguo para crear productos de amplificación y detectar los productos de amplificación. La detección de los productos de amplificación se correlaciona con un número de copias de la región genómica y la presencia o ausencia de un agente infeccioso en la muestra mixta.

5 En otros determinados aspectos, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de variación en el número de copias (VNC) de una región genómica, la presencia o ausencia de uno o más polimorfismos, y la presencia o ausencia de un agente infeccioso en una muestra mixta de un individuo usando un ensayo único, comprendiendo el ensayo las etapas de: introducir un primer conjunto de oligonucleótidos de
10 secuencia fija en una muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci 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 de secuencia fija se hibriden específicamente con regiones complementarias en uno o más loci con un posible polimorfismo; introducir un tercer conjunto de oligonucleótidos de secuencia fija en la
15 muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en ácidos nucleicos indicativos de un agente infeccioso; ligar los oligonucleótidos hibridados para crear un producto de ligamiento contiguo complementario a los ácidos nucleicos; amplificar el producto de ligamiento contiguo para crear productos de amplificación y detectar los productos de amplificación. La detección de los productos de amplificación se correlaciona con un número de copias de la región genómica, la
20 presencia o ausencia de un polimorfismo en uno o más loci y la presencia o ausencia de un agente infeccioso en la muestra mixta.

En estos aspectos, los ensayos también pueden comprender introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de puente se hibriden específicamente con regiones
25 complementarias en los ácidos nucleicos, en los que el uno o más oligonucleótidos de puente son complementarios a una región de los ácidos nucleicos entre las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto.

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

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 loci seleccionados en el ADNlc que corresponde a un primer cromosoma de interés en una muestra mixta; amplificar dos o más loci seleccionados en el ADNlc que corresponde a un segundo cromosoma de interés en la muestra mixta, determinar la frecuencia relativa de las regiones seleccionadas del primer y el segundo cromosomas de interés, comparar la frecuencia relativa de las regiones seleccionadas del primer y el segundo cromosomas de interés e identificar la presencia o ausencia de una aneuploidía basándose en las frecuencias relativas comparadas de las regiones seleccionadas. En un aspecto específico, los loci del primer y el segundo cromosomas se amplifican en una reacción única, y preferiblemente en una reacción única contenida dentro de un recipiente único.

Preferiblemente, el sistema de ensayo detecta la presencia o ausencia de loci en muestras que pueden obtenerse fácilmente de un sujeto, o de la sangre de un sujeto, tal como plasma, suero y similares. En un aspecto general, el sistema de ensayo utiliza la detección de regiones seleccionadas en el ADNlc en una muestra mixta. En un aspecto más específico, el sistema de ensayo utiliza la detección de regiones seleccionadas en el 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 loci. El número de copias dentro de una región genómica puede determinarse basándose en la detección de cantidades de loci seleccionados y en la comparación con las cantidades de loci seleccionados de otra región genómica y/o con las cantidades de loci 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 significativa estadísticamente de sujetos "normales" genéticamente, es decir sujetos que no tienen una VNC asociada con el loci particular interrogado en el sistema de ensayo.

En un aspecto preferido de la invención, los productos de amplificación que corresponden a los ácidos nucleicos seleccionados se aíslan como moléculas individuales para el análisis de los loci seleccionados. Estos productos de amplificación individuales se aíslan unos de otros, y preferiblemente se aíslan de manera física (por ejemplo, en un sustrato o en recipientes individuales). Las moléculas individuales pueden amplificarse además después del aislamiento para realizar copias idénticas, múltiples del producto de amplificación, una parte del mismo o un ácido

nucleico complementario al producto de amplificación o una parte del mismo. La detección de las moléculas individuales o el producto de amplificación puede hacerse a través de secuenciación.

5 En algunos aspectos de la divulgación, el producto de ligamiento no se amplifica, sino que se detecta directamente después de la hibridación, por ejemplo, usando técnicas de secuenciación de molécula única.

10 Por tanto, en un aspecto específico, la invención proporciona un sistema de ensayo para la detección de la presencia o ausencia de una alteración cromosómica y polimorfismos en uno o más loci en una muestra mixta usando un ensayo único, 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 ácidos nucleicos que corresponden a un cromosoma particular; introducir un segundo 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 ácidos nucleicos con un posible polimorfismo; introducir uno o más oligonucleótidos de puente en condiciones que permiten que los oligonucleótidos de puente se hibriden específicamente con regiones complementarias en los ácidos nucleicos, en el que uno o más oligonucleótidos de puente son complementarios a una región de los ácidos nucleicos entre e inmediatamente adyacente a las regiones complementarias a los oligonucleótidos de secuencia fija de cada conjunto; ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los ácidos nucleicos; amplificar los productos de ligamiento contiguos para crear productos de amplificación que tienen la secuencia de los loci; aislar productos de amplificación individuales y analizar los productos de amplificación individuales para determinar la secuencia de todos o una parte de los productos de amplificación individuales. El análisis de los productos de amplificación individuales se correlaciona con la presencia o ausencia de una alteración cromosómica y la situación genética de uno o más loci en la muestra mixta. Por ejemplo, los niveles de los ácidos nucleicos que corresponden a un cromosoma particular pueden compararse usando también ácidos nucleicos que corresponden a otro cromosoma, o pueden compararse con niveles de referencia para el cromosoma que se va a interrogar.

20 En un aspecto preferido, los productos de amplificación individuales se analizan a través de la determinación de secuencia. En otros aspectos de la divulgación los productos de amplificación individuales se analizan usando técnicas de hibridación.

30 Una característica de la presente invención es que el número de copias de los loci seleccionados puede detectarse usando métodos de detección no polimórficos, es decir, métodos de detección que no dependen de la presencia o ausencia de un polimorfismo particular para identificar la región de ácidos nucleicos seleccionada. En un aspecto preferido, los sistemas de ensayo de detección utilizan métodos de detección no polimórficos para "contar" los números relativos de loci seleccionados presentes en una muestra mixta. Estos números pueden utilizarse 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 utilizarse para determinar, si estadísticamente, los 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 un trastorno, para determinar el pronóstico de una enfermedad o un trastorno, para asistir en la determinación de las posibles opciones de tratamiento, etc.

40 En algunos aspectos, los métodos para la determinación de la variación en el número de copias de múltiples loci seleccionados de dos o más cromosomas en una muestra. Los niveles de los diferentes loci seleccionados que corresponden a cromosomas específicos pueden cuantificarse de manera individual y compararse para determinar la presencia o ausencia de una aneuploidía cromosómica en una o más fuentes de células en una muestra mixta. Las regiones cuantificadas de manera individual pueden recibir 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.

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

55 En aún otros aspectos, las frecuencias relativas de los loci 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.

60 En un aspecto particular, los loci seleccionados se aíslan antes de la detección. Los loci seleccionados pueden aislarse a partir de la muestra mixta usando cualquier medio que aisle de manera selectiva los ácidos nucleicos particulares presentes en la muestra mixta para el análisis, por ejemplo, hibridación, amplificación u otra forma de aislamiento basado en secuencia de los ácidos nucleicos a partir de la muestra mixta. Después del aislamiento, los ácidos nucleicos seleccionados se distribuyen de manera individual en un formato de detección adecuado, por ejemplo, en un micromatriz o en una célula de flujo, para la determinación de la secuencia y/o las 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.

5 Después del aislamiento y la distribución de los ácidos nucleicos seleccionados en un formato adecuado, las secuencias seleccionadas se identifican, por ejemplo, a través de la determinación de secuencia de la secuencia seleccionada.

10 En un aspecto específico, la invención proporciona un sistema 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 loci seleccionados de un primer y un segundo cromosoma de interés en la muestra mixta, amplificar dos o más loci seleccionados del primer y el 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 loci seleccionados del primer y el segundo cromosomas de interés e identificar la presencia o ausencia de una aneuploidía fetal basándose en las frecuencias relativas comparadas de los loci seleccionados.

20 En algunos aspectos específicos, las frecuencias relativas de los loci de una región genómica se calculan de manera individual, y las frecuencias relativas de los loci individuales se comparan para determinar la presencia o ausencia de una alteración cromosómica. En otros aspectos específicos, las frecuencias relativas de los loci seleccionados se usan para determinar una frecuencia del cromosoma de un primer y un 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 del primer y el segundo cromosomas de interés.

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

35 Aunque el sistema de ensayo se usa preferiblemente para detectar el ADNc en una muestra mixta, en determinados aspectos el ADN de interés que va a analizarse usando el sistema de ensayo de la divulgación comprende ADN directamente de los diferentes tipos de células en lugar de una muestra mixta que contiene ADN de los tipos de células principales y secundarios. Tales muestras pueden obtenerse de diversas fuentes dependiendo del ADN objetivo. Por ejemplo, pueden derivarse células fetales para el análisis de muestras tales como líquido amniótico, placenta (por ejemplo, las vellosidades coriónicas) y similares. Las muestras de órganos del donante pueden obtenerse en un individuo mediante biopsia. Los organismos infecciosos pueden aislarse directamente de un individuo y analizarse después del aislamiento. El ADN puede extraerse de células cancerosas o tejidos y usarse para el análisis.

45 Una característica de la invención es que los ácidos nucleicos analizados en el sistema de ensayo no requieren diferencias polimórficas entre las fuentes de células para determinar la frecuencia, y por tanto las VNC potenciales, en loci de una muestra mixta. Otra característica de la invención es que la mayoría sustancial de los ácidos nucleicos aislados de la muestra mixta y detectados en el sistema de ensayo proporciona información relevante a la presencia, cantidad y/o naturaleza polimórfica de un locus particular en la muestra mixta. Esto asegura que la mayoría de ácidos nucleicos analizados en el sistema de ensayo de la invención son informativos.

50 En algunos aspectos, se determinan múltiples loci para cada región genómica bajo interrogación, y la cantidad de las regiones seleccionadas presentes en la muestra mixta se suman de manera individual para determinar la frecuencia relativa de un locus en una muestra mixta. Esto incluye la 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 separadas, aunque en determinados aspectos esta información puede obtenerse además de la información de frecuencias relativas en la muestra en su conjunto.

60 En aspectos preferidos, se detectan los ácidos nucleicos seleccionados que corresponden a loci informativos y se suman para determinar la frecuencia relativa de una región genómica en la muestra mixta. Las frecuencias que son más altas que las esperadas para los loci de una primera región genómica cuando se comparan con los loci de un segundo locus en una muestra mixta son indicativas de una VNC de la primera región genómica en la muestra mixta.

65 La comparación de regiones genómicas puede ser una comparación de una parte o de todo un cromosoma. Por ejemplo, la región genómica detectada para la VNC puede ser un cromosoma completo en el feto (por ejemplo, los cromosomas 18 y 21), en los que la probabilidad de que ambos sean aneuploides es mínima. Esto también puede

ser una comparación de cromosomas en la que uno es posiblemente aneuploide (por ejemplo, el 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 utilizar dos o más cromosomas que son posiblemente aneuploides y uno o más cromosomas de referencias.

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

15 En otro aspecto, el sistema de ensayo de la invención analiza múltiples ácidos nucleicos que representan loci seleccionados en cromosomas de interés, y se analiza la frecuencia relativa de cada ácido nucleico seleccionado de la muestra y se cuantifica independientemente para determinar una cantidad relativa para cada locus seleccionado en la muestra. La suma de los loci en la muestra se compara para determinar estadísticamente si existe una VNC para uno o más loci en una región genómica de una fuente celular en una muestra mixta.

20 En otro aspecto, se analizan subconjuntos de loci en cada cromosoma para determinar si existe una alteración cromosómica. La frecuencia de los loci puede sumarse para un cromosoma particular, y la suma de los loci puede usarse para determinar la aneuploidía. Este aspecto de la invención suma las frecuencias de los loci individuales en cada región genómica y entonces compara la suma de los loci en una región genómica de un cromosoma frente a una región genómica de otro cromosoma para determinar si existe una alteración cromosómica. Los subconjuntos de loci pueden elegirse de manera aleatoria pero con suficiente número de loci para producir un resultado estadísticamente significativo para determinar si existe una alteración cromosómica. Pueden realizarse análisis
25 múltiples de diferentes subconjuntos de loci dentro de una muestra mixta para producir más poder estadístico. En otro aspecto, pueden seleccionarse loci particulares que se sabe que tienen menos variación entre muestras, o limitando los datos usados para la determinación de la frecuencia cromosómica, por ejemplo, ignorando los datos de loci con una frecuencia muy alta o muy baja dentro de una muestra.

30 En un aspecto particular, las cantidades medidas de uno o más loci particulares se normalizan para tener en cuenta las diferencias en la cantidad de loci en la muestra. Esto puede realizarse normalizando para cualquier variación conocida de fuentes tales como el sistema de ensayo (por ejemplo, temperatura, diferencias de lote de reactivos), biología subyacente de la muestra (por ejemplo, contenido de ácidos nucleicos), diferencias de operador o cualquier otra variable.

35 En determinados aspectos específicos, la determinación del porcentaje relativo de ácidos nucleicos de la fuente secundaria en una muestra mixta puede ser beneficiosa para realizar el sistema de ensayo, ya que proporcionará información importante sobre la presencia estadística relativa de loci que puede ser indicativa de variación en el número de copias dentro de la fuente secundaria en esa muestra. La determinación de los loci contribuidos a la muestra mixta de la fuente secundaria puede proporcionar información usada para calcular las diferencias estadísticamente significativas en las frecuencias para regiones genómicas de interés. Tales loci pueden por tanto proporcionar dos formas de información en el ensayo (puede usarse información alélica para determinar el porcentaje de una contribución celular secundaria en una muestra mixta y puede usarse una suma de la información alélica para determinar la frecuencia total relativa de ese locus en una muestra mixta). La información alélica no es
40 necesaria para determinar la frecuencia total relativa de ese locus.
45

50 En otro aspecto específico, el sistema de ensayo de la invención puede utilizarse para determinar el posible mosaicismos en una población celular, y si se deben realizar pruebas de confirmación adicionales para confirmar la identificación de mosaicismos 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 puede ayudar en la cuantificación del nivel estimado de mosaicismos. El mosaicismos puede confirmarse posteriormente usando otros métodos de prueba que pueden distinguir aneuploidía parcial o total en mosaico en células o tejido específicos.

55 En aún otro aspecto específico, el sistema de ensayo de la invención puede utilizarse para determinar la contaminación en una muestra, con las especies secundarias que representan especies contaminantes.

60 En otro aspecto, los oligonucleótidos para un ácido nucleico seleccionado dado pueden conectarse en los extremos no específicos de la secuencia de manera que una sonda circular o unimolecular puede unirse a los mismos. 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.

65 Una importante característica del ensayo es que los productos de amplificación se analizan directamente sin la necesidad del enriquecimiento de regiones polimórficas de la muestra mixta inicial. Por tanto, la invención actual permite la detección tanto de la VNC como polimorfismos a partir de una muestra materna sin una etapa de enriquecimiento polimórfico intermedia antes de la determinación de secuencia de los loci seleccionados.

Otra importante característica del ensayo es que tanto la detección de la VNC como del polimorfismo se determinan usando enfoque específico de detección y amplificación seleccionada. Esto permite que la mayoría de la información recogida en el ensayo sea útil para la determinación de la VNC y/o el polimorfismo interrogados en el locus de interés, y evita la necesidad de generar lecturas de secuencia que deben alinearse con una secuencia de referencia.

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

10 **Breve descripción de las figuras**

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

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

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

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

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

La figura 6 ilustra los elementos usados para una detección de aneuploidía y polimorfismo para dos cohortes de muestras maternas.

25 La figura 7 es un informe del paciente y datos e información de la muestra para un subconjunto de una segunda cohorte de sujetos embarazados.

La figura 8 ilustra la detección de aneuploidía del cromosoma 21 alcanzada usando un aspecto de la invención para una primera cohorte.

30 La figura 9 ilustra la detección de aneuploidía del cromosoma 18 alcanzada usando un aspecto de la invención para una primera cohorte.

35 La figura 10 ilustra la detección de aneuploidía del cromosoma 21 alcanzada usando un aspecto de la invención para una segunda cohorte.

La figura 11 ilustra la detección de aneuploidía del cromosoma 18 alcanzada usando un aspecto de la invención para una segunda cohorte.

40 **Definiciones**

Se pretende que los términos usados en el presente documento que tengan el significado simple y ordinario tal como lo entienden los expertos habituales en la técnica. Se pretende que las siguientes definiciones ayuden al lector a entender la presente invención, pero no se pretende que varíen o limiten de otra manera el significado de tales términos a menos que se indique específicamente.

50 El término "ácido nucleico amplificado" es cualquier molécula de ácido nucleico cuya cantidad se ha aumentado al menos dos veces mediante cualquier método de amplificación o replicación de ácidos nucleicos realizado *in vitro* en comparación con su cantidad de partida en una muestra mixta.

El término "producto de amplificación" tal como se usa en el presente documento se refiere al producto que resulta de una reacción de amplificación usando el producto de ligamiento contiguo como molde, o el producto que resulta de una reacción de amplificación usando una molécula complementaria al producto de ligamiento contiguo como molde.

55 El término "alteración cromosómica" se refiere a cualquier variación genética que afecta a todo o una parte de un cromosoma más grande que un locus único. Las variantes genéticas pueden incluir pero no limitarse a cualquier VNC tal como duplicaciones o deleciones, translocaciones, inversiones y mutaciones. Los ejemplos de alteraciones 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 triple X, síndrome XYY, trisomía 8, trisomía 16, síndrome de Turner, translocación de Robertsonian, síndrome de DiGeorge y síndrome de Wolf-Hirschhorn.

65 Los términos "complementario" o "complementariedad" se usan en referencia a moléculas de ácido nucleico (es decir, una secuencia de nucleótidos) que se relacionan mediante reglas de apareamiento de bases. Los nucleótidos complementarios son, generalmente, A y T (o A y U) o C y G. Se dice que dos moléculas de ADN o ARN

monocatenario son sustancialmente complementarias cuando los nucleótidos de una cadena, alineados de manera óptima y con las inserciones o deleciones de nucleótidos apropiadas, se emparejan 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 cadena de ARN o ADN se hibridará en condiciones de hibridación selectiva con su complemento. Las condiciones de hibridación selectiva 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_m).

El término “variación en el número de copias” o “VNC” se usa de manera intercambiable en el presente documento, son alteraciones del ADN de un genoma que dan como resultado una célula que tiene un número anómalo de copias de uno o más loci en el ADN. Las VNC que son clínicamente relevantes pueden limitarse a un gen único o incluyen un conjunto contiguo de genes. Una VNC también puede corresponder a regiones relativamente más grandes del genoma que se han delecionado, invertido o duplicado en determinados cromosomas, hasta incluir 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 a la cantidad o “recuentos” de regiones genéticas presentes en una muestra.

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 los cambios de nucleótidos que pueden producirse fuera de la secuenciación tal como síntesis de oligos, amplificación y cualquier otro aspecto del ensayo. Estos índices de corrección pueden ser índices independientes que son secuencias separadas, o pueden incorporarse dentro de otras regiones para ayudar a confirmar la precisión de las técnicas experimentales usadas, por ejemplo, un índice de corrección puede ser un subconjunto de secuencias usado para la amplificación universal o un subconjunto de nucleótidos de un locus de muestra.

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 un ensayo o prueba de diagnóstico en una muestra del paciente.

El término “rasgo de enfermedad” se refiere a un rasgo monogénico o poligénico asociado con un estado patológico, por ejemplo, una enfermedad, trastorno, síndrome o predisposición.

El término “hibridación” generalmente significa la reacción mediante la cual se produce el apareamiento de cadenas complementarias de ácido nucleico. El ADN es habitualmente bicatenario, y cuando las cadenas se separan se volverán a hibridar en condiciones apropiadas. Pueden formarse híbridos entre ADN-ADN, ADN-ARN o ARN-ARN. Pueden formarse entre una cadena corta y una cadena larga que contiene una región complementaria a la corta. También pueden formarse híbridos imperfectos, pero cuanto más imperfectos son, menos estables serán (y menos propensos a formarse).

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

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

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

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

El término “temperatura de fusión” o T_m se define comúnmente como la temperatura a la que una población de moléculas de ácido nucleico bicatenario se convierte en medio disociada en cadenas simples. La ecuación para calcular la T_m de ácidos nucleicos es bien conocida en la técnica. Tal como lo indica la bibliografía habitual, una estimación simple del valor de T_m puede calcularse mediante la ecuación: $T_m = 81,5 + 16,6(\log_{10}[\text{Na}^+]) - 0,41(\%[\text{G}+\text{C}]) - 675/n - 1,0m$, cuando un ácido nucleico está en disolución acuosa que tiene concentraciones de catión de 0,5 M o menos, el contenido de (G+C) es de 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)). Otra bibliografía incluye computaciones más sofisticadas, que tienen en cuenta las características estructurales así como las de secuencia para el cálculo de T_m .

“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 manera que cada sitio de la matriz comprende copias idénticas o sustancialmente idénticas de oligonucleótidos o polinucleótidos y se define espacialmente y no se solapa con otros sitios de miembros de la matriz; es decir, los sitios son espacialmente discretos. 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 unirse de manera covalente al soporte sólido, o pueden unirse de manera no covalente. La tecnología de micromatriz 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.

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

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 de células de interés, siendo una, una fuente principal y siendo la otra una fuente secundaria dentro de un sólo individuo. Las muestras mixtas incluyen muestras con material genómico tanto de una fuente principal como de 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. Las muestras mixtas se derivan preferiblemente de manera periférica, por ejemplo, de sangre, plasma, suero, etc.

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 gen único. Tales rasgos incluyen rasgos asociados con una enfermedad, trastorno o predisposición producido por una disfunción en un gen único. Los rasgos también incluyen características no patológicas (por ejemplo, presencia o ausencia de moléculas de superficie celular en un tipo de célula específico).

