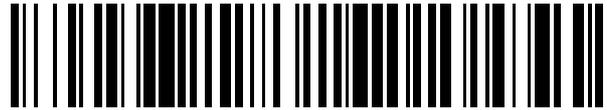


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 772 029**

51 Int. Cl.:

C12Q 1/68

(2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.05.2013 PCT/US2013/041906**

87 Fecha y número de publicación internacional: **28.11.2013 WO13177086**

96 Fecha de presentación y número de la solicitud europea: **20.05.2013 E 13726373 (7)**

97 Fecha y número de publicación de la concesión europea: **25.12.2019 EP 2852680**

54 Título: **Métodos y procesos para la evaluación no invasiva de variaciones genéticas**

30 Prioridad:

21.05.2012 US 201261649841 P
20.12.2012 US 201261740377 P
20.12.2012 US 201261740368 P
01.03.2013 US 201313782857
01.03.2013 US 201313782883

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