El término alelo “no materno” significa un alelo con un polimorfismo y/o una mutación que se encuentra en un alelo fetal (por ejemplo, un alelo con una mutación o un 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 loci seleccionados, se entiende la 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 trifosfatos de ribonucleósido ATP, UTP, CTG, GTP y trifosfatos de desoxiribonucleósido tales como dATP, dCTP, dITP, dUTP, dGTP, dTTP, o derivados de los mismos. Tales derivados incluyen, por ejemplo, [α S]dATP, 7-deaza-dGTP y 7-deaza-dATP, y derivados de nucleótidos que confieren resistencia a la nucleasa en la molécula de ácido nucleico que los contiene. El término nucleótido tal como se usa en el presente documento también se refiere a trifosfatos de didesoxiribonucleósido (ddNTP) y sus derivados. Los ejemplos ilustrados de trifosfatos de didesoxiribonucleósido incluyen, pero no se limitan a, ddATP, ddCTP, ddGTP, ddITP y ddTTP.

Según la presente invención, un “nucleótido” puede estar no 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, 9ª ed., Molecular Probes, Inc., Eugene OR (2002); Keller y Manak, DNA Probes, 2ª 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 bibliografía: Fung *et al.*, patente estadounidense número 4.757.141; Hobbs, Jr., *et al.*, patente estadounidense número 5.151.507; Cruickshank, patente estadounidense número 5.091.519; Menchen *et al.*, patente estadounidense número 5.188.934; Begot *et al.*, patente estadounidense número 5.366.860; Lee *et al.*, patente estadounidense número 5.847.162; Khanna *et al.*, patente estadounidense número 4.318.846; Lee *et al.*, patente estadounidense número 5.800.996; Lee *et al.*, patente estadounidense número 5.066.580; Mathies *et al.*, patente estadounidense número 5.688.648; y similares. El marcado también puede llevarse a cabo con puntos cuánticos, tal como se da a conocer en las siguientes patentes y publicaciones de patente: patentes estadounidenses números 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 de 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, verde Oregón, rojo Texas, cianina y ácido 5-(2'-aminoetil)aminonaftaleno-1-sulfónico (EDANS). Los ejemplos específicos de nucleótidos marcados de manera fluorescente incluyen [R6G]dUTP, [TAMRA]dUTP, dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, ddATP, [dR110]ddCTP, [dTAMRA]ddGTP y ddTTP disponibles de Perkin Elmer, Foster City, Calif. Los desoxinucleótidos FluoroLink, 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, EN y nucleótidos marcados con cromosoma E, 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, verde Oregón 488-5-dUTP, verde rodamina-5-UTP, verde rodamina-5-dUTP, tetrametilrodamina-6-UTP, tetrametilrodamina-6-dUTP, rojo Texas-5-UTP, rojo Texas-5-dUTP y rojo Texas-12-dUTP disponibles de Molecular Probes, Eugene, OR.

Los términos "oligonucleótidos" u "oligos" tal como se usa en el presente documento se refieren a oligómeros lineales de monómeros de ácidos nucleicos modificados o naturales, incluyendo desoxiribonucleótidos, ribonucleótidos, formas anómericas de los mismos, monómeros de ácidos nucleicos peptídicos (PNA), monómeros de ácidos nucleicos bloqueados (LNA), y similares, o una combinación de los mismos, capaces de unirse específicamente a un polinucleótido monocatenario por medio de un patrón regular de interacciones monómero a monómero, tales como apareamiento de bases de tipo Watson-Crick, apilamiento de bases, apareamiento de bases de tipos 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 desde unas pocas unidades monoméricas, por ejemplo, 8-12, hasta varias decenas de unidades monoméricas, por ejemplo, 100-200 o más. Pueden prepararse moléculas adecuadas de ácido nucleico mediante el método de la fosforamidita descrito por Beaucage y Carruthers (*Tetrahedron Lett.*, 22:1859-1862 (1981)), o mediante el método del triéster según Matteucci, *et al.* (*J. Am. Chem. Soc.*, 103:3185 (1981)), o mediante otros métodos químicos tales como usando 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 un polimorfismo en más de un gen único. Tales rasgos incluyen rasgos asociados con una enfermedad, trastorno, síndrome o predisposición producido 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 juntos en una cadena larga, usando otra cadena como molde. Existen dos tipos generales de polimerasa: ADN polimerasas, que sintetizan ADN, y ARN polimerasas, que sintetizan ARN. Dentro de estas dos clases, numerosos subtipos de polimerasas, dependiendo de 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 ADN no específico en exceso. Se añaden cebadores al ADN seleccionado, donde los cebadores inician el copiado del ADN seleccionado usando nucleótidos y, normalmente, polimerasa Taq o similares. Sometiendo a ciclos la temperatura, el ADN seleccionado se desnaturaliza y se copia de manera repetida. Una copia única del ADN seleccionado, incluso si se mezcla con otro, ADN aleatorio, puede amplificarse para obtener miles de millones 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, los métodos de amplificación lineal pueden usarse como una alternativa la PCR.

El término "polimorfismo" tal como se usa en el presente documento se refiere a cualquier cambio genético o

variantes 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 solo gen, mutaciones puntuales, repeticiones de trinucleótidos, indels y similares.

5 Generalmente, un “cebador” es un oligonucleótido usado para, por ejemplo, cebar la extensión, ligamiento y/o síntesis del ADN, tales como en la etapa de síntesis de la reacción en cadena de la polimerasa o en las técnicas de extensión de cebadores usadas en determinadas reacciones de secuenciación. Un cebador también puede usarse en técnicas de hibridación como medio para proporcionar complementariedad de un locus a un oligonucleótido de captura para la detección de un locus específico.

10 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 estudio científico, de naturaleza académica o comercial, incluyendo el desarrollo de agentes terapéuticos farmacéuticos y/o biológicos. Las herramientas de investigación de la invención no se pretende que sean terapéuticas o que estén sujetas 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 una prueba regulatoria.

20 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 análisis de múltiples muestras). Por tanto el índice de muestra puede usarse para ayudar en la identificación de locus para el multiplexado de diferentes muestras en un recipiente de reacción único, de manera que cada muestra pueda identificarse basándose en su índice de muestra. En un aspecto preferido, hay un índice de muestra único para cada muestra en un conjunto de muestras, y las muestras se reúnen durante la secuenciación. Por ejemplo, si se reúnen doce muestras dentro de una reacción de secuenciación única, hay al menos doce índices de muestra de manera que cada muestra se marca de manera única.

30 El término “locus seleccionado” tal como se usa en el presente documento se refiere a un locus que corresponde a unos loci interrogados, por ejemplo, para el número de copias, la presencia o ausencia de uno o más polimorfismos, la presencia o ausencia de un organismo infeccioso, etc. Tales loci seleccionados pueden aislarse directamente y amplificarse a partir de la muestra para la detección, por ejemplo, basándose en la hibridación y/u otras técnicas basadas en secuencias, o pueden amplificarse usando la muestra como un molde antes de la detección de la secuencia. Las regiones de ácidos nucleicos para su uso en los sistemas de ensayo de la presente invención pueden seleccionarse en función de la variación del nivel de ADN entre individuos, basándose en la especificidad para un cromosoma particular, basándose en el contenido de CG y/o en las condiciones de amplificación requeridas de los loci seleccionados, u otras características que resultarán evidentes para un experto en la técnica tras leer la presente divulgación.

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

45 El término “unirse 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 un cebador de ácido nucleico, un 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 el doble de la desviación estándar de cualquier señal generada como resultado de interacciones no deseadas (antecedentes).

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

65

Descripción detallada de la invención

Los sistemas de ensayo y métodos descritos en el presente documento pueden emplear, a menos que se indique lo contrario, técnicas y descripciones convencionales de biología molecular (incluyendo técnicas recombinantes), biología celular, bioquímica, tecnología de secuenciación y micromatriz, que están dentro de la experiencia de los expertos en la técnica. Tales técnicas convencionales incluyen síntesis de matrices de polímero, hibridación y ligamiento de oligonucleótidos, secuenciación de oligonucleótidos y detección de hibridación usando un marcador. Se pueden tener ilustraciones específicas de técnicas adecuadas mediante la referencia a los ejemplos en el presente documento. Sin embargo, también pueden usarse, por supuesto, procedimientos convencionales equivalentes. Tales técnicas y descripciones convencionales pueden encontrarse en manuales de laboratorio habituales tales como Green, *et al.*, Eds., *Genome Analysis: A Laboratory Manual Series* (volúmenes 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, *ADN 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., *Bioquímica* (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, se entenderá que esta invención no se limita a métodos, composiciones, objetivos y usos específicos descritos, ya que, por supuesto, esto puede variar. También se entenderá que la terminología usada en el presente documento tiene el propósito de describir aspectos particulares solamente y no se pretende que limite el alcance de la presente invención, que se limitará solamente mediante las reivindicaciones adjuntas.

Se observará que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un," "una," y "el/la" incluyen los referentes plurales 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 referencias a etapas y métodos equivalentes conocidos por los expertos en la técnica, y así sucesivamente.

Cuando se proporciona un intervalo de valores, se entenderá que cada valor intermedio entre el límite superior y el inferior de ese intervalo (y cualquier otro valor declarado o intermedio en ese intervalo establecido) está incluido dentro de la invención. Cuando el intervalo indicado incluye límites superiores e inferiores, los intervalos que excluyen cualquiera de los límites incluidos también se incluyen en la invención.

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

En la siguiente descripción, se exponen numerosos detalles específicos para proporcionar una comprensión más completa de la presente invención. Sin embargo, será evidente para un experto en la materia que la presente invención puede ponerse en práctica sin uno o más de estos detalles específicos. En otros casos, las características y procedimientos bien conocidos por los expertos en la técnica no se han descrito para evitar dificultar la comprensión de la invención.

45 **La invención en general**

La presente invención proporciona métodos de ensayo único con la capacidad de detectar variaciones en el número de copias, polimorfismos y enfermedad infecciosa en una muestra mixta a partir de un sólo individuo. Los métodos de ensayo comprenden identificar el número de copias de loci seleccionados (incluyendo múltiples loci que se asocian con alteraciones cromosómicas) y la detección de polimorfismos que se encuentran en una fuente principal y/o una fuente secundaria. Estos métodos son útiles para cualquier muestra mixta que contenga material genómico (por ejemplo, ADN) de una fuente principal y una secundaria que están presentes en un sólo individuo.

El uso de loci seleccionados en los métodos de ensayo de la invención proporciona amplificación de loci para la detección de la variación en el número de copias. Una clara ventaja de la invención es que los loci seleccionados que corresponden 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, técnicas de determinación de secuenciación de alto rendimiento y PCR digital. Las sondas de selección pueden diseñarse frente a cualquier número de loci para cualquier cromosoma. Aunque la amplificación de la muestra mixta antes de la identificación y cuantificación de las regiones de selección de ácidos nucleicos no es obligatoria, puede realizarse amplificación limitada antes de la detección, en particular si las cantidades iniciales de ácido nucleico están limitadas.

La figura 1 es un diagrama de flujo simplificado de las etapas generales utilizadas en los sistemas de ensayo de la invención. La figura 1 muestra el método 100, en el que en una primera etapa 110, se proporciona una muestra de

- ácidos nucleicos mixta para el análisis. La muestra mixta puede prepararse a partir de prácticamente cualquier muestra ya que tales técnicas son conocidas por los expertos en la técnica (véanse, 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)). Dependiendo del tipo de muestra mixta elegida, pueden realizarse etapas de procesamiento y/o purificación adicionales para obtener fragmentos de ácido nucleico de una pureza o tamaño deseado, usando métodos de procesamiento incluyendo pero sin limitarse a sonicación, nebulización, purificación de gel, sistemas de purificación de la PCR, escisión de la nucleasa, o una combinación de estos métodos. En un aspecto preferido, se usan muestras que comprenden ADN libre de células (ADNlc).
- 15 En la etapa 120, se introduce un primer conjunto de oligonucleótidos de secuencia fija en la muestra de ácidos nucleicos mixta, en condiciones que permiten que el primer conjunto de oligonucleótidos de secuencia fija se hibride con la muestra de ácidos nucleicos mixta. El primer conjunto de oligonucleótidos de secuencia fija comprende secuencias de ácido nucleico que son complementarias a un locus seleccionado o loci en la muestra mixta, que tal como se describirá en detalle en el presente documento son útiles para determinar las variaciones en el número de copias y/o las alteraciones cromosómicas. Las secuencias de ácido nucleico capaces de determinar las variaciones en el número de copias y/o alteraciones cromosómicas incluyen secuencias que permiten la identificación de alteraciones cromosómicas tales como duplicaciones o deleciones, aneuploidías, translocaciones o inversiones.
- 20 En la etapa 130, se introduce un segundo conjunto de oligonucleótidos de secuencia fija en la muestra de ácidos nucleicos mixta y el primer conjunto de oligonucleótidos de secuencia fija en condiciones que permite que el segundo conjunto de oligonucleótidos de secuencia fija se hibride con la muestra de ácidos nucleicos mixta. El segundo conjunto de oligonucleótidos de secuencia fija comprende secuencias de ácido nucleico que son complementarias a un locus seleccionado o loci en la muestra mixta, capaz de detectar polimorfismos. Pueden incluirse etapas de lavado opcionalmente entre las etapas 120 y 130, y 130 y 140.
- 25 Aunque la invención se describe como los dos conjuntos de oligos introducidos en la muestra mixta de manera secuencial, el orden de los conjuntos puede invertirse respecto al descrito en las figuras o, en aspectos preferidos, pueden introducirse simultáneamente.
- 30 En la etapa 140, se ligan el primer y el segundo conjuntos de oligonucleótidos de secuencia fija que se han hibridado con regiones adyacentes de los loci seleccionados en la muestra mixta, y en la etapa 150, se amplifican los oligonucleótidos ligados. Entonces se detectan y se analizan los oligonucleótidos ligados y amplificados, que permiten la determinación de variaciones en el número de copias o alteraciones cromosómicas y la identificación de polimorfismos en la etapa 160.
- 35 Los conjuntos de ácidos nucleicos de secuencia fija se diseñan para hibridar al menos dos regiones separadas en un locus seleccionado. En aspectos preferidos, dos o más oligos separados se usan para hibridar con estas regiones para proporcionar ácidos nucleicos adyacentes complementarios al locus seleccionado. En algunos aspectos, sin embargo, puede usarse una sonda única que comprende dos o más regiones no adyacentes distintas que son complementarias a los loci seleccionados incluyendo sondas precirculares tales como las denominadas "sondas candado" o "sondas de inversión molecular (MIP)".
- 40 La presente invención proporciona un sistema mejorado sobre técnicas más aleatorias tales como secuenciación paralela de manera masiva, secuenciación al azar y el uso de PCR digital aleatoria que otros han usado para detectar las VNC. Estos enfoques anteriormente mencionados se basan en la secuenciación de todos o una población estadísticamente significativa de fragmentos de ADN en una muestra, seguido por la cartografía de estos fragmentos o si no asociando los fragmentos a sus cromosomas apropiados. Los fragmentos identificados se comparan entonces entre sí o con cualquier otra referencia (por ejemplo, composición cromosómica normal) para determinar las VNC en cromosomas particulares. Estos métodos son intrínsecamente ineficientes en comparación con la presente invención, ya que los cromosomas primarios de interés sólo constituyen una minoría de datos que se generan a partir de la detección de tales fragmentos de ADN en las muestras mixtas.
- 45 Los ensayos de la presente invención proporcionan detección dirigida de los loci seleccionados, que proporciona información tanto del contenido de la región seleccionada (es decir, presencia de una región polimórfica) como de la información sobre la frecuencia de la región detectada en una muestra (con o sin detectar cualquier posible polimorfismo en esa región). Esta característica clave proporciona la capacidad de detectar tanto el número de copias de regiones seleccionadas como la presencia o ausencia de polimorfismos en una región seleccionada como un conjunto de datos único del rendimiento de un ensayo multiplexado de la invención.
- 50 Las técnicas que dependen de un muestreo muy amplio de ADN en una muestra proporcionan una cobertura muy amplia del ADN analizado, pero en realidad muestrean el ADN contenido dentro de una muestra en una base de IX o
- 55
- 60
- 65

menor (es decir, submuestreo). En cambio, la amplificación selectiva usada en los presentes ensayos se diseña específicamente para proporcionar profundidad de cobertura de los loci particulares de interés, y proporciona un "supermuestreo" de tales loci seleccionados con una cobertura de secuencia promedio de preferiblemente 2X o más, más preferiblemente cobertura de secuencia de 100X de más, incluso más preferiblemente cobertura de secuencia de 1000X o más de los loci seleccionados (incluyendo de una o más fuentes secundarias) presentes en la muestra mixta inicial.

Una clara ventaja de la divulgación es que pueden analizarse los productos de ligamiento que resultan de los ensayos que corresponden a alteraciones cromosómicas y/o polimorfismos y alteraciones cromosómicas usando una variedad de técnicas de detección y cuantificación, incluyendo pero sin limitarse a técnicas de hibridación, técnicas de determinación de secuenciación de alto rendimiento y PCR digital.

Los sistemas de ensayo de la invención proporcionan un uso más eficiente y económico de los datos, y la mayoría sustancial de secuencias analizadas después de la amplificación de la muestra da como resultado una información afirmativa sobre la presencia de una VNC particular en la muestra mixta. Por tanto, a diferencia de las técnicas que se basan en la secuenciación paralela de manera masiva o "recuento" digital aleatorio de regiones de cromosomas y la posterior identificación de datos relevantes a partir de tales, el sistema de ensayo de la invención proporciona un uso mucho más eficiente de la colección de datos que los enfoques aleatorios enseñados por otros en la técnica.

20 Métodos de ensayo

Los sistemas de ensayo de la invención utilizan un esquema general tal como se describió anteriormente, aunque pueden emplearse muchas configuraciones y variaciones diferentes, algunas de las cuales se describen a continuación y más de las cuales se ejemplifican en el documento estadounidense con número de serie 61/371605 presentada el 6 de agosto de 2010 y el documento estadounidense con número de serie 13/013.732.

La figura 2 ilustra un primer esquema general para un sistema de ensayo basado en ligamiento de la invención. Los oligonucleótidos 201, 203 de secuencia fija comprenden regiones 209 y 211 de cebadores universales, respectivamente y regiones complementarias a los locus 205 y 207 seleccionados, respectivamente. Sin embargo, además, el sistema de ensayo en la figura 2 emplea una región 221 de índice de muestra en el primer oligonucleótido 201 de secuencia fija. En determinados aspectos, todas o una parte de las secuencias de los loci seleccionados se detectan directamente usando las técnicas descritas, por ejemplo, mediante técnicas de determinación o hibridación de secuencia. En el ejemplo de la figura 2, se asocia un índice de muestra con el primer oligonucleótido 201 de secuencia fija. La detección de los índices puede identificar una secuencia de una muestra específica en un sistema de ensayo altamente multiplexado.

En la etapa 202, los oligonucleótidos 201, 203 de secuencia fija se introducen en la etapa 202 en la muestra 200 mixta y se les permite que se unan específicamente al locus 215 seleccionado. Después de 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 oligo de puente y se le permite que se hibride en la etapa 204 con la región del locus 215 entre los oligonucleótidos primero 201 y el segundo 203 de secuencia fija. Los oligonucleótidos unidos se ligan en la etapa 206 para crear un ácido nucleico contiguo que se expande sobre y es complementario al locus de interés. En determinados aspectos de la invención, los oligonucleótidos de puente son de entre 2-45 nucleótidos de longitud. En un aspecto específico, los oligonucleótidos de puente son de entre 3-9 nucleótidos de longitud. En aún otro aspecto específico, los oligonucleótidos de puente son de entre 10-30 nucleótidos de longitud.

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

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

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

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

Cada uno de los oligonucleótidos de secuencia fija comprende una región complementaria al locus 305, 307 seleccionado, y regiones 309, 311 de cebadores universales usadas para amplificar los diferente loci seleccionados después de la selección y/o el aislamiento inicial de los loci seleccionados a partir de la muestra mixta. Las regiones de cebadores universales se ubican en los extremos de los oligonucleótidos 301, 303 y 323 de secuencia fija que flanquean los índices y las regiones complementarias al ácido nucleico de interés, por tanto conservan las secuencias específicas de ácido nucleico y los índices de muestra en los productos de cualquier método de amplificación universal. Los oligonucleótidos 301, 303, 323 de secuencia fija se introducen en la etapa 302 en una alícuota de la muestra genética 300 y se les permite que se unan específicamente a los loci 315 o 325 seleccionados. Después de 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 oligos de puente que corresponden a un A/T SNP 313 ó un G/C SNP 333 se introducen y se les permite unirse en la etapa 304 a la región del locus 315 o 325 seleccionado entre las regiones primera 305 y segunda 307 complementarias de ácido nucleico de los oligonucleótidos de secuencia fija. Alternativamente, los oligos 313, 333 de puente pueden introducirse en la muestra simultáneamente con los oligonucleótidos de secuencia fija. Los oligonucleótidos unidos se ligan en la etapa 306 en la mezcla de reacción única para crear un ácido nucleico contiguo que se extiende sobre y es complementario al locus seleccionado.

Después del ligamiento, las reacciones separadas pueden combinarse preferiblemente para las etapas de amplificación universal y detección. Se introducen cebadores 317, 319 universales en las reacciones combinadas en la etapa 308 para amplificar las regiones de molde ligadas y crear en la etapa 310 el primer y el segundo oligos de secuencia fija ligados y productos 327, 329 de oligos de puente que comprenden la secuencia del locus seleccionado que representa ambos SNP en el locus seleccionado. Estos productos 327, 329 de ligamiento se detectan y cuantifican a través de la determinación de secuencia del producto de ligamiento, 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 oligo de puente puede hibridarse con una región que no es directamente adyacente a la región complementaria a uno o ambos de los oligos de secuencia fija, y es necesaria una etapa intermedia que requiere la extensión de uno o más de los oligos antes del ligamiento. 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 413 de puente. Cada uno de los oligonucleótidos de secuencia fija comprende una región complementaria al locus 405, 407 seleccionado, y secuencias de cebadores, preferiblemente secuencias 409, 411 de cebadores universales, es decir, regiones de oligos complementarias a los cebadores universales. Las secuencias 409, 411 de cebadores se ubican en o cerca de los extremos de los oligonucleótidos 401, 403 de secuencia fija, y por tanto conservan las secuencias específicas de ácido nucleico en los productos de cualquier método de amplificación universal. Los oligonucleótidos 401, 403 de secuencia fija se introducen en la etapa 402 en la muestra mixta 400 y se permite que se unan específicamente a las partes complementarias del locus 415 de interés. Después de la hibridación, los oligonucleótidos de secuencia fija no hibridados se separan preferiblemente del resto de la muestra genética (no mostrado).

El oligonucleótido de puente se introduce entonces en la etapa 404 y se permite que se una a la región del locus 415 seleccionado entre los oligonucleótidos primero 401 y segundo 403 de secuencia fija. Alternativamente, el oligo de

puente puede introducirse simultáneamente con los oligonucleótidos de secuencia fija. En este aspecto a modo de ejemplo, el oligo de puente se hibrida con una región directamente adyacente a la primera región 405 de oligos de secuencia fija, pero está separado por uno o más nucleótidos de la región complementaria del segundo oligonucleótido 407 de secuencia fija. Después de la hibridación de la secuencia fija y los oligos de puente, el oligo 5 413 de puente se extiende en la etapa 406, por ejemplo, usando una polimerasa y dNTP, para rellenar el hueco entre el oligo 413 de puente y el segundo oligo 403 de secuencia fija. Después de la extensión, los oligonucleótidos unidos se ligan en la etapa 408 para crear un ácido nucleico contiguo que se extiende sobre y es complementario al locus 415 de interés. Después del ligamiento, se introducen cebadores 417, 419 universales en la etapa 410 para 10 amplificar los oligos primero, segundo y de puente para crear en la etapa 412 productos 423 de amplificación que comprenden la secuencia del locus seleccionado de interés. Los productos 423 de amplificación se aíslan, detectan y cuantifican 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

15 Los sistemas de ensayo utilizan sondas de ácido nucleico diseñadas para identificar, y preferiblemente aislar, regiones seleccionadas de ácidos nucleicos en una muestra mixta. Algunas de las sondas identifican secuencias de interés en loci seleccionados interrogados para el número de copias (es decir, frecuencia de loci), y otras sondas 20 identifican secuencias que corresponden a polimorfismos de interés (es decir, contenido de loci) en ácidos nucleicos que corresponden a una fuente principal o una fuente secundaria en una muestra mixta.

En aspectos específicos, los sistemas de ensayo de la invención emplean 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 loci seleccionados analizados. Esto está en contraste 25 directo al enfoque de amplificación aleatoria usado por otros que emplean, por ejemplo, secuenciación paralela de manera masiva, ya que tales técnicas de amplificación implican generalmente amplificación aleatoria de todo o una parte sustancial del genoma. Además, aunque la muestra inicial puede enriquecerse usando métodos tales como amplificación general para aumentar el número de copias de ácidos nucleicos en la muestra mixta, preferiblemente no se usan etapas de enriquecimiento antes de las etapas de hibridación, ligamiento y amplificación usadas para 30 identificar los loci de interés.

En un aspecto general, el usuario de la invención analiza múltiples loci seleccionados en diferentes cromosomas. Cuando se analizan múltiples loci para una muestra, una realización preferida es para amplificar todos los loci 35 seleccionados para cada muestra en un recipiente de reacción. La frecuencia o cantidad de los múltiples loci seleccionados se analizan para determinar si existe una alteración cromosómica, y se analiza la presencia o ausencia de un polimorfismo para determinar la presencia o ausencia o el cálculo de probabilidad de una alteración cromosómica en una fuente en la muestra mixta.

En aspectos preferidos, pueden amplificarse múltiples loci seleccionados de dos o más muestras en un recipiente de 40 reacción único, y analizarse la información simultáneamente en un conjunto de datos único, por ejemplo, a través de la determinación de secuencia. Los datos resultantes se analizan entonces para separar los resultados para las diferentes muestras y se usan para determinar la presencia de ausencia de VNC y/o la presencia de ausencia de polimorfismos para muestras individuales.

45 En un aspecto, se identifican alteraciones cromosómicas en el sistema de ensayo de la invención usando múltiples loci seleccionados en múltiples cromosomas, y la frecuencia de los loci seleccionados en los múltiples cromosomas se compara para identificar un aumento de la probabilidad de aneuploidía basándose en las razones de los cromosomas. La normalización o estandarización de las frecuencias puede realizarse para uno o más loci 50 seleccionados.

En otro aspecto, el sistema de ensayo suma las frecuencias de los loci seleccionados en dos o más cromosomas y entonces compara la suma de los loci seleccionados en un cromosoma frente a otro cromosoma para determinar si 55 existe una aneuploidía cromosómica. En otro aspecto, el sistema de ensayo analiza subconjuntos de loci seleccionados en dos o más cromosomas para determinar si existe una aneuploidía cromosómica para uno de los dos cromosomas. La comparación puede realizarse dentro de los mismos o diferentes cromosomas.

En determinados aspectos, los datos usados para determinar la frecuencia de los loci seleccionados pueden excluir 60 datos atípicos que parecen ser debidos a error experimental, o que tienen niveles elevados o reducidos basados 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 loci seleccionados que se encuentran en una abundancia particularmente baja en una o más muestras.

65 En otro aspecto pueden elegirse subconjuntos de loci de manera aleatoria pero con un número suficiente de loci para producir un resultado estadísticamente significativo para determinar si existe una alteración cromosómica. Pueden realizarse múltiples análisis de diferentes subconjuntos de loci dentro de una muestra mixta para producir

más poder estadístico. Por ejemplo, si existen 100 regiones seleccionadas para el cromosoma 21 y 100 regiones seleccionadas para el cromosoma 18, pueden realizarse una serie de análisis para evaluar menos de 100 regiones para cada uno de los cromosomas. En este ejemplo, los loci seleccionados no se excluyen de manera selectiva.

- 5 La cantidad de diferentes ácidos nucleicos detectables en determinados cromosomas puede variar dependiendo de un número de factores, incluyendo la representación general de loci en diferentes fuentes de células en muestras maternas, las tasas de degradación de los diferentes ácidos nucleicos que representan diferentes loci en muestras mixtas, métodos de preparación de muestras, y similares. Por tanto, en otro aspecto, la cantidad de loci particulares en un cromosoma se suma para determinar la cantidad de loci para diferentes cromosomas en la muestra. Las frecuencias de loci se suman para un cromosoma particular, y la suma de los loci se usa para determinar la aneuploidía. Este aspecto de la invención suma las frecuencias de los loci individuales en cada cromosoma y entonces compara la suma de los loci en un cromosoma frente a otro cromosoma para determinar si existe una alteración cromosómica.
- 10
- 15 Los ácidos nucleicos analizados usando los sistemas de ensayo de la invención se amplifican preferiblemente de 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 tal amplificación selectiva diseñados para aislar regiones pueden elegirse por diversas razones, pero se diseñan preferiblemente para 1) amplificar de manera eficaz una región del cromosoma de interés; 2) tener un intervalo de expresión predecible de fuentes maternas y/o fetales en diferentes muestras mixtas; 3) ser distintivo para el 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.
- 20

25 El sistema de ensayo de la invención detecta tanto aneuploidías fetales como alteraciones cromosómicas específicas a través de la identificación y cuantificación de loci específicos de interés. Tales alteraciones adicionales 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 secuencias seleccionadas de ácido nucleico presentes en el mismo o cromosomas separados, y al menos uno de los loci seleccionados se asocia con la alteración alélica fetal. Las secuencias de los dos loci seleccionados y el número de copias de los dos loci seleccionados se comparan entonces para determinar si está presente la alteración cromosómica, y si es así, la naturaleza de la alteración.

30

35 Si bien gran parte de la descripción contenida en el presente documento describe la detección de aneuploidía contando la abundancia de loci en uno o más posibles cromosomas aneuploides y la abundancia de loci en uno o más cromosomas normales, pueden usarse las mismas técnicas para detectar variaciones en el número de copias en las que la variación en el número de copias se produce solamente en una parte de un cromosoma. En esta detección de las variaciones en el número de copias, se comparan múltiples loci dentro de la posible ubicación de la variación en el número de copias para múltiples loci fuera de la posible ubicación de la variación en el 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 regiones nucleicas dentro de la delección de 22q11 y dos o más loci fuera de la delección de 22q11. Los loci fuera de la delección de 22q11 pueden estar en otra región del cromosoma 22 o pueden estar en un cromosoma completamente diferente. La abundancia de cada loci se determina mediante los métodos descritos en esta solicitud.

40

45

En algunos aspectos puede usarse una amplificación universal para amplificar los loci. En algunos aspectos, los loci para cada muestra se someten a ensayo en una reacción única en un recipiente único. En otros aspectos, los loci de múltiples muestras pueden someterse a ensayo en una reacción única en un recipiente único.

- 50 Determinados aspectos de la invención pueden detectar una delección, incluyendo los límites de tales delecciones. En algunos aspectos, al menos 24 loci seleccionados pueden usarse dentro de la región de la posible delección y al menos 24 loci seleccionados pueden usarse fuera de la región de la posible delección. En otro aspecto al menos 48 loci seleccionados pueden usarse dentro de la región de la posible delección y al menos 48 loci seleccionados pueden usarse fuera de la región de la posible delección. En otro aspecto al menos 48 loci seleccionados pueden usarse dentro de la región de la posible delección y al menos 96 loci seleccionados pueden usarse fuera de la región de la posible delección. En otro aspecto al menos 48 loci seleccionados pueden usarse dentro de la región de la posible delección y al menos 192 loci seleccionados pueden usarse fuera de la región de la posible delección. En un aspecto preferido al menos 384 loci seleccionados pueden usarse dentro de la región de la posible delección y al menos 384 loci seleccionados pueden usarse fuera de la región de la posible delección. Los loci dentro de la delección se suman entonces como los loci fuera de la delección. Estas sumas se comparan entonces entre sí para determinar la presencia o ausencia de una delección. Opcionalmente, las sumas se colocan en una razón y esa razón puede compararse con una razón promedio creada a partir de una población normal. Cuando la razón para una muestra se encuentra fuera estadísticamente de una razón esperada, se detecta la delección. El umbral para la detección de una delección puede ser el doble o más, preferiblemente cuatro o más veces la variación calculada en la población normal.
- 55
- 60
- 65

Polimorfismos asociados con enfermedades o predisposiciones

- Los sistemas de ensayo de la invención se utilizan para detectar polimorfismos, tales como los asociados con un trastorno de predisposición o enfermedad autosómica recesiva o dominante. Dada la naturaleza multiplexada de los sistemas de ensayo de la invención, esta detección tiene lugar en el mismo ensayo que la detección de alteraciones cromosómicas. Por tanto, un sistema de ensayo único puede proporcionar información diagnóstica en diferentes clases de mutaciones genéticas. Por consiguiente, como los sistemas de ensayo preferidos de la invención son altamente multiplexados y capaces de interrogar cientos o incluso miles de ácidos nucleicos dentro de una muestra mixta, en determinados aspectos es deseable interrogar la muestra para marcadores de ácidos nucleicos dentro de la muestra mixta, por ejemplo, ácidos nucleicos asociados con riesgo genético o que identifican la presencia o ausencia de organismos infecciosos. Por tanto, los sistemas de ensayo proporcionan la detección de tales ácidos nucleicos junto con la detección de ácidos nucleicos para la determinación del número de copias dentro de una muestra mixta.
- Por tanto, el sistema de ensayo de la invención puede usarse para detectar polimorfismos en una muestra mixta, en el que tales polimorfismos se asocian con genes asociados con trastornos autosómicos recesivos, mutaciones asociadas con trastornos autosómicos dominantes; polimorfismos asociados con el riesgo de desarrollar una enfermedad y/o progresión de la enfermedad (por ejemplo, metástasis) e indicadores de pronóstico.
- En otros aspectos específicos, el sistema de ensayo de la invención puede usarse para detectar mutaciones o polimorfismos fetales en una muestra materna, en el que tales mutaciones o polimorfismos se asocian 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 de susceptibilidad al cáncer, (por ejemplo mutaciones en BRCA1 o II o en p53); polimorfismos asociados con un riesgo aumentado de desarrollar enfermedades de aparición tardía, tal como el polimorfismo del gen apoE3 asociado con el riesgo de Alzheimer.

Además de la detección de alteraciones cromosómicas y mutaciones o polimorfismos de un solo gen asociados con predisposiciones, trastornos o enfermedades monogénicas o poligénicas, los sistemas de ensayo de la invención pueden identificar agentes infecciosos en la muestra mixta.

Amplificación seleccionada

Pueden usarse varios 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 los números de copias de un locus de interés en una muestra mixta de una manera que permite la conservación de información sobre el contenido inicial del locus en la muestra mixta. Aunque no todas las combinaciones de amplificación y análisis se describen en este documento en detalle, es bien conocido dentro de la experiencia de los expertos en la técnica utilizar diferentes métodos de amplificación y/o herramientas analíticas para aislar y/o analizar los ácidos nucleicos de la región compatibles con esta memoria descriptiva, y tales variaciones serán evidentes para un experto en la técnica al 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 ADN Amplification, ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), reacción en cadena de la ligasa (LCR) (Wu y Wallace, Genomics 4:560, 1989; Landegren *et al.*, Science 241:1077, 1988), amplificación de desplazamiento de cadena (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 vinculada (LLA) (patente estadounidense n.º 6.027.923), y similares, replicación de secuencia autosostenida (Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87, 1874 (1990) y documento WO90/06995), amplificación selectiva de secuencias de polinucleótidos diana (patente estadounidense n.º 6.410.276), reacción en cadena de la polimerasa cebada con secuencias 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 amplificación de secuencia basada en ácidos 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: replicasa Qbeta, descrito en la solicitud de patente PCT n.º PCT/US87/00880, métodos de amplificación isotérmica 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 el documento estadounidense con número de serie 09/854.317 y en la publicación estadounidense n.º 20030143599.

En algunos aspectos, el ADN se amplifica mediante PCR múltiplex específica de locus. En un aspecto preferido, el ADN se amplifica usando ligamiento con adaptador y PCR de cebador único. Otros métodos de amplificación disponibles, tales como PCR equilibrada (Makrigiorgos, *et al.*, Nature Biotechnol, 20:936-9 (2002)) y métodos de amplificación isotérmica tales como amplificación basada en secuencias de ácidos nucleicos (NASBA) y replicación

de secuencia autosostenida (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 el ADN de cualquier longitud siempre que contenga el locus de interés en su secuencia.

5 La longitud de un locus seleccionado amplificado de una región genómica de interés es lo suficientemente larga como para proporcionar suficiente información de secuencia para distinguirla de otros ácidos nucleicos que se amplifican y/o se seleccionan. En general, un ácido nucleico amplificado tiene al menos aproximadamente 16 nucleótidos de longitud, y más típicamente, un ácido nucleico amplificado tiene al menos aproximadamente 20 nucleótidos de longitud. En un aspecto preferido de la invención, un ácido nucleico amplificado tiene al menos aproximadamente 30 nucleótidos de longitud. En un aspecto más preferido de la invención, un ácido nucleico amplificado tiene al menos aproximadamente 32, 40, 45, 50 ó 60 nucleótidos de longitud. En otros aspectos de la invención, un ácido nucleico amplificado puede tener aproximadamente 100, 150 ó hasta 200 de longitud.

15 En determinados aspectos, la amplificación seleccionada comprende una etapa de amplificación lineal inicial. Esto 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 la cuantificación precisa del ADN o una fracción del ADN (por ejemplo contribución del ADN fetal en una muestra materna).

20 Por tanto, en un aspecto, se realiza un número limitado de ciclos de amplificación lineal específica de secuencia en la muestra mixta de partida que comprende el ADNlc. El número de ciclos es generalmente menor que el usado para una amplificación de PCR típica, por ejemplo 5-30 ciclos o menos. Los cebadores o sondas pueden diseñarse para amplificar segmentos o regiones genómicas específicas. Los cebadores o sondas pueden modificarse con una etiqueta de extremo en el extremo 5' (por ejemplo con biotina) o en otro lugar a lo largo del cebador o la sonda, de modo que los productos de amplificación puedan 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 reacción única produce múltiples fragmentos de ADN de diferentes regiones. Los productos de amplificación de la amplificación lineal podrían amplificarse adicionalmente con métodos de PCR habituales o con amplificación lineal adicional.

25 Por ejemplo, puede aislarse ADNlc de sangre, plasma o suero de una mujer embarazada e incubarse con cebadores frente a un número establecido de loci que corresponden a cromosomas de interés. Preferiblemente, el número de cebadores usados para la amplificación lineal inicial será de 12 o más, más preferiblemente de 24 o más, más preferiblemente de 36 o más, incluso más preferiblemente de 48 o más e incluso más preferiblemente de 96 o más. Cada uno de los cebadores corresponde a un locus único, 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 mediante la unión a avidina o estreptavidina en un sustrato sólido. Luego, los productos se someten a procedimientos bioquímicos adicionales, como una amplificación adicional con otros cebadores y/o técnicas de detección, como la determinación de secuencias y la hibridación.

40 Las eficacias de la amplificación lineal pueden variar entre los sitios y entre los ciclos, de modo que en determinados sistemas puede usarse la normalización para garantizar que los productos de la amplificación lineal sean representativos del material de partida del contenido de ácido nucleico. Una persona que practica el sistema de ensayo de la invención puede utilizar información de diversas muestras para determinar la variación en los niveles de ácido nucleico, incluida la variación en diferentes loci en muestras individuales y/o entre los mismos loci en diferentes muestras después de la amplificación lineal inicial limitada. Dicha información puede usarse en la normalización para evitar la distorsión de los niveles iniciales de contenido de ADN.

Amplificación universal

55 En aspectos preferidos de la invención, los loci amplificados de manera selectiva se amplifican de manera adicional preferiblemente a través de la amplificación universal de todos o sustancialmente todos los diversos loci que van a analizarse usando los sistemas de ensayo de la invención. Las regiones de cebadores universales se añaden a los oligonucleótidos de secuencia fija para que los loci amplificados de manera selectiva puedan amplificarse adicionalmente en una reacción de amplificación universal única. Estas secuencias de cebadores universales pueden añadirse a las regiones de ácidos nucleicos durante el procedimiento de amplificación selectiva, es decir, los cebadores para amplificación selectiva tienen secuencias de cebadores universales que flanquean un locus. Alternativamente, pueden añadirse adaptadores que comprenden secuencias de amplificación universales a los extremos de los ácidos nucleicos seleccionados como adaptadores después de la amplificación y aislamiento de los ácidos nucleicos seleccionados de la muestra mixta.

65 En un aspecto a modo de ejemplo, los ácidos nucleicos se amplifican inicialmente a partir de una muestra mixta usando cebadores complementarios a regiones seleccionadas de los cromosomas de interés, seguidos de una

etapa de amplificación universal para aumentar el número de loci para el análisis. Esta introducción de regiones de cebadores a los productos de amplificación iniciales de una muestra mixta permite una amplificación universal controlada posterior de todos o una parte de los ácidos nucleicos seleccionados antes o durante el análisis, por ejemplo determinación de secuencia.

5 El sesgo y la variabilidad pueden introducirse durante la amplificación del ADN, tal como la observada durante la reacción en cadena de la polimerasa (PCR). En los casos en los que una reacción de amplificación es multiplexada, existe la posibilidad de que los loci se amplifiquen a diferentes velocidades o eficacia. Parte de esto puede deberse a la variedad de cebadores en una reacción múltiplex, teniendo algunos mejor eficacia (es decir, hibridación) que otros, o funcionando algunos mejor en condiciones experimentales específicas debido a la composición de la base. Cada conjunto de cebadores para un locus dado puede comportarse de manera diferente según el contexto de secuencia del cebador y el molde de ADN, las condiciones del tampón y otras condiciones. Una amplificación de ADN universal para un sistema de ensayo multiplexado generalmente introducirá menos sesgo y variabilidad.

15 En consecuencia, en un aspecto, se realiza un pequeño número (por ejemplo 1-10, preferiblemente 3-5) de ciclos de amplificación seleccionada o enriquecimiento con ácido nucleico de la muestra inicial en una reacción de mezcla multiplexada, seguido de amplificación universal usando los cebadores universales introducidos. El número de ciclos que usan 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 la amplificación universal después de un número menor de ciclos de amplificación, se reduce el sesgo de tener determinados loci que se amplifiquen a mayores velocidades que otros.

25 Opcionalmente, el sistema de ensayo incluirá una etapa entre la amplificación seleccionada y la amplificación universal para retirar los ácidos nucleicos en exceso que no se amplifican específicamente en la amplificación seleccionada.

30 El producto completo o una alícuota del producto de la amplificación seleccionada puede usarse para la amplificación universal. Las mismas o diferentes condiciones (por ejemplo, polimerasa, tampones, y similares) pueden usarse en las etapas de amplificación, por ejemplo, para asegurar que el sesgo y la variabilidad no se introducen de manera inadvertida debido a las condiciones experimentales. Además, pueden usarse variaciones en las concentraciones del cebador para limitar de manera efectiva el número de ciclos de amplificación específica de secuencia.

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

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

45 Algunos aspectos utilizan reacciones acopladas para la detección múltiplex de secuencias de ácido nucleico en las que los oligonucleótidos de una fase temprana de cada procedimiento contienen secuencias que pueden ser usadas por oligonucleótidos de una fase posterior del procedimiento. Pueden usarse procedimientos a modo de ejemplo para amplificar y/o detectar ácidos nucleicos en muestras, 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.

50 En determinados aspectos, el sistema de ensayo de la divulgación utiliza una de las siguientes técnicas de amplificación universales y selectivas combinadas: (1) reacción de detección de la ligasa ("LDR") acoplada con reacción en cadena de la polimerasa ("PCR"); (2) PCR primaria acoplada con PCR secundaria acoplada con LDR; y (3) PCR primaria acoplada con PCR secundaria. Cada uno de estos aspectos de la divulgación tiene una aplicabilidad en la detección de determinadas características del ácido nucleico. Sin embargo, cada uno requiere el uso de reacciones acopladas para la detección múltiplex de diferencias de secuencia de ácido nucleico en los que los oligonucleótidos de una fase temprana de cada procedimiento contienen secuencias que pueden ser utilizadas por oligonucleótidos de una fase posterior del procedimiento.

60 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 la reacción en cadena de la ligasa (LCR) para la detección de secuencias específicas de nucleótidos en una variedad de muestras de ácido nucleico.

65 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 detección de ácido nucleico.

5 Barany *et al.*, patentes estadounidenses n.ºs 7.556.924 y 6.858.412, describen el uso de sondas precirculares (también denominadas “sondas candado” o “sondas multiinversión”) con LDR acoplada y reacción en cadena de la polimerasa (“PCR”) para la detección de ácido nucleico.

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

Willis *et al.*, patentes estadounidenses n.ºs 7.700.323 y 6.858.412, describen el uso de sondas precirculares en amplificación, detección y genotipificación de ácido nucleico multiplexada.

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

20 En algunos casos, un ensayo único puede emplear una combinación de los métodos anteriormente descritos. Por ejemplo, algunos de los loci pueden detectarse usando oligonucleótidos de secuencia fija que se unen a regiones complementarias, adyacentes en un locus, mientras que otros loci pueden detectarse usando loci de puente en el mismo ensayo. En otro ejemplo, algunos de los loci pueden detectarse usando oligonucleótidos de secuencia fija que se unen a regiones complementarias, adyacentes en un locus, mientras que otros loci pueden requerir una etapa de extensión de cebador para unir los oligonucleótidos de secuencia fija.

25 En un aspecto preferido, los productos de amplificación se multiplexan, tal como se describió anteriormente. En un aspecto preferido, los productos de amplificación múltiplex se cuantifican por análisis de los productos de amplificación. En un aspecto preferido, una muestra representativa de moléculas individuales de los procedimientos de amplificación se aísla de las otras moléculas para un análisis adicional. Para obtener una muestra representativa de moléculas individuales, el número promedio de moléculas por locus debe exceder el ruido de muestreo creado por la reacción multiplexada. En un aspecto, el número promedio por locus es mayor de 100. En otro aspecto, el número promedio por locus es mayor de 500. En otro aspecto, el número promedio por locus es mayor de 1000.

35 Las moléculas individuales del producto de amplificación se aíslan preferiblemente de manera física de las otras moléculas de una manera que permite que los diferentes productos de amplificación se distingan entre sí en el análisis. En un aspecto preferido, este aislamiento se produce sobre un sustrato sólido. La molécula aislada puede asociarse con una dirección física o identificable particular ya sea antes del análisis, o la dirección puede darse a conocer por los productos de amplificación particulares basados en el resultado del análisis. El sustrato puede ser una superficie plana o una superficie tridimensional tal como un perla.

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

45 En un aspecto preferido, la longitud completa del producto de amplificación o una parte del producto de amplificación puede analizarse utilizando la determinación de secuencia. El número de bases que necesitan determinarse debe ser suficiente para identificar de forma única el producto de amplificación como perteneciente a un locus y/o alelo específico. En un aspecto preferido, el producto de amplificación se analiza a través de la determinación de la secuencia del producto de amplificación seleccionado.

50 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, incluyendo, pero sin limitarse 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, que se incorporan como referencia, secuenciación mediante métodos de síntesis, por ejemplo, Nyren *et al.*, patentes estadounidenses n.ºs 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, 7.459.311, 6.828.100 y 6.210.891 y métodos de determinación de secuenciación basado en ligamiento, 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.

65 La información de la secuencia puede determinarse utilizando métodos que determinan muchas (típicamente de miles a miles de millones) de secuencias de ácido nucleico de una manera intrínsecamente paralela, donde muchas secuencias se leen preferiblemente en paralelo utilizando 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 ligamiento (por ejemplo, tal como se comercializa en la tecnología SOLiD™, Life Technology, Inc., Carlsbad, CA); secuenciación mediante síntesis usando nucleótidos modificados (tal como se comercializa 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 de California, Inc., Menlo Park, CA), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., San Francisco Sur, CA); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain Vista, CA); tecnologías de secuenciación basadas en nanoporos (por ejemplo, tal como se desarrollan por Oxford Nanopore Technologies, LTD, Oxford, UK), y métodos de secuenciación altamente paralelizados similares.

Alternativamente, en otro aspecto, la longitud completa del producto de amplificación o una parte del producto de amplificación puede analizarse usando técnicas de hibridación. Los métodos para realizar ensayos de hibridación de polinucleótidos para detección han sido bien desarrollados en la técnica. Los procedimientos y condiciones del ensayo de hibridación variarán según la aplicación y se seleccionarán de acuerdo con los métodos de unión generales conocidos, incluidos los mencionados en: Maniatis *et al.* Molecular Cloning: A Laboratory Manual (2ª ed. Cold Spring Harbor, N.Y., 1989); Berger y Kimmlos 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). Los métodos y aparatos para llevar a cabo reacciones de hibridación controladas y repetidas se ha descrito 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).

Los métodos y aparatos para la detección de señales y el procesamiento de datos de intensidad se dan a conocer 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 la solicitud de patente estadounidense 60/364.731 y en la solicitud PCT PCT/US99/06097(publicada como el documento WO99/47964).

Minimización de la variación dentro de y entre las muestras

Un desafío con la detección de anomalías cromosómicas en un feto mediante la detección en una muestra mixta es que los ácidos nucleicos de la fuente secundaria están presentes en una abundancia mucho menor que los ácidos nucleicos del tipo de célula normal.

La variación entre los niveles encontrados entre las muestras y/o para los loci dentro de una muestra puede minimizarse con 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 compararla frente a un cromosoma presente en abundancia supuestamente anómala, como la aneuploidía, en la misma muestra. Si bien el uso de uno de tales cromosomas "normales" como cromosoma de referencia puede ser suficiente, también es posible usar dos o más cromosomas normales como cromosomas de referencia internos para aumentar el poder estadístico de la cuantificación.

Un método para usar una referencia interna es calcular una razón de la abundancia de los cromosomas supuestamente anómalos con respecto a la abundancia de los cromosomas normales en una muestra, denominada razón cromosómica. Al calcular la razón cromosómica, la abundancia o los recuentos de cada uno de los loci para cada cromosoma se suman juntos para calcular los recuentos totales para cada cromosoma. Los recuentos totales para un cromosoma se dividen luego entre los recuentos totales de un cromosoma diferente para crear una razón cromosómica para esos dos cromosomas.

Alternativamente, puede calcularse una razón cromosómica para cada cromosoma sumando primero los recuentos de cada uno de los loci para cada cromosoma, y luego dividiendo la suma de un cromosoma entre la suma total de dos o más cromosomas. Una vez calculada, la razón cromosómica se compara con la razón cromosómica promedio de una población normal.

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 utiliza la media. Al desarrollar el conjunto de datos para la razón cromosómica de la población normal, se calcula la variación normal de los cromosomas medidos. Esta variación puede expresarse de varias maneras, más típicamente como el 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 de la muestra cae estadísticamente fuera de la razón cromosómica promedio de 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 aceptables de falsos positivos y falsos negativos para

el ensayo deseado. 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 otro ejemplo, es seis o más veces la variación de la razón cromosómica. En el ejemplo anterior, la razón cromosómica se determina sumando los recuentos de loci por cromosoma. Normalmente, se usa el mismo número de loci seleccionados para cada cromosoma. Un método alternativo para generar la razón cromosómica será calcular los recuentos promedio de los loci 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, como un promedio ajustado o ponderado.

Una vez que se han calculado los recuentos promedio para cada cromosoma, los recuentos promedio para cada cromosoma pueden dividirse entre sí 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 describe anteriormente. Tal como se destacó anteriormente, la capacidad de detectar una aneuploidía en una muestra materna donde el posible ADN tiene una abundancia relativa baja depende en gran medida de la variación en las mediciones de diferentes loci seleccionados en el ensayo. Pueden usarse numerosos métodos analíticos que reducen esta variación y, por tanto, mejoran la sensibilidad de este método para detectar la aneuploidía. Un método para reducir la variabilidad del ensayo es aumentar el número de loci seleccionados utilizados para calcular la abundancia de los cromosomas. En general, si la variación medida de un locus seleccionado único de un cromosoma es $X\%$ e Y los diferentes loci seleccionados se miden en el mismo cromosoma, la variación de la medición de la abundancia cromosómica se calcula sumando o promediando la abundancia de cada locus seleccionado en ese cromosoma se dividirá aproximadamente $X\%$ entre $Y^{1/2}$. Dicho de otra manera, la variación de la medición de la abundancia del cromosoma será aproximadamente la variación promedio de la medición de la abundancia de cada locus seleccionado dividida entre la raíz cuadrada del número de loci.

En un aspecto preferido de esta invención, el número de loci medidos para cada cromosoma es de al menos 24. En otro aspecto preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es de al menos 48. En otro aspecto preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es de al menos 100. En otro aspecto preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es de al menos 200. Existe un coste incremental para medir cada locus y, por tanto, es importante minimizar el número de cada locus seleccionado. En un aspecto preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es menor que 2000. En un aspecto preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es menor que 1000. En un aspecto más preferido de esta invención, el número de loci seleccionados medidos para cada cromosoma es de al menos 48 y menos de 1000. En un aspecto, después de la medición de la abundancia para cada locus seleccionado, puede usarse un subconjunto de los loci seleccionados para determinar la presencia o ausencia de la aneuploidía. Existen muchos métodos habituales para elegir el subconjunto de loci seleccionados. Estos métodos incluyen la exclusión de valores atípicos, en la que los loci 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 más alto medido por la abundancia. En otro aspecto, el percentil puede ser el 10% más bajo y más alto medido por la abundancia. En otro aspecto, el percentil puede ser el 25% más bajo y más alto, medido por la abundancia.

Otro método para elegir el subconjunto de loci seleccionados incluye la eliminación de regiones que quedan fuera de algún límite estadístico. Por ejemplo, los loci seleccionados que quedan fuera de una o más desviaciones estándar de la abundancia media pueden eliminarse del análisis. Otro método para elegir el subconjunto de loci 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 supere la prueba de esperanza. Para minimizar adicionalmente la variación en el ensayo, puede aumentarse el número de veces que se mide cada locus seleccionado. Tal como se ha comentado, 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 intencionalmente cada locus seleccionado múltiples veces. En general, al contar eventos, la variación en el conteo está determinada por las estadísticas de Poisson, y la variación de conteo es típicamente igual a uno dividido entre la raíz cuadrada del número de recuentos. En un aspecto preferido de la invención, los loci seleccionados se miden cada uno en promedio al menos 100 veces. En un aspecto preferido de la invención, los loci seleccionados se miden cada uno en promedio al menos 500 veces. En un aspecto preferido de la invención, los loci seleccionados se miden cada uno en promedio al menos 1000 veces. En un aspecto preferido de la invención, los loci seleccionados se miden cada uno en promedio al menos 2000 veces. En un aspecto preferido de la invención, los loci seleccionados se miden cada uno en promedio al menos 5000 veces.

En otro aspecto, los subconjuntos de loci pueden elegirse de manera aleatoria pero con un número suficiente de loci para producir un resultado estadísticamente significativo para determinar si existe una alteración cromosómica. Pueden realizarse múltiples análisis de diferentes subconjuntos de loci dentro de una muestra mixta para producir más poder estadístico. En este ejemplo, puede o no ser necesario retirar o eliminar cualquier locus antes del análisis aleatorio. Por ejemplo, si hay 100 loci seleccionados para el cromosoma 21 y 100 loci seleccionados para el cromosoma 18, podría realizarse una serie de análisis que evalúen menos de 100 loci para cada uno de los cromosomas.

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 loci para cada muestra se interrogan en una reacción única en un recipiente único. 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

Los sistemas de ensayo de la divulgación permiten la detección de alteraciones específicas de tumores cuantitativas y cualitativas de ADNlc, como la integridad de la cadena de ADN, la frecuencia de mutaciones, las alteraciones de los microsatélites y la metilación de genes, como marcadores de diagnóstico, pronóstico y monitorización en pacientes con cáncer. La capacidad de combinar tal detección de alteraciones de un gen único (incluidas las mutaciones puntuales, los indels y la variación del número de copias) con la detección de la VNC proporciona métodos potentes para ayudar con el diagnóstico clínico, los tratamientos, la predicción de resultados y la 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 divulgación se usa para propósitos de diagnóstico por ejemplo para detectar la presencia y/o la naturaleza de un tumor maligno en un paciente o para proporcionar una estimación cuantitativa de la carga tumoral en un paciente. El ADN tumoral circulante y los microARN se han asociado con determinados cánceres, como el cáncer de pulmón (Roth C *et al.*, *Mol Oncol.* Junio de 2011; 5 (3): 281-91. Documento Epub de 24 de febrero de 2011). Las variaciones en el número de copias también se han detectado en determinados cánceres, como HER2 amplificado y receptor de estrógeno en el ADNlc en pacientes con cáncer de mama. (Page K., *Br J Cancer.* 12 de abril de 2011; 104 (8): 1342-8. Documento Epub de 22 de marzo de 2011).

En otros aspectos de la divulgación, el sistema de ensayo se utiliza en pacientes con cáncer para monitorizar una respuesta al tratamiento y/o seguir el progreso de la enfermedad, por ejemplo para medir alteraciones de un gen único y ADNlc en pacientes que reciben quimiorradioterapia (CRT). Para determinados cánceres, se ha demostrado que el índice de integridad de ADNlc puede asociarse de manera significativa e independiente 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 del 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. Documento Epub de 29 de julio de 2010, que describe variaciones genéticas útiles para la determinación de la respuesta al tratamiento y la supervivencia en el cáncer. Por ejemplo, la detección de los niveles de ADNlc combinada con la detección de mutaciones en el gen K-RAS y/o el gen p53 proporciona una herramienta potente, relativamente no invasiva para medir el pronóstico de diversos cánceres, incluido el cáncer de ovario, el cáncer de endometrio y los linfomas. Dobrzycka B *et al.*, *Ann Oncol.* Mayo de 2011; 22 (5): 1133-40. Documento Epub de 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. Documento Epub de 28 de noviembre de 2008. Este análisis puede ayudarse adicionalmente usando herramientas como Varietas, un portal de base de datos funcional para la identificación de la variación genética y la asociación con los resultados del tratamiento y el pronóstico. Paananen J *et al.*, base de datos (Oxford). 29 de julio de 2010; 2010: baq016.

Uso de sistemas de ensayo para la detección de muestras mixtas de pacientes de trasplante

Los sistemas de ensayo de la divulgación pueden usarse para monitorizar la salud de los órganos en un paciente de trasplante que utiliza una combinación de detección de ADNlc y detección de SNP o mutaciones en uno o más genes únicos. Los órganos trasplantados tienen genomas que son distintos del genoma de un paciente receptor, y la salud del órgano puede detectarse usando un sistema de ensayo. Por ejemplo, se ha demostrado que el rechazo celular agudo se asocia con niveles aumentados de manera significativa de ADN libre de células del genoma del donante en receptores de trasplantes de corazón. Snyder TM *et al.*, *Proc Natl Acad Sci U S A.* 12 de abril de 2011; 108 (15): 6229-34. Documento Epub de 28 de marzo de 2011. Además, las quimiocinas y las moléculas de adhesión median en el rechazo del aloinjerto reclutando leucocitos en el aloinjerto, y los SNP localizados en la interleucina (IL)-8, CXCR1, CXCR2, se ha demostrado que se correlacionan con los resultados del aloinjerto. Ro H. *et al.*, *Transplantation.* 15 de enero de 2011; 91 (1): 57-64. Así, los sistemas de ensayo de la divulgación 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 diagnóstico o pronóstico más complejas.

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 beneficioso para analizar los productos de amplificación, ya que el porcentaje de ADN fetal en la muestra proporciona información importante sobre la presencia estadística esperada de los cromosomas, y la variación de esa expectativa puede ser indicativa de aneuploidía fetal. Esto puede ser especialmente útil en

5 circunstancias en las que el nivel de ADN fetal en una muestra materna es bajo, ya que el porcentaje de contribución fetal puede usarse para determinar la significación estadística cuantitativa en las variaciones de los niveles de loci seleccionados en una muestra materna. En otros aspectos, la determinación del porcentaje relativo de ADN fetal en una muestra materna puede ser beneficiosa para estimar el nivel de certeza o poder en la detección de una aneuploidía fetal.

10 En algunos aspectos específicos, la contribución fetal relativa del ADN materno en el alelo de interés puede compararse con la contribución paterna en ese alelo para determinar la concentración de ADN fetal aproximada en la muestra. En otros aspectos específicos, puede usarse la cantidad relativa de secuencias derivadas únicamente de manera paterna (por ejemplo, secuencias del cromosoma Y o polimorfismos paternos específicos) para determinar la concentración relativa de ADN fetal en una muestra materna.

15 Otro enfoque a modo de ejemplo para determinar el porcentaje de contribución fetal en una muestra materna a través del análisis de fragmentos de ADN con diferentes patrones de metilación del ADN entre el ADN fetal y el materno. En un aspecto preferido, el ADN amplificado del ADN libre de plasma es mediante reacción en cadena de la polimerasa (PCR). También pueden usarse otros mecanismos para la amplificación, incluidos los descritos con más detalle en el presente documento, como será evidente para un experto en la técnica tras leer la presente divulgación.

20 En aspectos particulares, el porcentaje de ADN fetal libre en la muestra materna puede determinarse mediante PCR utilizando ADN diluido en serie aislado de la muestra materna, que puede cuantificar con precisión el número de genomas que comprenden los genes amplificados.

25 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 loci específicos del cromosoma Y y la comparación con el contenido de ADN materno calculado. Las cantidades de un locus específico del cromosoma Y amplificado, tal como una región del gen de la región Y que determina el sexo (SRY), que se encuentra en el cromosoma Y y, por tanto, es representativo del ADN fetal, puede determinarse a partir de la muestra y compararse con uno o más loci seleccionados amplificados que están presentes tanto en el ADN materno como en el ADN fetal y que preferiblemente no son de un cromosoma que se cree que es potencialmente aneuploide en el feto, por ejemplo, una región autosómica que no está en el cromosoma 21 ó 18. Preferiblemente, esta etapa de amplificación se realiza en paralelo con la etapa de amplificación selectiva, aunque puede realizarse antes o después de la amplificación selectiva dependiendo de la naturaleza del ensayo multiplexado.

35 En un aspecto específico, el porcentaje de ADN fetal libre en plasma materno se calcula para un feto masculino usando la siguiente fórmula: porcentaje de ADN fetal libre = $(n.^{\circ} \text{ de copias del gen unido a Y} \times 2 \times 100) / (n.^{\circ} \text{ de copias de gen autosómico})$, en la que 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 utiliza para normalizar el hecho de que solo existe 1 copia del gen unido a Y en comparación con dos copias del gen autosómico en cada genoma, fetal o materno.

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

45 En determinados aspectos de la divulgación, la determinación de la contribución del ADN de una fuente menor puede ser útil para determinar la variación del número de copias de los loci en esas muestras. Por ejemplo, en cada muestra de origen materno, el ADN de un feto tendrá aproximadamente el 50% de sus loci heredados de la madre y el 50% de los loci heredados del padre. Determinar los loci contribuidos al feto a partir de fuentes paternas puede permitir la estimación del ADN fetal en una muestra materna y, por tanto, proporcionar información usada para calcular las diferencias estadísticamente significativas en las frecuencias cromosómicas para los cromosomas de interés.

50 En determinados aspectos, la determinación de polimorfismos de fuente secundaria requiere un análisis de mutaciones y/o SNP dirigido para identificar la presencia del ADN de la fuente secundaria en una muestra mixta. En algunos aspectos, el uso de genotipificación previa es útil, por ejemplo, la genotipificación del donante de un trasplante, la genotipificación del padre y la madre en una muestra materna. Pero, en general, esta información relacionada con la genotipificación previa no es necesaria antes de realizar el ensayo, y preferiblemente la genotipificación se realiza simultáneamente con la determinación del número de copias de los loci seleccionados dentro de una muestra mixta.

60 En un aspecto preferido, el porcentaje de ácidos nucleicos de fuente secundaria en una muestra mixta puede cuantificarse usando la detección de SNP multiplexada sin usar el conocimiento genotípico previo. En este aspecto, se usan dos o más loci polimórficos seleccionados con un SNP conocido en cada región. En un aspecto preferido, los loci polimórficos seleccionados se amplifican. En un aspecto preferido, la amplificación es universal. En una realización preferida, los loci polimórficos seleccionados se amplifican en una reacción en un recipiente. Cada alelo de los loci polimórficos seleccionados en la muestra materna se determina y cuantifica. En un aspecto preferido, la secuenciación de alto rendimiento se utiliza para tal determinación y cuantificación. Los loci se identifican donde los

genotipos de la fuente principal y secundaria son diferentes, por ejemplo, el genotipo del donante es homocigótico y el genotipo del receptor es heterocigótico. 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 loci 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 loci que cumplen con este requisito se usan para determinar la concentración de ácido nucleico de la fuente secundaria a través del análisis estadístico. En un aspecto, la concentración se determina sumando los alelos de baja frecuencia de dos o más loci juntos, dividiendo entre la suma de los alelos de alta frecuencia y multiplicando por dos. En otro aspecto, el porcentaje de ácido nucleico de la fuente secundaria se determina promediando los alelos de baja frecuencia de dos o más loci, dividiendo entre el promedio de los alelos de alta frecuencia y multiplicando por dos.

Para muchos alelos, las secuencias de ácidos nucleicos de la fuente principal y la secundaria menor pueden ser homocigóticas e idénticas, y como esta información no lo distingue, no es útil en la determinación de ácido nucleico de la fuente secundaria en una muestra mixta. La presente invención utiliza información alélica en la que existe una diferencia distinguible entre las fuentes celulares (por ejemplo, un alelo fetal que contiene al menos un alelo que difiere del alelo materno) en los cálculos de porcentajes de ácido nucleico de la fuente secundaria. Los datos pertenecientes a regiones alélicas que son las mismas para la fuente principal y la fuente secundaria no se seleccionan para el análisis, o se eliminan de los datos pertinentes antes de la determinación del porcentaje para no saturar los datos útiles.

Métodos a modo de ejemplo para la cuantificación de ADN fetal en plasma materno pueden encontrarse, por ejemplo, en Chu *et al.*, Prenat Diagn 2010; 30:1226-1229.

En un aspecto, los loci seleccionados pueden excluirse si la cantidad o la frecuencia de la región parece ser un valor atípico debido a un error experimental, o por un sesgo genético idiopático dentro de una muestra particular. En otro aspecto, los ácidos nucleicos seleccionados pueden someterse a ajustes estadísticos o matemáticos, tales como normalización, estandarización, agrupación o transformación antes de la suma o la determinación del promedio. En otro aspecto, los ácidos nucleicos seleccionados pueden experimentar tanto la normalización como la exclusión de errores experimentales de datos antes de la suma o determinación del promedio.

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

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

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 contenga una aneuploidía. En un aspecto, se usan métodos de detección de aneuploidía que utilizan análisis de segmentos de ADN aleatorios, tal como el que se describe 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 loci seleccionados en una muestra mixta incluyen ambas regiones para la determinación del contenido de ADN de la fuente secundaria así como regiones no polimórficas de dos o más cromosomas para detectar una alteración cromosómica en una reacción única. La reacción única ayuda a minimizar el riesgo de contaminación o sesgo que puede introducirse durante diversas etapas en el sistema de ensayo que, de lo contrario, pueden distorsionar los resultados cuando se utiliza un contenido de ADN de la fuente secundaria para ayudar a determinar la presencia o ausencia de una alteración cromosómica. En otros aspectos, un locus o regiones seleccionadas pueden utilizarse tanto para la determinación del contenido de ADN de la fuente secundaria como para la detección de alteraciones cromosómicas de la fuente menor secundaria. La utilización de las mismas regiones tanto para el contenido de ADN como para la detección de alteraciones cromosómicas puede ayudar adicionalmente a minimizar cualquier sesgo debido a un error experimental o contaminación.

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la técnica una divulgación y descripción completas de cómo hacer y usar la presente invención, y no se pretende limitar el alcance de lo que los inventores consideran como su invención, tampoco se pretende representar o implicar que los experimentos a continuación son todos o los únicos experimentos realizados. Los expertos en la materia apreciarán que pueden realizarse numerosas

variaciones y/o modificaciones a la invención tal como se muestra en los aspectos específicos sin apartarse del alcance de la invención tal como se describe y define ampliamente en las reivindicaciones adjuntas. Por tanto, los aspectos actuales deben considerarse en todos los aspectos como ilustrativos y no restrictivos.

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

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

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

Estos ensayos se diseñaron basándose en secuencias genómicas humanas, y cada interrogación consistió en dos oligos de secuencia fija por locus seleccionado interrogado en el ensayo. El primer oligo, complementario a la región 3' de una región genómica, comprendía los siguientes elementos de oligo secuenciales (de 5' a 3'): una secuencia de cebado de PCR universal común a todos los ensayos: TACACGGCGTTATGCGTCGAGAC (SEQ ID NO:1); un índice de identificación de nueve nucleótidos específico para el locus seleccionado; una secuencia específica de alelo/locus o locus de 9 bases que actúa como un índice de locus en el primer conjunto independiente de SNP y un índice de locus/alelo en el segundo conjunto específico de polimorfismo; un nucleótido de ruptura 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 los casos en que se detectó un SNP o una mutación en esta parte del locus genómico seleccionado, el conjunto de interrogación específico de alelo consistió en dos primeros cebadores de ligamiento 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 oligos se diseñaron para cada ácido nucleico seleccionado para proporcionar una T_m uniforme predicha con una variación de dos grados a través de todas las interrogaciones en el conjunto de ensayo.

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

En los casos en los que se detectó un SNP o una mutación en el locus genómico seleccionado, el conjunto de interrogación específico de alelo consistió en dos cebadores de ligamiento en tándem, cada uno con un índice de locus/alelo diferente y una base específica de alelo diferente en la posición de mutación/SNP. Se diseñó este segundo oligo de secuencia fija para cada ácido nucleico seleccionado para proporcionar una T_m 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_m como el primer conjunto de oligos.

En determinados aspectos probados, se usaron uno o más oligos de puente que eran complementarios a la secuencia del locus genómico entre la región complementaria al primer y el segundo oligos de secuencia fija usados para cada locus seleccionado. En los aspectos específicos probados, se usó más de un oligo de puente para cubrir el hueco entre los oligonucleótidos de secuencia fija, y puede diseñarse el uno o más oligo de puente opcionalmente para identificar una o más mutaciones o SNP en la secuencia. La longitud de los oligonucleótidos de puente utilizados en los sistemas de ensayo varió entre 5 y 36 pares de bases.

Se sintetizaron todos los oligonucleótidos utilizados en los formatos de ligamiento en tándem usando la química de fase sólida convencional. Se sintetizaron los segundos oligos de secuencia fija y los oligonucleótidos de puente con restos fosfato 5' para permitir el ligamiento a los extremos terminales hidroxilo 3' de oligonucleótidos adyacentes.

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

Se obtuvo el ADN genómico de un hombre caucásico (NA12801) o una mujer caucásica (NA11995) de Coriell Cell Repositories (Camden, Nueva Jersey) y se fragmentó mediante cizalladura acústica (Covaris, Woburn, MA) a un tamaño de fragmento medio de aproximadamente 200 pb.

Se sometió a biotilación el ADN de Coriell usando procedimientos habituales. Brevemente, se reparó en el extremo el ADN fragmentado de Covaris generando la siguiente reacción en un microtubo de 1,5 ml: 5 µg de ADN, 12 µl de tampón de ligasa T4 10X (Enzymatics, Beverly MA), 50 U de polinucleótido cinasa T4 (Enzymatics, Beverly MA) y H₂O hasta 120 µl. Se incubó esto a 37°C durante 30 minutos. El ADN se diluyó utilizando Tris 1mM EDTA 10 mM, pH 8,5, hasta la concentración final deseada de ~2 ng/µl.

Se colocaron 5 µl de ADN en cada pocillo de una placa de 96 pocillos, y se selló la placa con un sellador de placa adhesivo y se centrifugó durante 10 segundos a 250 x g. Luego se incubó la placa a 95°C durante 3 minutos, se enfrió hasta 25°C y se volvió a centrifugar durante 10 segundos a 250 x g. Se preparó una mezcla maestra de biotilación en un microtubo de 1,5 ml hasta la concentración final de: tampón IX TdT (Enzymatics, Beverly, MA), 8 U de TdT (Enzymatics, Beverly, MA), CoCl₂ 250 µM, 0.01 nmol/µl de biotina-16-dUTP (Roche, Nutley NJ) y H₂O hasta 1,5 ml. Se tomaron alícuotas de la mezcla maestra de 15 µl en cada pocillo de una placa de 96 pocillos, y se selló la placa con un sellador de placa adhesivo. Se centrifugó la placa durante 10 segundos a 250 x g y se incubó a 37°C durante 60 minutos. Tras la incubación, se centrifugó de nuevo la placa durante 10 segundos a 250 x g y se añadieron a cada pocillo 7,5 µl de mezcla de precipitación (1 µg/µl de Dextran Blue, NaOAC 3 mM).

Se selló la placa con un sellador de placa adhesivo y se mezcló usando un agitador tipo vórtex de placas IKA durante 2 minutos a 3000 rpm. Se añadieron 27,5 µl de isopropanol a cada pocillo, se selló la placa con un sellador de placa 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 en un pañuelo de celulosa de laboratorio absorbente. Se secó la placa al aire durante 5 minutos y se volvió a suspender el sedimento en 30 µl de Tris 10 mM, pH 8,0, EDTA 1 mM.

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

Se sometieron a prueba numerosos formatos de ensayo de ligamiento en tándem que utilizan el ADN biotilado para ilustrar la prueba de concepto para los sistemas de ensayo de la invención, y demostraron la capacidad para realizar la detección dirigida, altamente multiplexada, de un gran número de loci independientes usando la serie de diferentes formatos de ensayo. Los sistemas de ensayo a modo de ejemplo de la invención se diseñaron para comprender 96 o más interrogaciones por loci en una muestra genética, y en los casos en los que se detectaron SNP, los formatos de ensayo utilizaron 192 o más interrogaciones separadas, cada una utilizando la detección de diferentes alelos por 96 loci en muestras genéticas. Los ejemplos descritos para cada formato de ensayo utilizaron dos conjuntos diferentes de oligonucleótidos de secuencia fija y/o oligos de puente (tal como se describe en el ejemplo 1), que comprenden un total de 96 o 192 reacciones de interrogación para los loci seleccionados dependiendo de si se identificaron o no los SNP.

Un primer formato de ensayo a modo de ejemplo usó oligos de secuencia fija específicos de locus y oligos de puente, en los que existía un hueco de una base entre el primer oligo de secuencia fija y los oligos de puente, y un segundo hueco de una base entre los oligos puente y el segundo oligo de secuencia fija. Cada uno de los dos huecos abarcaba dos SNP diferentes. En este formato, se usó una ADN polimerasa para incorporar cada una de las bases de SNP, y se usó la ligasa para sellar las mellas formadas de ese modo. La discriminación de la base de SNP derivó de la fidelidad de la incorporación de la base por la polimerasa, y en el caso de una incorporación errónea, la tendencia de la ligasa a no sellar mellas adyacentes a las bases que presentan apareamiento erróneo.

El segundo formato de ensayo a modo de ejemplo usó dos oligonucleótidos de secuencia fija específicos del locus sin un oligo puente, donde existía un hueco de base de ~15-35 entre los oligos de secuencia fija, y en el que el hueco se extendía sobre 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 este modo. La discriminación de la base de SNP derivó de la fidelidad de la incorporación de la base mediante la polimerasa, y en el caso de una incorporación errónea, la tendencia de la ligasa a no sellar las mellas adyacentes a las bases que presentan apareamiento erróneo.

Un tercer formato de ensayo a modo de ejemplo usó oligos de secuencia fija primero y segundo específicos de alelo sin un oligo de puente, donde había un hueco de base de ~15-35 entre el primer y el segundo oligos de secuencia fija, y en el hueco se extendía sobre uno o más SNP. Se usaron dos oligos de secuencia fija primero específicos de alelo separados y dos oligos de secuencia fija segundo específicos de alelo separados. 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 la base de SNP derivó de la especificidad de hibridación, la tendencia de la polimerasa a la no corrección de lectura para no extender los cebadores apareados con apareamientos erróneos cerca del extremo 3', y la tendencia de la ligasa a no sellar las mellas adyacentes a las bases que presentan apareamiento erróneo.

Un cuarto formato a modo de ejemplo usó oligos de secuencia fija específicos de alelo y un oligo de puente específico de locus. En este formato, dos oligos de secuencia fija separados complementarios al extremo 3' del loci de interés, el primero con una base 3' específica para un alelo del SNP objetivo, y el segundo con una base 3' específica para el otro alelo del SNP objetivo. De manera similar, se usaron dos oligos de secuencia fija separados, el primero con una base 5' específica para un alelo de un segundo SNP objetivo, y el segundo con una base 5' específica para el otro alelo del segundo SNP objetivo. Los oligos de puente eran complementarios a la región directamente adyacente a las regiones del locus complementarias al primer y segundo oligos de secuencia fija, y por tanto no se necesitó polimerasa antes del ligamiento. Se usó la ligasa para sellar las mellas entre los oligos de secuencia fija y el oligo de puente. La discriminación de la base de SNP en este formato de ensayo derivó de la

especificidad de hibridación y la tendencia de la ligasa a no sellar las mellas adyacentes a las bases que presentan apareamiento erróneo. Se sometió a prueba este formato a modo de ejemplo utilizando la ligasa T4 o la ligasa Taq para la creación del molde contiguo, y ambas demostraron ser eficaces en la reacción tal como se describe a continuación.

5 Un quinto formato a modo de ejemplo usó oligos 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 por la hibridación de estos oligos. En este formato, no se requirió polimerasa, y se usó una ligasa para sellar la mella única entre los oligos.

10 Un sexto formato a modo de ejemplo usó oligos de secuencia fija específicos de alelo y oligos de puente específicos de locus, en los que había un corto hueco de bases de cinco bases entre la región de loci complementaria a los oligos de secuencia fija. El oligo de puente específico para el locus en este ejemplo fue un mero 5 complementario a las regiones directamente adyacentes a las regiones complementarias al primer y segundo oligos de secuencia fija. En este formato, no se requirió polimerasa, y se usó una ligasa para sellar las dos mellas entre los oligos.

15 Un séptimo formato a modo de ejemplo usó oligos de secuencia fija específicos de locus y un oligo de puente específico de locus, donde había un hueco de bases más corto de cinco bases que contenía un SNP en la región complementaria al oligo de puente. Se incluyeron los oligos de puente específicos de alelo que correspondían a los posibles SNP en la reacción de hibridación y ligamiento. En este formato, no se requirió polimerasa, y se usó una ligasa para sellar las dos mellas entre los oligos. La discriminación de la base de SNP en este formato de ensayo derivó de la especificidad de hibridación y la tendencia de la ligasa a no sellar las mellas adyacentes a las bases que presentan apareamiento erróneo.

20 Un octavo formato a modo de ejemplo usó oligos de secuencia fija específicos de locus y dos oligos de puente específicos de locus adyacentes, donde había un hueco de 10 bases entre las regiones complementarias al primer y el segundo oligos de secuencia fija. Se incluyeron los oligos de puente específicos de locus en la reacción de ligamiento, con el hueco que requiere dos meros 5 contiguos para cerrar el hueco. En este formato, no se requirió polimerasa, y se usó una ligasa para sellar las tres mellas entre los oligos.

25 Para cada uno de los formatos de ensayo descritos anteriormente, se creó una reserva equimolar (40 nM cada uno) de conjuntos del primer y el segundo oligonucleótidos de secuencia fija específicos de loci o alelo a partir de los oligos preparados tal como se expone en el ejemplo 2. Se creó de la misma forma una reserva equimolar separada (20 μ M cada uno) de oligonucleótidos de puente para los procedimientos de ensayo basados en las secuencias de los loci genómicos seleccionados.

30 Se transfirieron 100 μ g de perlas de estreptavidina a los pocillos de una placa de 96 pocillos y se retiró el sobrenadante. Se añadieron 60 μ l de tampón BB2 (Tris 100 mM, pH 8,0, EDTA 10 mM, NaCl₂ 500 mM, formamida al 58%, Tween-80 al 0,17%), 10 μ l de reserva de oligo de secuencia fija 40 nM y 30 μ l del ADN de molde biotinilado preparado en el ejemplo 2 a las perlas. Se selló la placa con un sellador de placa adhesivo y se agitó con vórtex a 3000 rpm hasta que se volvieron a suspender las perlas. Se aparearon los oligos al ADN de molde mediante incubación a 70°C durante 5 minutos, seguido de un enfriamiento lento hasta temperatura ambiente

35 Se colocó la placa en una placa magnética de barra elevada durante 2 minutos para atraer las perlas magnéticas y el ADN asociado al lado de los pocillos. Se eliminó el sobrenadante pipeteando y se reemplazó con 50 μ l de BB2 al 60% (v/v en agua). Se volvieron a suspender las perlas mediante agitación con vórtex, se colocaron de nuevo en el imán y se retiró el sobrenadante. Se repitió este procedimiento de lavado de perlas una vez usando 50 μ l de BB2 al 60% y se repitió dos veces más con 50 μ l de tampón de lavado (Tris 10 mM, pH 8,0, EDTA 1 mM, NaCl₂ 50 mM).

40 Se volvieron a suspender las perlas en 37 μ l de una mezcla de reacción de ligamiento que consistía en 1x tampón de ligasa Taq (Enzymatics, Beverly, MA), 1U de ligasa Taq y reserva de puente de oligos 2 μ M (dependiendo del formato del ensayo), y se incubaron a 37°C durante una hora. Cuando fue apropiado, y dependiendo del formato del ensayo, se incluyó en esta mezcla una polimerasa termoestable de no corrección de lectura más 200 nM de cada dNTP. Se colocó la placa en una placa magnética de barra elevada durante 2 minutos para atraer las perlas magnéticas y el ADN asociado al lado de los pocillos. Se retiró el sobrenadante pipeteando y se reemplazó con 50 μ l de tampón de lavado. Se volvieron a suspender las perlas agitando con vórtex, se colocaron de nuevo en el imán y se retiró el sobrenadante. Se repitió una vez el procedimiento de lavado.

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

Ejemplo 4: Amplificación universal de productos ligados en tándem

65

Se amplificaron los ácidos nucleicos polimerizados y/o ligados utilizando cebadores de PCR universales complementarios a las secuencias universales presentes en el primer y segundo oligos de secuencia fija hibridados a los loci 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 que consistía en 25 µl de producto de ligamiento eluido más tampón 1X Pfusion, betaína 1 M, 400 nM de cada dNTP, 1 U de ADN polimerasa termoestable correctora de errores Pfusion (Thermo Fisher, Waltham MA) y los siguientes pares de cebadores:

TAATGATACGGCGACCACCGAGATCTACACGGCGTTATGCGTCGAGA (SEQ ID NO:3) y

10 TCAAGCAGAAGACGGCATAACGAGATXAAACGACGCGATCATCGGTCC CCGCAA (SEQ ID NO:4), en la que X representa uno de los 96 índices de muestra diferentes utilizados para identificar de forma única muestras individuales antes de la agrupación y secuenciación. Se llevó a cabo la PCR en condiciones rigurosas utilizando un termociclador BioRad Tetrad™.

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

Ejemplo 5: Detección y análisis de loci seleccionados

20 Se secuenciaron los productos de PCR purificados de cada formato de ensayo en un solo carril de un portaobjetos en un Illumina HiSeq™ 2000 (Illumina, San Diego, CA). Las series de secuenciación suelen dar lugar a lecturas en bruto de ~100 M, de las cuales ~85 M (85%) se mapean para las estructuras de ensayo esperadas. Esto se tradujo en un promedio de ~885 K lecturas/muestra en todo el experimento, y (en el caso de un experimento que usa 96 loci), 9,2 K lecturas/repeticón/locus en 96 loci. Se analizaron las lecturas mapeadas en recuentos de repetición/locus/alelo, y se calcularon diversas métricas para cada condición, incluyendo:

Rendimiento: una métrica de la proporción de ADN de entrada que se consultó en la secuenciación, calculada como el número promedio de lecturas únicas por locus (solo contando lecturas de índices de identificación únicas por repetición/locus) dividido entre el número total de equivalentes genómicos contenidos en el ADN de entrada.

35 Intervalo de frecuencia de locus del percentil 80: una métrica de la variabilidad de frecuencia de locus en los datos de secuenciación, se interpretó como el intervalo de plegado que abarca el 80% de los loci. Se calculó sobre la distribución de lecturas totales por locus, en todos los loci, como el percentil 90º de lecturas totales por locus dividido entre el percentil 10º de las lecturas totales por locus.

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

40 Estos resultados se resumen en la Tabla 1:

Tabla 1: Sumario de resultados de los formatos de ensayo de ligamiento en tándem

FORMATO DE ENSAYO	OLIGO DE SECUENCIA FIJA (1º y/o 2º)	OLIGO DE 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	lig Taq	5,0%	5,9	0,92%
4	ESPECÍFICO DE ALELO	Específico de locus	lig T4	5,3%	4,4	0,95%
5	ESPECÍFICO DE LOCUS	No	lig Taq	22,5%	1,7	N/A
6	ESPECÍFICO DE LOCUS	Específico de locus	lig Taq	12,5	2,9	N/A
7	ESPECÍFICO DE LOCUS	Específico de alelo	lig Taq	14,3	2,8	0,20%
8	ESPECÍFICO DE LOCUS	2 específicos	lig Taq	18,5%	2,8	N/A

		de locus				
--	--	----------	--	--	--	--

La tabla 1 indica que el ensayo de ligamiento en tándem específico del locus usando un ADN de molde convertido en oligo de puente en un producto dirigido con alto rendimiento (-10%), con una alta proporción de producto derivado de loci objetivos (el 15% de las lecturas no contenía las estructuras de ensayo esperadas), con un sesgo de locus limitado (el 80% de los loci se encuentran dentro de un intervalo de concentración de ~5 veces) y con alta precisión de SNP (tasa de error de SNP del 0,2%). El ensayo de ligamiento en tándem específico del locus sin el uso de un oligo de puente produjo rendimientos reducidos y sesgos sustanciales del locus, pero aún así produjo datos de genotipado de SNP de alta precisión. El ensayo de ligamiento en tándem específico de alelo con un oligo de puente produjo rendimientos intermedios en comparación con el ensayo de locus específico que usa tanto la ligasa T4 como la Taq, pero aún produjo un sesgo de locus limitado y datos de genotipado de SNP de alta precisión. El ensayo de ligamiento en tándem específico de alelo sin un puente produjo rendimientos reducidos y sesgo de locus sustancial, pero aún así produjo datos de genotipado de SNP de alta precisión.

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

Ejemplo 6: Detección de aneuploidía en muestras del paciente de sujetos embarazados

Los sistemas de ensayo de la invención se usaron en la detección de polimorfismos y alteraciones cromosómicas en dos cohortes separadas de hembras preñadas. 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. Las aneuploidías cromosómicas se detectaron mediante 576 ensayos de los cromosomas 21 y 576 de los cromosomas 18, se agruparon y se analizaron en una sola reacción, tal como se indica a continuación.

Los elementos usados en los ensayos de detección de aneuploidía se ilustran en la figura 6. El ADNlc 601 aislado de muestras maternas se usó como molde para la hibridación, el ligamiento y la amplificación de múltiples loci seleccionados tanto del cromosoma 21 como del cromosoma 18 en cada muestra materna. Se hibridaron tres oligonucleótidos a cada locus seleccionado para crear productos de ligamiento para amplificación y detección. El oligonucleótido de secuencia fija izquierda (o primera) comprendía una región complementaria a un locus 609 seleccionado y una primera región 611 de cebador universal. El oligonucleótido 605 de secuencia fija derecha (o segunda) comprendía una segunda región complementaria al locus 613 seleccionado y una segunda región 615 de cebador universal. Los oligonucleótidos 607 de puente usados se diseñaron de modo que cada uno se hibridara con las regiones de puente de dos o más loci seleccionados usados en el ensayo de detección de aneuploidía. Cuando los oligonucleótidos 603, 605 de secuencia fija y el oligonucleótido 607 de puente se hibridaron con la región complementaria en el ADNlc 601, sus extremos formaron dos mellas. Tras el ligamiento de los oligonucleótidos hibridados con el ADNlc, se creó un producto de ligamiento para cada locus seleccionado que comprende 603, 605 y 607 que se usó como molde para los cebadores 619, 621 de amplificación.

Luego se usaron dos cebadores 619, 621 de amplificación que comprenden regiones complementarias a las regiones primera y segunda de cebador universal, respectivamente, para amplificar el producto de ligamiento. Este producto de amplificación comprendía la secuencia del locus seleccionado. El cebador de amplificación derecho también comprendía un índice 617 de muestra para identificar la muestra particular a partir de la cual se obtuvo el locus en el ensayo multiplexado. La amplificación con 96 cebadores 629 de amplificación derecho diferentes permitió la combinación y la secuenciación simultánea de 96 productos de amplificación diferentes en un solo carril.

Los cebadores 619, 621 de amplificación también contenían una secuencia 623 de agrupación izquierda (TAATGATACGGCGACCACCGA) (SEQ ID NO: 7) y una secuencia 625 de agrupación derecha (ATCTCGTATGCCGTTCTGCTTGA) (SEQ ID NO: 8) que soportaba la amplificación de agrupación para la secuenciación usando el sistema el Illumina HiSeq™ 2000 (Illumina, San Diego, CA). Se usó un cebador 627 de secuenciación que comprende la primera secuencia de cebador universal para determinar la secuencia del producto de amplificación, y se usó un segundo cebador 629 de secuenciación para determinar el índice 617 de muestra del producto de amplificación.

En resumen, se recogieron aproximadamente 10 ml de sangre periférica de cada paciente en un tubo BCT (Streck, Omaha, NE), que se envió por mensajería durante la noche a Tandem Diagnostics. Se aisló plasma de tubos BCT en el plazo de 72 h de la recogida de sangre mediante centrifugación a 1600 g durante 10 m. El plasma se transfirió a un segundo tubo y se centrifugó a 16000 g durante 10 m para eliminar las células restantes. 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 realizó en lotes

multiplexados de hasta 96 muestras de pacientes de ADNlc por método de sistema de matriz.

El ADNlc aislado de las muestras maternas en cada pocillo se precipitó con biotina y se resuspendió en 30 μ l de TE como en el ejemplo 3 anterior. El ADN de molde biotinilado se mezcló con 100 μ g de perlas magnéticas recubiertas con estreptavidina MyOneC1 (Life Technologies, Carlsbad, CA), 60 μ l de tampón BB2 (Tris 100 mM, pH 8,0, EDTA 10 mM, NaCl₂ 500 mM), formamida al 58%, Tween-80 al 0,17%) y 10 μ l de oligonucleótidos de secuencia fija 605 derecha y 603 izquierda 40 nM agrupados. La mezcla se calentó hasta 70°C y se enfrió durante 2 horas. Luego, las perlas se inmovilizaron magnéticamente en el lado del pocillo, se lavaron dos veces con 50 μ l de BB2 al 60% (v/v con H₂O), se lavaron dos veces más con 50 μ l de tampón de lavado (Tris 10 mM, pH 8,0, EDTA 1 mM, NaCl₂ 50 mM), y luego se resuspendió en una reacción de 50 μ l que contenía 1U ligasa Taq (Enzymatics, Beverly MA), tampón ligasa IX Taq (Enzymatics) y 10 μ M de un oligonucleótido 607 de puente de 5 meros 5'-fosforilado. La mezcla se incubó a 37°C durante 1 hora. Las perlas se inmovilizaron magnéticamente de nuevo en el lado del pocillo, se lavaron dos veces con 50 μ l de tampón de lavado y luego se resuspendieron en 30 μ l de TE.

Los productos de ligamiento se eluyeron de las perlas inmovilizadas mediante incubación a 95°C durante 3 minutos. Los productos de ligamiento eluidos se amplificaron mediante 26 ciclos de PCR en una reacción de 50 μ l que contenía 1U Pfusion (Finnzymes), betaína 1 M, tampón IX Pfusion y cebadores de amplificación izquierdo y derecho 400 nM (619, 621 respectivamente). El cebador derecho contenía un índice de muestra de 7 bases (617) que permitía la secuenciación multiplexada de 96 muestras en el sistema HiSeq2000 (Illumina, San Diego, CA). La secuencia del oligo de secuencia fija izquierda fue:

TAATGATACGGCGACCACCGAGATCTACACCGGCGTTATGCGTCGAGA

C

(SEQ ID NO: 5)

Y la secuencia del oligo de secuencia fija derecha fue:

TCAAGCAGAAGACGGCATAACGATNNNNNNNAAACGACGCGATCATCG

GTCCCCGCAAT (SEQ ID NO:6)

Los productos de amplificación de una sola placa de 96 pocillos se agruparon en igual volumen, y los productos de amplificación agrupados se purificaron con perlas SPRI AMPureXP™ (Beckman-Coulter, Danvers, MA) según las instrucciones del fabricante. Cada biblioteca agrupada purificada se usó como molde para la amplificación de agrupamiento en una celda de flujo del kit de agrupación Illumina TruSeq v2 SR (Illumina, San Diego, CA) según los protocolos del fabricante. El portaobjetos se procesó en un sistema Illumina HiSeq™ 2000 (Illumina, San Diego, CA) para producir 56 bases de secuencia específica de locus a partir de un cebador 623 de secuencia izquierdo y se obtuvo una lectura separada de 8 bases de secuencia específica de muestra del cebador 625 de secuencia derecho. Se recogió un promedio de 903K lecturas en bruto por muestra. Se asignó un promedio de 876K (97%) de las lecturas a las estructuras de ensayo esperadas.

La figura 7 muestra datos ejemplares para un subconjunto de las muestras del paciente de la segunda cohorte, que se analizaron en un solo ensayo multiplexado en un solo carril de una ejecución de secuenciación. Inicialmente, se ejecutaron 96 muestras diferentes en esta ejecución en particular, pero más tarde se excluyeron seis muestras de este conjunto analítico porque no cumplían con los umbrales de control de calidad de la muestra.

Se calculó una media recortada para cada cromosoma 18 y cromosoma 21 para las muestras en base a las lecturas producidas en el ensayo. La media recortada se calculó eliminando el 10% de los recuentos altos y bajos para cada cromosoma por muestra. Los productos de amplificación detectados correspondientes a los diversos loci seleccionados se usaron para calcular una métrica de proporción del cromosoma 21 y una métrica de proporción del cromosoma 18 para cada muestra. Para la proporción del cromosoma 21, esto se calculó como la media recortada de los recuentos en los 384 locus seleccionados del cromosoma 21 dividida entre la suma de los recuentos medios recortados para los 576 locus del cromosoma 21 y 576 locus del cromosoma 18 para cada muestra.

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 de la segunda cohorte. Estos recuentos se usaron para calcular las puntuaciones z de la proporción de cromosomas para el cromosoma 21 y el cromosoma 18.

Brevemente, las puntuaciones z se calcularon escalando la mediana por recuento de locus a un valor común (por ejemplo, 1000) para cada muestra, y los recuentos a escala se transformaron por la base logarítmica 2. Se realizó un modelo lineal logarítmico de RMA y un pulido de la mediana (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 cromosómicos, los efectos de locus, los efectos de muestra y los residuos. Los efectos cromosómicos estimados se ajustaron a un valor común, por ejemplo, 0, y se calculó $2^{(\text{efecto cromosómico} + \text{efecto de muestra} + \text{residuo})}$ para cada locus para crear recuentos normalizados. Las puntuaciones Z se escalaron usando una censura iterativa para que tuvieran una media de 0 y una desviación estándar de 1.

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

Los datos proporcionados en la figura 7 se combinaron con los datos de las muestras restantes de la segunda cohorte para determinar las puntuaciones z para el cromosoma 21 y el cromosoma 18, se ilustran en las figuras 10 y 11, respectivamente. Las muestras normales se muestran como rombos de color gris oscuro, y las muestras con trisomía se muestran como rombos de color gris claro. 125/125 muestras normales tenían puntuaciones $z < 3$, 36/36 (100%) muestras T21 y 8/8 (100%) muestras T18 tenían puntuaciones $z > 3$. La puntuación z de muestras T18 media fue de 9,5 y el intervalo fue de 5,1-19,8. La puntuación z de T21 media fue de 11,4 y el intervalo fue de 3,4-21,8.

Además de la detección de aneuploidía en estas cohortes, también se determinaron polimorfismos específicos para estas muestras en un mismo ensayo. Se obtuvo información específica para los loci individuales, así como información polimórfica más general, como el número de locus en los que el locus fetal mostraba un polimorfismo de nucleótido único en un alelo diferente de los polimorfismos de nucleótido único en el locus materno (figura 7, n.º Locus Poli. dif.). Esta determinación también identificó la presencia de polimorfismos específicos en el genoma fetal. Por ejemplo, el estado de tres polimorfismos ejemplares se determinó usando una combinación de oligos de puente que se diseñaron para unirse a tanto el residuo A como T en las siguientes regiones polimórficas a modo de ejemplo:

Tabla 2: Polimorfismos individuales consultados usando el ensayo de invención

Cromosoma	Ubicación	RSID
Ch01	01_010303942	rs11582123
TTTACATGTCTTTGGGCATTTAGGT[A/T]GAGTGAAATCTAGGCCTTGCAAATC (SEQ ID NO:7)		
Ch03	03_098690592	rs2470750
TTGTGTAACGTTAACCTCAGGGACCA[A/T]GAGATGTACTTAGTATTAATTTGCC (SEQ ID NO:8)		
Ch04	04_055495793	rs6815910
GGAAGAAGTGCAGTGTAGTAGACAAC[A/T]CTGGCATTGTGTTTTGTGAAGTGGG (SEQ ID NO:9)		

Tabla 3: Estado materno y fetal predicho para SNP rs11582123.

Muestra	SNP	Recuentos		Estado fetal predicho	Estado materno predicho
		A	T		
1	rs11582123	294	26	A/T	A/A
2		181	134	A/A	A/T
4		34	330	A/T	T/T
5		241	21	A/T	A/A
6		166	134	A/T	A/T
7		137	182	T/T	A/T
8		199	135	A/A	A/T
9		0	267	T/T	T/T
10		0	284	T/T	T/T
11		151	154	A/T	A/T
12		294	1	A/T	A/A
13		131	114	A/A	A/T

ES 2 718 111 T3

14	118	159	T/T	A/T
15	257	10	A/T	A/A
16	309	31	A/T	A/A
17	20	289	A/T	T/A
18	137	166	T/T	A/T
19	138	143	A/T	A/T
20	24	242	A/T	T/T
21	140	161	A/T	A/T
22	159	118	A/A	A/T
23	119	122	A/T	A/T
24	0	250	T/T	T/T
25	0	285	T/T	T/T
26	120	130	A/T	A/T
28	134	113	A/A	A/T
29	109	118	A/T	A/T
30	0	271	T/T	T/T
31	148	139	A/T	A/T
32	29	253	A/T	T/T
33	0	304	T/T	T/T
34	0	278	T/T	T/T
35	103	188	T/T	A/T
36	18	269	A/T	T/T
37	279	34	A/T	A/A
38	0	250	T/T	T/T
39	0	263	T/T	T/T
40	136	142	A/T	A/T
41	147	145	A/T	A/T
42	15	270	A/T	T/T
43	44	222	A/T	T/T
44	140	159	T/T	A/T
45	0	259	T/T	T/T
46	1	304	T/T	T/T
47	162	127	A/A	A/T
48	0	335	T/T	T/T
49	1	247	T/T	T/T
50	153	154	A/T	A/T
51	118	182	T/T	A/T
52	145	134	A/T	A/T
53	146	132	A/T	A/T
54	7	319	A/T	T/T
55	152	174	T/T	A/T
56	1	319	T/T	T/T
57	147	150	A/T	A/T
58	136	157	T/T	A/T
59	83	162	A/A	T/T
60	14	215	A/T	T/T

ES 2 718 111 T3

61	157	121	A/A	A/T
62	281	0	A/A	A/A
63	0	260	T/T	T/T
64	0	305	T/T	T/T
65	18	252	A/T	T/T
66	0	303	T/T	T/T
67	99	161	T/T	A/T
68	141	127	A/T	A/T
69	0	237	T/T	T/T
70	0	315	T/T	T/T
71	132	139	A/T	A/T
73	112	120	A/T	A/T
75	1	268	T/T	T/T
76	166	123	A/A	A/T
78	0	245	T/T	T/T
79	12	264	A/T	T/T
80	15	281	A/T	T/T
81	21	269	A/T	T/T
82	108	160	T/T	A/T
83	106	144	T/T	A/T
84	137	135	A/T	A/T
85	115	151	T/T	A/T
86	0	262	T/T	T/T
87	0	269	T/T	T/T
89	0	284	T/T	T/T
90	0	261	T/T	T/T
91	143	137	A/T	A/T
92	0	308	T/T	T/T
93	1	256	T/T	T/T
94	158	105	A/A	A/T
95	149	103	A/A	A/T

Tabla 4: Estado materno y fetal predicho para SNP rs2470750.

Muestra	SNP	Recuentos A	Recuentos T	Estado fetal predicho	Estado materno predicho
1	rs2470750	243	15	A/T	A/A
2		265	0	A/A	A/A
4		170	107	A/A	A/T
5		196	30	A/T	A/A
6		141	144	A/T	A/T
7		139	137	A/T	A/T
8		272	0	A/A	A/A
9		218	0	A/A	A/A
10		216	0	A/A	A/A
11		228	0	A/A	A/A
12		6	224	A/T	T/T

ES 2 718 111 T3

13	126	93	A/A	A/T
14	125	123	A/T	A/T
15	234	20	A/T	A/A
16	147	113	A/A	A/T
17	235	2	A/T	A/A
18	129	142	A/T	A/T
19	132	114	A/A	A/T
20	214	0	A/A	A/A
21	1	245	T/T	T/T
22	141	111	A/A	A/T
23	135	128	A/A	A/T
24	121	160	T/T	A/T
25	209	21	A/T	A/A
26	0	239	T/T	T/T
27	203	4	A/T	A/A
28	101	115	T/T	A/T
29	212	10	A/T	A/A
30	86	101	T/T	A/T
31	118	116	A/T	A/T
32	135	121	A/A	A/T
33	111	128	T/T	A/T
34	120	118	A/T	A/T
35	246	0	A/A	A/A
36	113	115	A/T	A/T
37	96	126	T/T	A/T
38	107	88	A/A	A/T
39	241	0	A/A	A/A
40	116	118	A/T	A/T
41	135	89	A/A	A/T
42	129	85	A/A	A/T
43	0	205	T/T	T/T
44	138	88	A/A	A/T
45	129	86	A/A	A/T
46	108	123	T/T	A/T
47	14	246	A/T	T/T
48	129	148	T/T	A/T
49	108	110	A/T	A/T
50	120	124	A/T	A/T
51	212	22	A/T	A/T
52	237	0	A/A	A/A
53	104	147	T/T	A/T
54	134	126	A/T	A/T
55	128	82	A/A	A/T
56	225	5	A/T	A/A
57	213	11	A/T	A/A
58	125	116	A/T	A/T

ES 2 718 111 T3

59	226	1	A/A	A/A
60	103	119	T/T	A/T
61	84	91	T/T	A/T
62	130	104	A/A	A/T
63	251	0	A/A	A/A
64	243	0	A/A	A/A
65	127	115	A/A	A/T
66	113	104	A/A	A/T
67	26	190	A/T	T/T
68	80	83	A/T	A/T
69	122	132	T/T	A/T
70	0	235	T/T	T/T
71	90	123	T/T	A/T
73	174	0	A/A	A/A
75	0	233	T/T	T/T
76	220	0	A/A	A/A
78	115	115	A/T	A/T
79	112	144	T/T	A/T
80	10	248	A/T	T/T
81	241	0	A/A	A/A
82	228	0	A/A	A/A
83	243	16	A/T	A/A
84	133	104	A/A	A/T
85	101	99	A/T	A/T
86	1	209	A/A	A/A
87	224	7	A/T	A/A
89	122	101	A/A	A/T
90	130	89	A/A	A/T
91	128	151	T/T	A/T
92	231	0	A/A	A/A
93	107	118	T/T	A/T
94	93	100	A/T	A/T
95	132	119	A/A	A/T

Tabla 5: Estado materno y fetal predicho para SNP rs6815910.

Muestra	SNP	Recuentos A	Recuentos T	Estado fetal predicho	Estado materno predicho
1	rs6815910	295	32	A/T	A/A
10		133	107	A/A	A/T
11		115	131	T/T	A/T
12		311	10	A/T	A/A
13		18	252	A/T	T/T
14		132	178	T/T	A/T
15		288	0	A/A	A/A
16		325	1	A/A	A/A
17		11	276	A/T	T/T

ES 2 718 111 T3

18	282	0	A/A	A/A
19	131	133	A/T	A/T
2	7	311	A/T	T/T
20	135	116	A/A	A/T
21	121	140	T/T	A/T
22	287	11	A/T	A/A
23	148	146	A/T	A/T
24	185	138	A/A	A/T
25	116	126	T/T	A/T
26	235	0	A/A	A/A
27	288	0	A/A	A/A
28	242	0	A/A	A/A
29	239	12	A/T	A/A
30	235	24	A/T	A/A
31	126	148	T/T	A/T
32	25	256	A/T	T/T
33	286	1	A/A	A/A
34	158	156	A/T	A/T
35	287	0	A/A	A/A
36	118	133	T/T	A/T
37	163	119	A/A	A/T
38	273	10	A/T	A/A
39	132	148	T/T	A/T
4	0	343	T/T	T/T
40	143	177	T/T	A/T
41	0	308	T/T	T/T
42	297	0	A/A	A/A
43	117	130	T/T	A/T
44	296	1	A/A	A/A
45	276	0	A/A	A/A
46	140	134	A/T	A/T
47	158	139	A/A	A/T
48	0	304	T/T	T/T
49	251	13	A/T	A/A
5	138	115	A/A	A/T
50	142	162	T/T	A/T
51	0	306	T/T	T/T
52	249	21	A/T	A/A
53	111	170	T/T	A/T
54	140	151	A/T	A/T
55	102	217	T/T	A/T
56	315	0	A/A	A/A
57	123	158	T/T	A/T
58	146	168	T/T	A/T
59	226	50	A/T	A/A
6	309	0	A/A	A/A

ES 2 718 111 T3

60	122	133	T/T	A/T
61	240	28	A/T	A/A
62	132	124	A/T	A/T
63	291	9	A/T	A/A
64	0	304	T/T	T/T
65	273	0	A/A	A/A
66	154	139	A/T	A/T
67	145	153	A/T	A/T
68	110	163	T/T	A/T
69	131	134	A/T	A/T
7	186	127	A/A	A/T
70	167	163	A/T	A/T
71	238	26	A/T	A/A
73	18	244	A/T	T/T
75	130	129	A/T	A/T
76	133	113	A/A	A/T
78	237	2	A/T	A/A
79	0	278	T/T	T/T
8	192	159	A/A	A/T
80	153	131	A/A	A/T
81	25	229	A/T	T/T
82	0	256	T/T	T/T
83	152	142	A/T	A/T
84	290	2	A/T	A/A
85	270	0	A/A	A/A
86	0	242	T/T	T/T
87	150	134	A/A	A/T
89	169	117	A/A	A/T
9	271	1	A/A	A/A
90	109	144	T/T	A/T
91	261	12	A/T	A/A
92	258	0	A/A	A/A
93	0	309	T/T	T/T
94	116	146	T/T	A/T
95	123	116	A/T	A/T

5 La ubicación de estos SNPS se denota usando dbSNP versión 132 y GRCH37/UCSC hg 19. Los datos para estos polimorfismos se obtuvieron en el mismo conjunto de datos que los datos de aneuploidía ilustrados en las figuras 10 y 11. Por tanto, un solo ensayo demostró la capacidad de identificar la aneuploidía fetal, las diferencias polimórficas entre los loci fetales y maternos, y la información real de SNP para los loci fetales seleccionados en un solo ensayo.

Lista de secuencias

- 10 <110> Tandem Diagnostics, Inc. Sparks, Andrew Oliphant, Arnold Zahn, Jacob Song, Ken Stuelpnagel, John
- <120> SISTEMAS DE ENSAYO PARA ANÁLISIS GENÉTICO
- <130> TAND009PCT
- 15 <150> 13/013.732
- <151> 25-01-2011

<150> 61/371605
 <151> 06-08-2010

5 <160> 11
 <170> PatentIn versión 3.5

<210> 1
 10 <211> 24
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

<220>
 15 <223> Oligonucleótido sintético

<400> 1
 tacaccggcg ttatgcgtcg agac 24

20 <210> 2
 <211> 24
 <212> ADN
 <213> ORGANISMO ARTIFICIAL

25 <220>
 <223> Oligonucleótido sintético

<400> 2
 attgcgggga ccgatgatcg cgtc 24

30 <210> 3
 <211> 48
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

35 <220>
 <223> Oligonucleótido sintético

<400> 3
 40 taatgatacg gcgaccaccg agatctacac cggcggtatg cgtcgaga 48

<210> 4
 <211> 53
 <212> ADN
 45 <213> ORGANISMO ARTIFICIAL

<220>
 <223> Oligonucleótido sintético

<220>
 50 <221> Índice
 <222> (26)..(26)
 <223> n es un índice de secuencia.

<220>
 55 <221> misc_feature
 <222> (26)..(26)
 <223> n es a, c, g o t

<400> 4
 60 tcaagcagaa gacggcatac gagatnaaac gacgcgatca tcggtccccg caa 53

<210> 5
 <211> 21
 <212> ADN
 65 <213> SECUENCIA ARTIFICIAL

<220>
 <223> Oligonucleótido sintético

<400> 5
 5 taatgatacg gcgaccaccg a 21

<210> 6
 <211> 25
 <212> ADN
 10 <213> SECUENCIA ARTIFICIAL

<220>
 <223> ATCTCGTATGCCGTCTTCTGCTTGA

<400> 6
 15 atctcgtatg ccgtcttctg ctgga 25

<210> 7
 <211> 48
 20 <212> ADN
 <213> SECUENCIA ARTIFICIAL

<220>
 <223> Oligonucleótido sintético

<400> 7
 25 aatgatacgg cgaccaccga gatctacacc ggcggtatgc gtcgagac 48

<210> 8
 <211> 60
 30 <212> ADN
 <213> SECUENCIA ARTIFICIAL

<220>
 <223> Oligonucleótido sintético

<220>
 <221> n
 <222> (26)..(32)
 40 <223> n = G, A, T o C

<220>
 <221> misc_feature
 <222> (26)..(32)
 45 <223> n es a, c, g o t

<400> 8
 tcaagcagaa gacggcatag gagatnnnnn nnaaacgacg cgatcatcgg tccccgcaat 60

<210> 9
 <211> 52
 <212> ADN
 <213> SECUENCIA ARTIFICIAL

<220>
 <223> Oligonucleótido sintético

<220>
 <221> w
 <222> (27)..(27)
 60 <223> w = A o T

<400> 9
 65 ttacatgtc ttgggcatt ttaggtwgag tgaaatctag gccttgcaaa tc 52

<210> 10

ES 2 718 111 T3

<211> 52
<212> ADN
<213> SECUENCIA ARTIFICIAL

5 <220>
<223> Oligonucleótido sintético

<220>
<221> w

10 <222> (27)..(27)
<223> w = A o T

<400> 10

15 ttgtgtaacg ttaacctcag ggaccawgag atgtacttag tattaattg cc 52

<210> 11
<211> 52
<212> ADN
<213> Secuencia artificial

20 <220>
<223> Oligonucleótido sintético

<220>

25 <221> w
<222> (27)..(27)
<223> w = A o T

<400> 11

30 ggaagaagtg cagtgtagta gacaacwctg gcattgtgtt ttggaactg gg 52

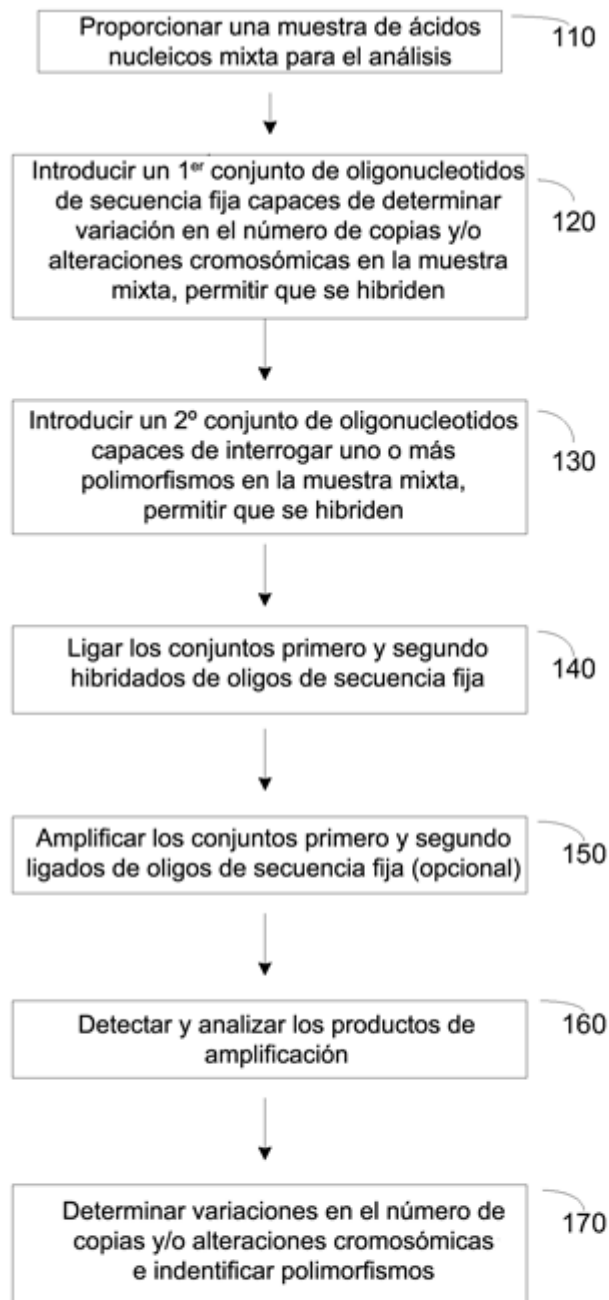
REIVINDICACIONES

1. Método de ensayo único para la detección de la presencia o ausencia de variación en el número de copias (VNC) de una región genómica y presencia o ausencia de uno o más polimorfismos en una muestra mixta usando un ensayo único, en el que la muestra mixta comprende ácidos nucleicos libres de células tanto fetales como maternas y en el que la muestra mixta es plasma materno o suero materno;

comprendiendo el método de ensayo las etapas de:

 - introducir un primer conjunto de oligonucleótidos de secuencia fija a la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en 24 o más loci en o asociados con un primer cromosoma, en el que al menos uno de los oligonucleótidos de secuencia fija comprende regiones de cebadores universales;
 - introducir un segundo conjunto de oligonucleótidos de secuencia fija a la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en 24 o más loci en o asociados con un segundo cromosoma, en el que al menos uno de los oligonucleótidos de secuencia fija comprende regiones de cebadores universales;
 - introducir un tercer conjunto de oligonucleótidos de secuencia fija a la muestra mixta en condiciones que permiten que los oligonucleótidos de secuencia fija se hibriden específicamente con regiones complementarias en 24 o más loci con un posible polimorfismo, en el que al menos uno de los oligonucleótidos de secuencia fija comprende regiones de cebadores universales;
 - ligar los oligonucleótidos hibridados para crear productos de ligamiento contiguos complementarios a los loci del primer cromosoma, a los loci del segundo cromosoma y a los loci con un posible polimorfismo;
 - amplificar los productos de ligamiento contiguos usando las regiones de cebadores universales para crear productos de amplificación; y
 - detectar y cuantificar los productos de amplificación mediante secuenciación de alto rendimiento, en el que los loci seleccionados se miden cada uno en promedio al menos 100 veces;
 - en el que la detección de los productos de amplificación se correlaciona con el número de copias de una o más regiones genómicas y la presencia o ausencia de un polimorfismo en uno o más loci en la muestra mixta.
2. Método de ensayo único según la reivindicación 1, en el que el método de ensayo comprende además, antes de la etapa de ligamiento de oligonucleótidos hibridados, la etapa de:
 - introducir uno o más oligonucleótidos de puente en condiciones que permitan que los oligonucleótidos de puente se hibriden específicamente con regiones en los loci entre las regiones complementarias a los oligonucleótidos de secuencia fija del conjunto.
3. Método de ensayo único según la reivindicación 1, en el que los oligonucleótidos de secuencia fija de cada conjunto se hibridan con regiones inmediatamente adyacentes en los loci.
4. Método de ensayo único según la reivindicación 1, en el que uno o más conjuntos de oligonucleótidos de secuencia fija comprenden una sonda precircular.
5. Método de ensayo único según la reivindicación 1, en el que los productos de amplificación se aíslan antes de la detección.
6. Método de ensayo único según la reivindicación 5, en el que los productos de amplificación se aíslan como moléculas individuales antes de la detección.
7. Método de ensayo único según la reivindicación 6, en el que los productos de amplificación aislados individuales se amplifican además para crear copias idénticas de todos o de una parte de los productos de amplificación individuales antes de la detección.
8. Método de ensayo único según la reivindicación 5, en el que los productos de amplificación aislados individuales se amplifican además para crear copias idénticas de moléculas complementarias a todos o una parte de los productos de amplificación individuales antes de la detección.
9. Método de ensayo único según la reivindicación 1, en el que al menos un locus interrogado por número de copias es diferente de todos los loci interrogados por polimorfismos.

10. Método de ensayo único según la reivindicación 9, en el que varios loci interrogados por número de copias son diferentes de los loci interrogados por polimorfismos.
- 5 11. Método de ensayo único según la reivindicación 1, en el que el primer cromosoma es el cromosoma 18.
12. Método de ensayo único según la reivindicación 1, en el que el primer cromosoma es el cromosoma 21.
- 10 13. Método de ensayo único según la reivindicación 1, en el que el segundo cromosoma es el cromosoma 18.
14. Método de ensayo único según la reivindicación 1, en el que el segundo cromosoma es el cromosoma 21.



100

FIGURA 1

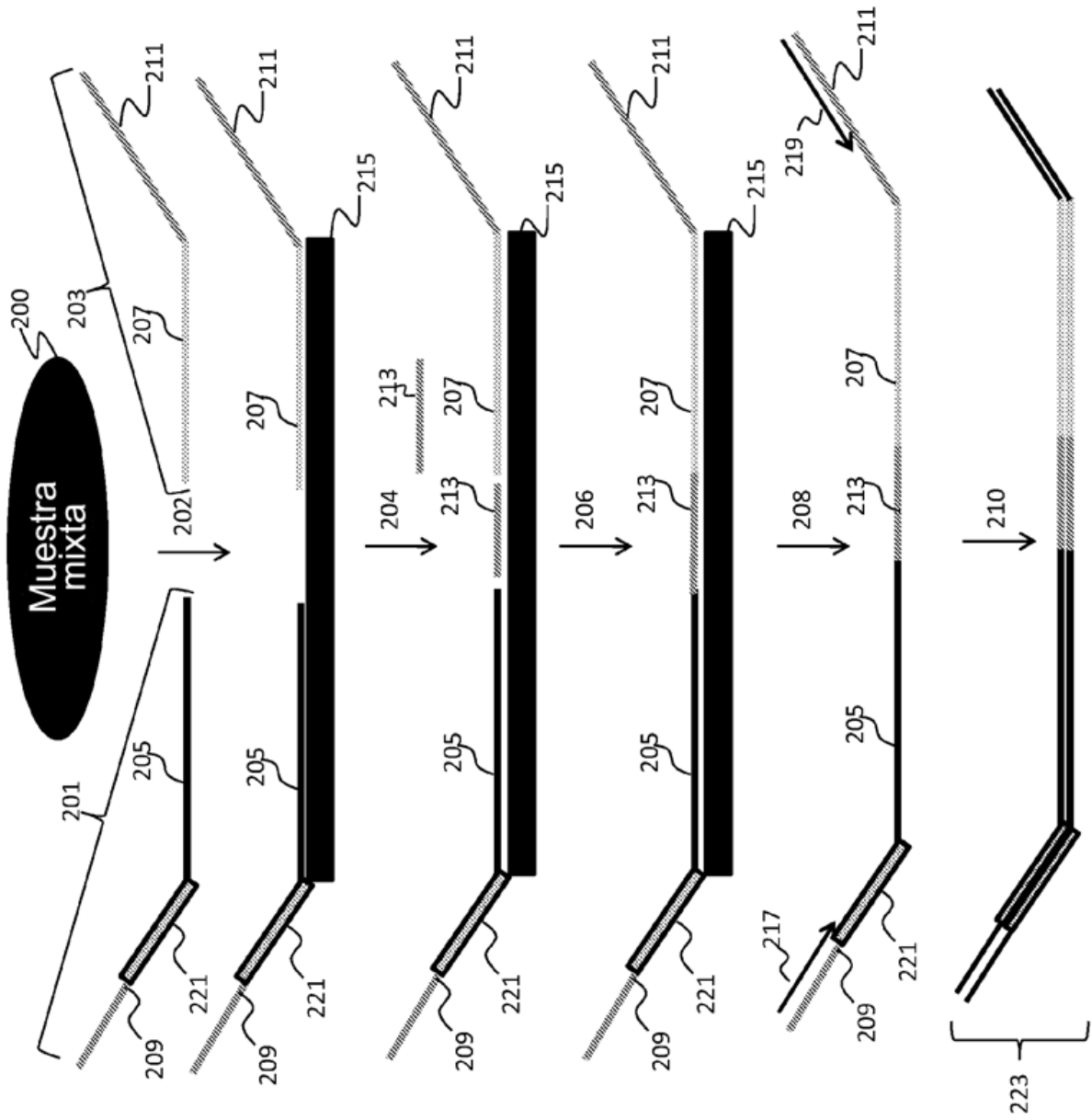


FIG. 2

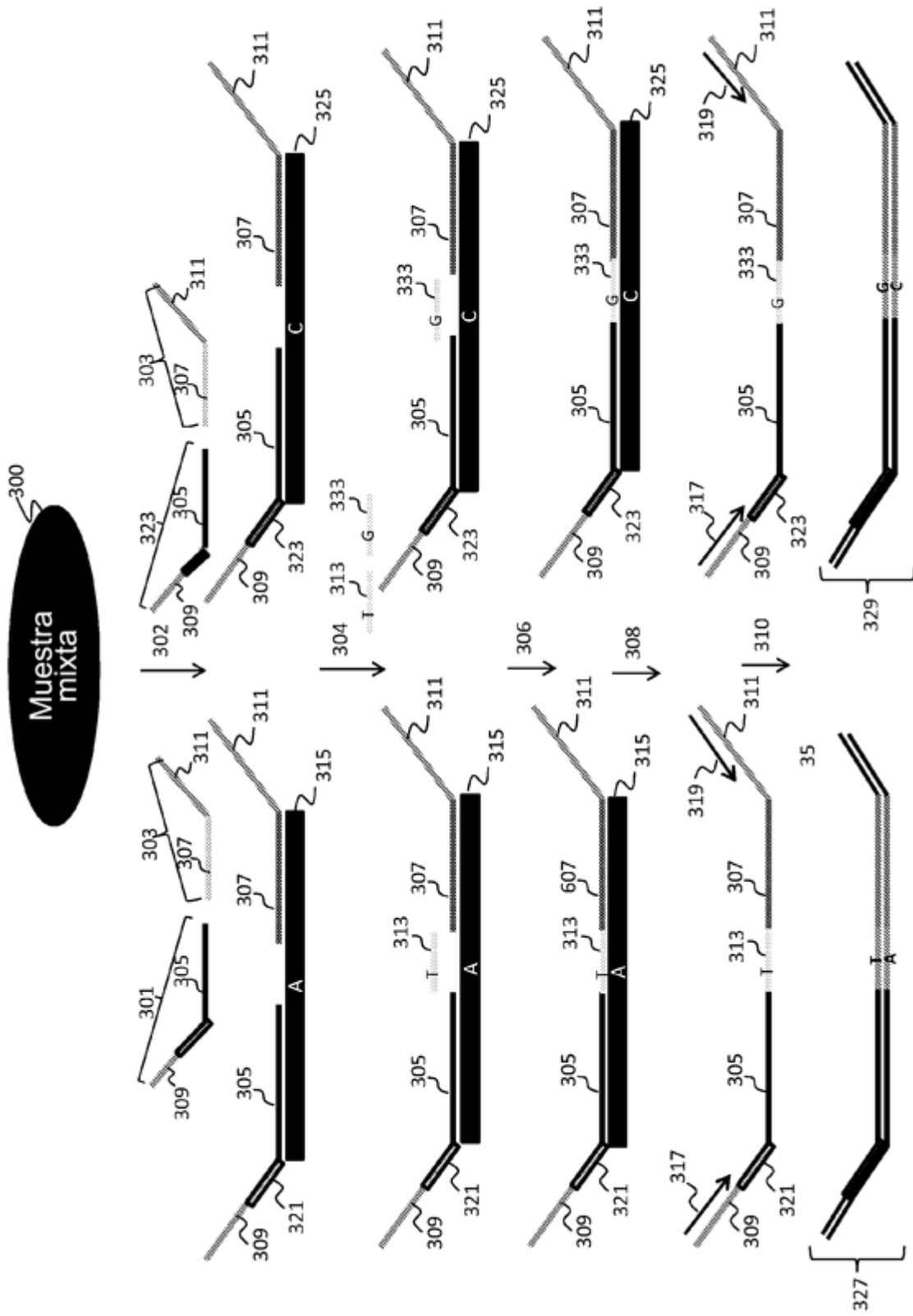


FIG. 3

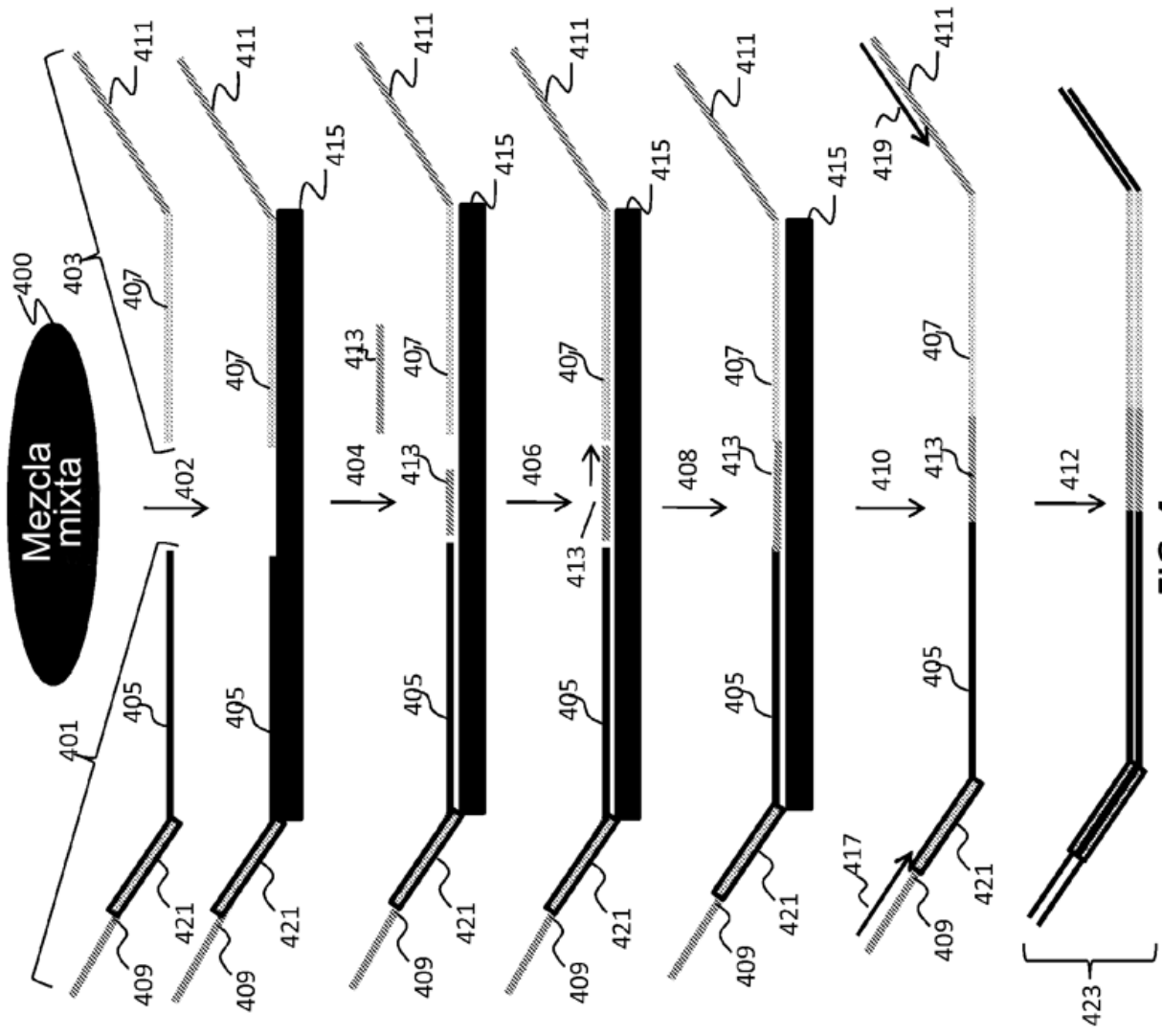


FIG. 4

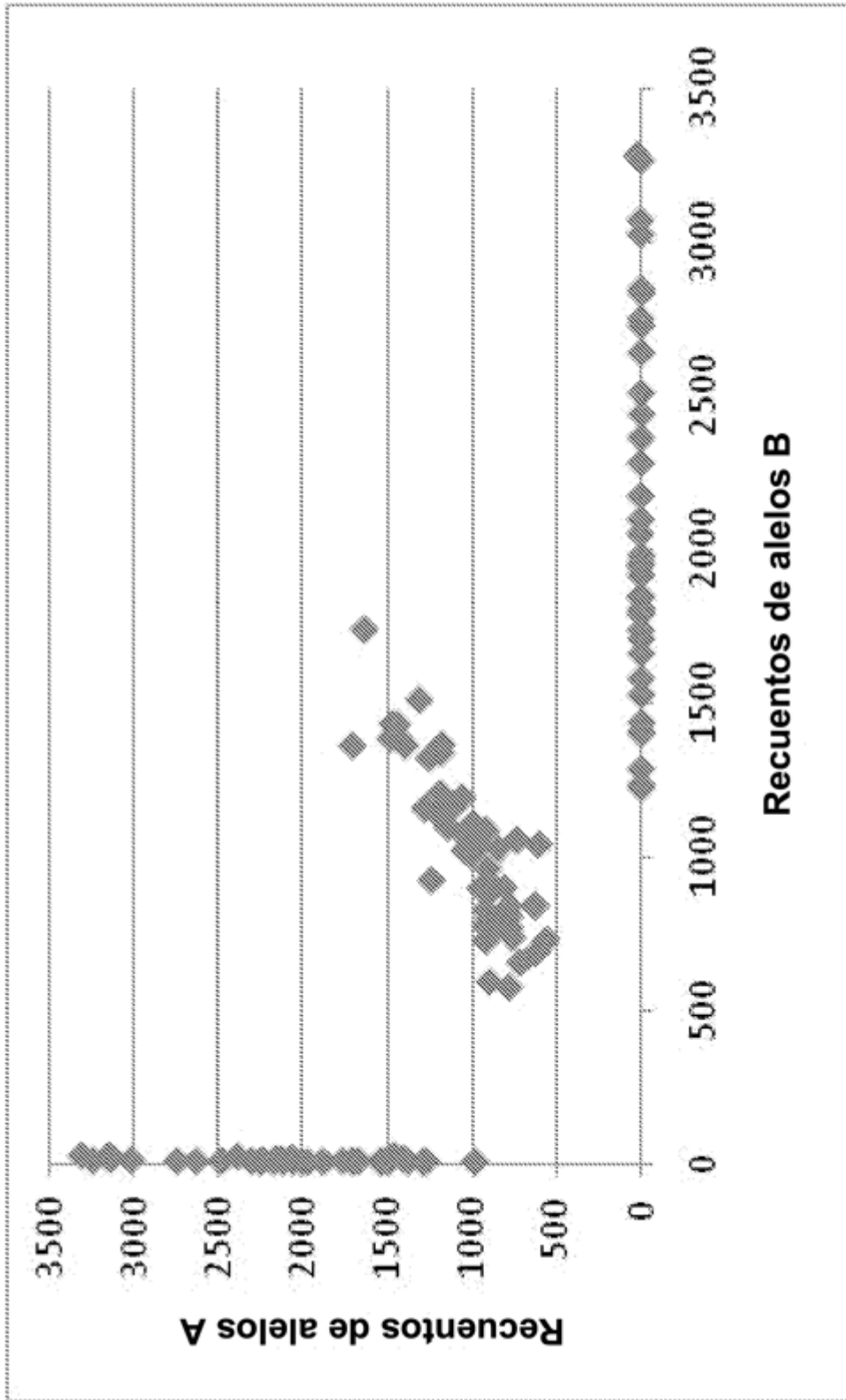


FIG. 5

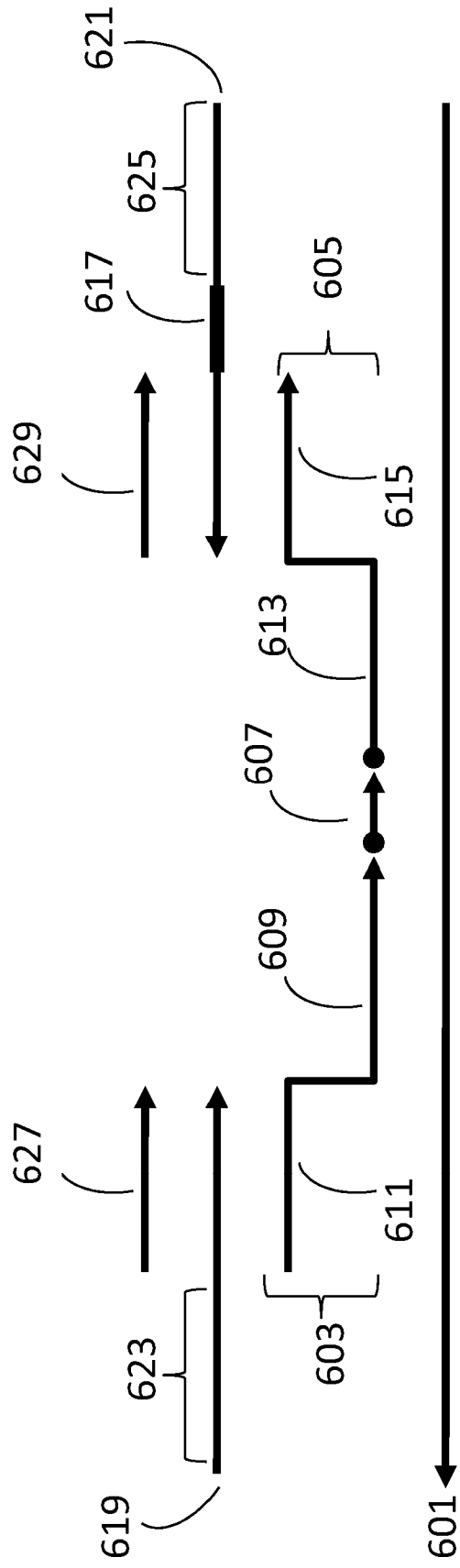


FIG. 6

Muestra	Etiqueta de muestra	Género fetal	Equív. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad gest. (día)	n.º de Loci Poli. dif.	% Fetal	Media 18	Media 21	Prop18	Prop21	Est. z 18	Est. z 21
1	TGTCCTCC	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. 7

22	CTCTATG	M	2002	46XX	46XY	30	123	30	0.08322	994.7936	999.994	0.498697	0.501303	-1.3432	1.343199
	Muestra	Etiqueta de muestra	Género fetal	Equiv. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad gest. (día)	% fetal	Media 18	Media 21	Prop18	Prop21	Est. z 18	Est. z 21
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	GAACCCG	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. 7 (cont.)

Muestra	Etiqueta de muestra	Género fetal	Equiv. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad gest. (día)	Loci informativos	% Fetal	Media 18	Media 21	Prop18	Prop21	Est. z 18	Est. z 21
43	AAGCGT	F	6984	46XX	46XX	37	130	36	0.228387	998.8644	995.701	0.500793	0.499207	0.027059	-0.02706
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	TTGGAT	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. 7 (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	Etiqueta de muestra	Género fetal	Equiv. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad gest. (día)	Loci informativos	% Fetal	Media 18	Media 21	Prop18	Prop21	Est. z 18	Est. z 21
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. 7 (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	
	Muestra	Etiqueta de muestra	Género fetal	Equiv. de genoma	Cariotipo materno	Cariotipo fetal	Edad de la madre	Edad gest. (día)	Loci informativos	% Fetal	Media 18	Media 21	Prop18	Prop21	Est. z 18	Est. z 21
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. 7 (cont.)

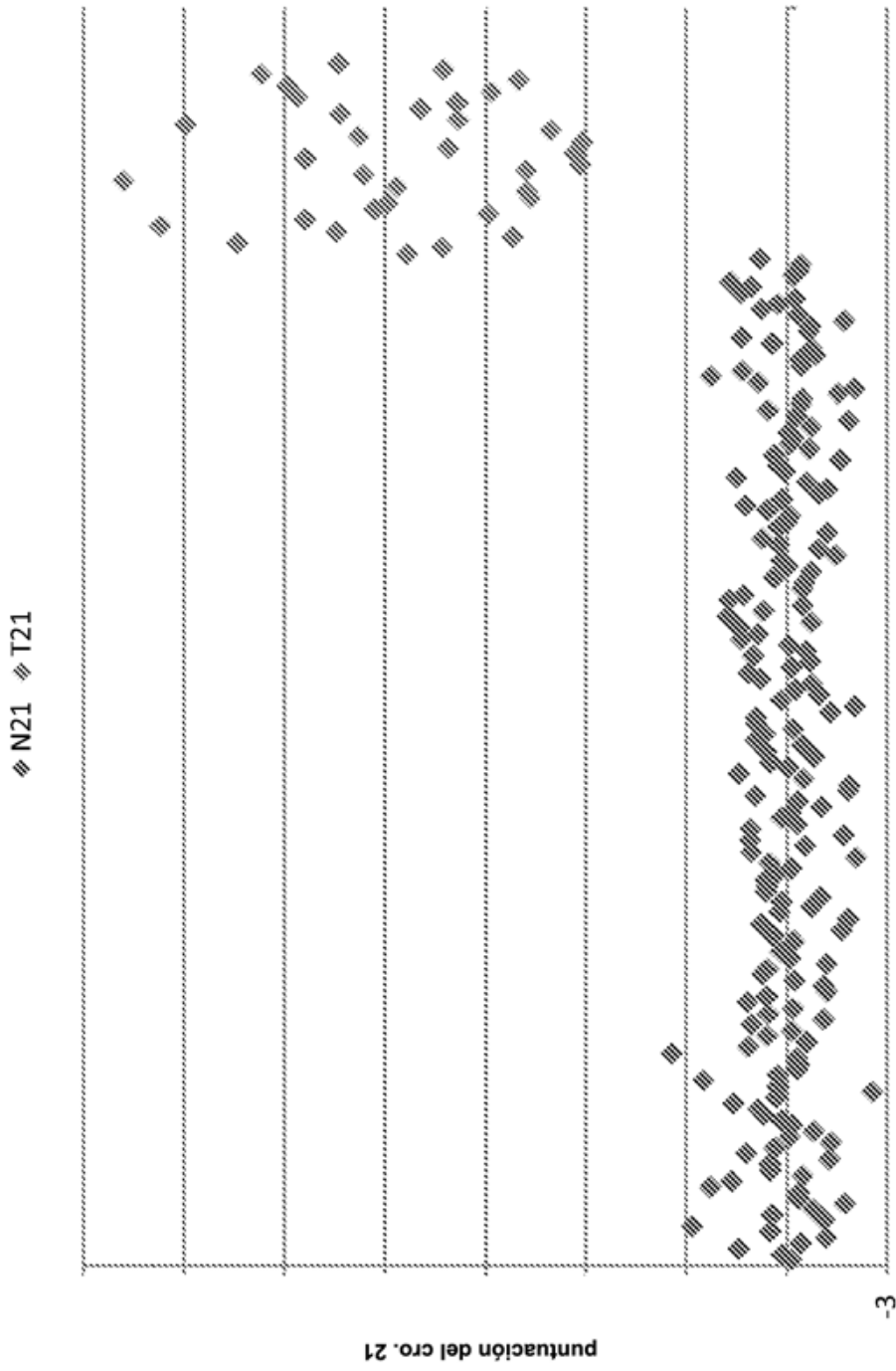


FIG. 8

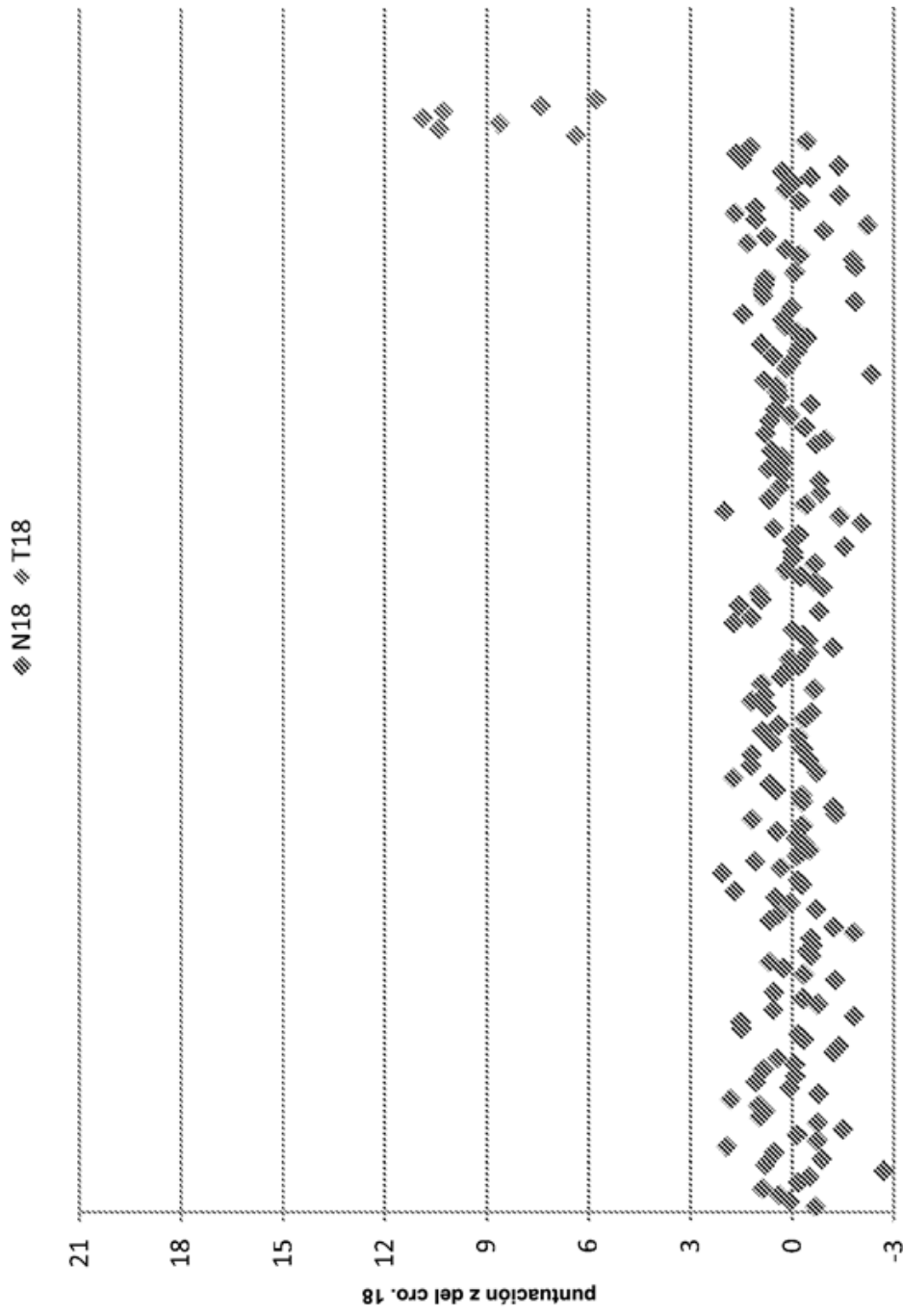


FIG. 9

◆ N21 ◆ T21

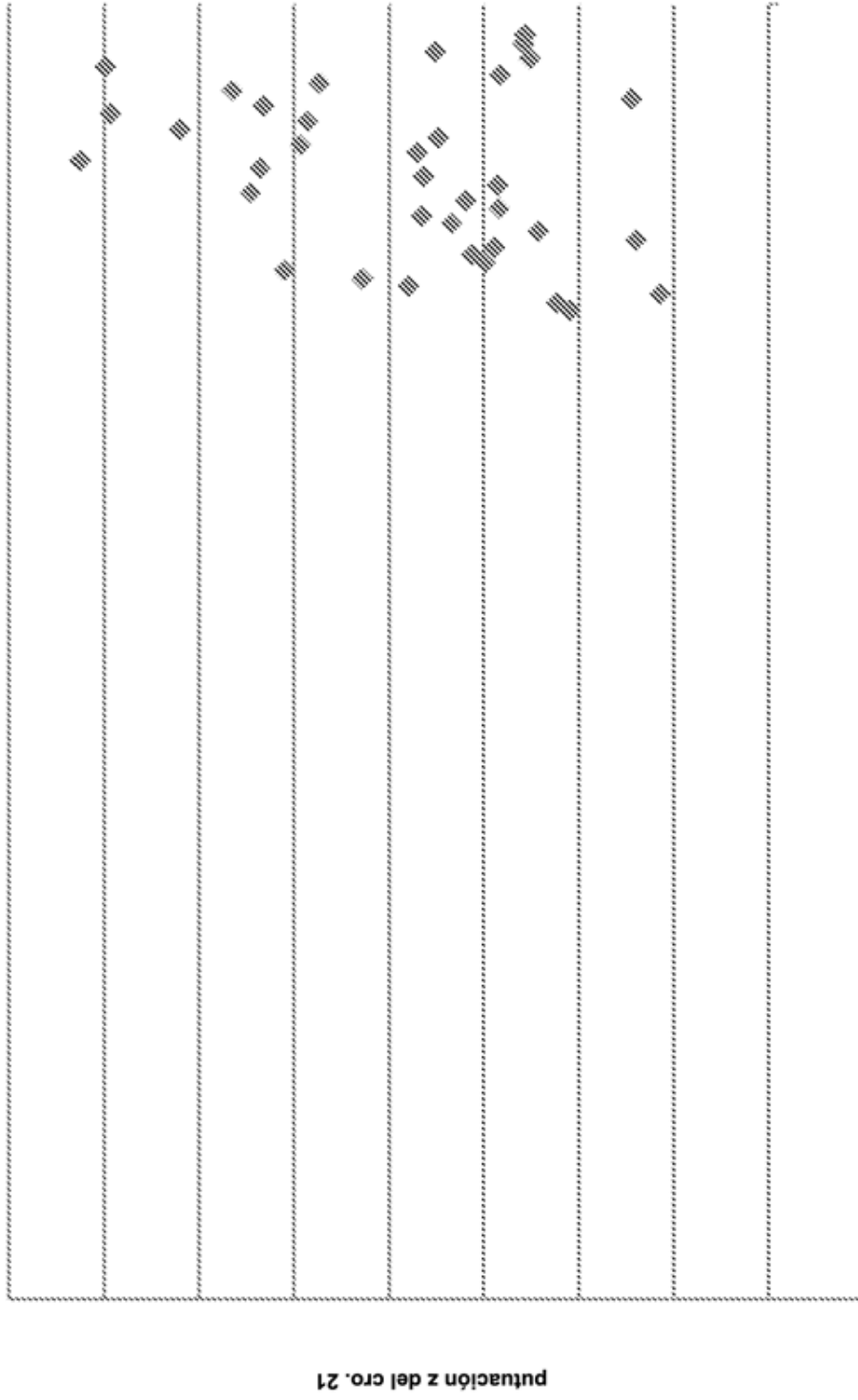


FIG. 10

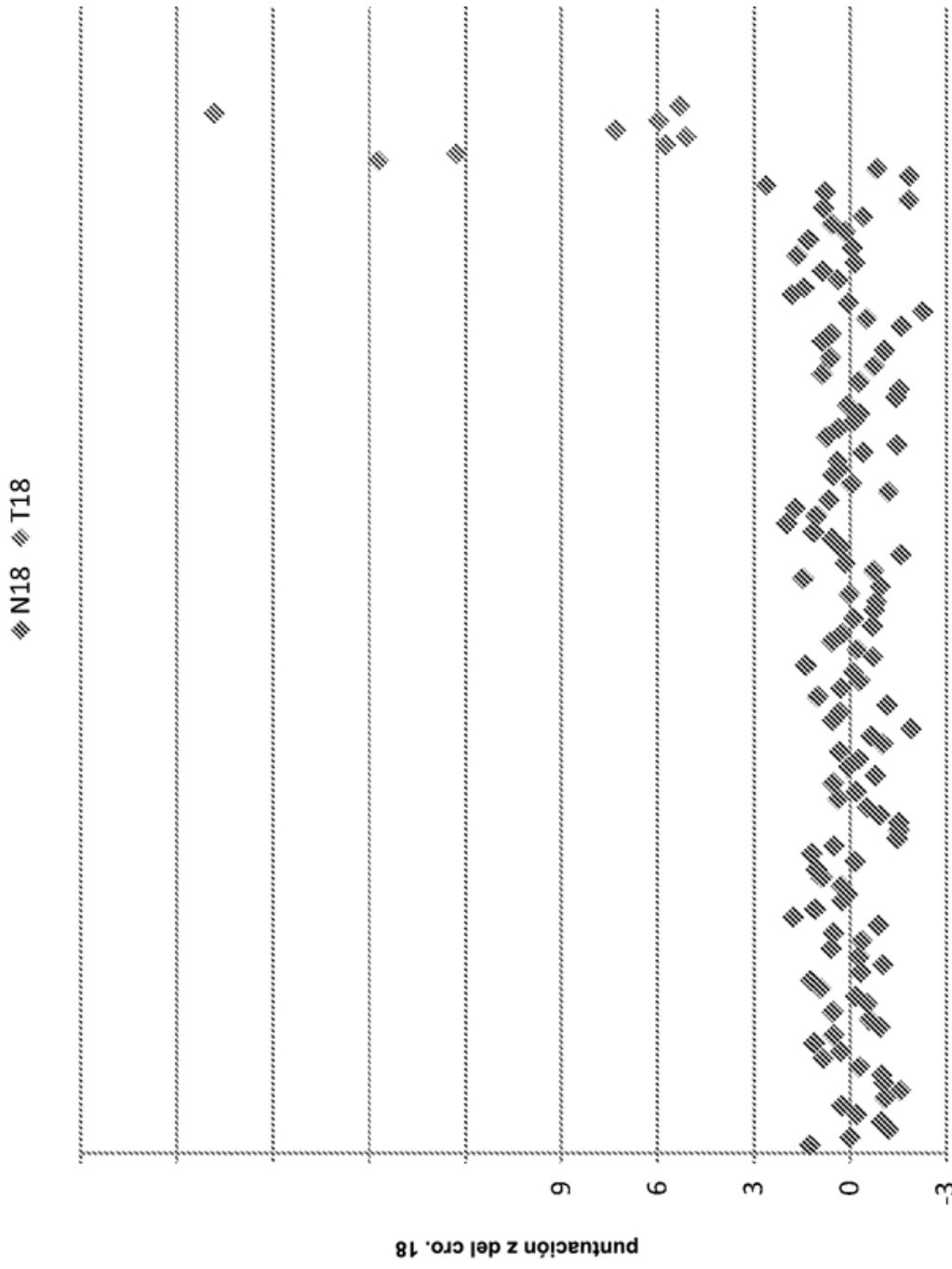


FIG. 11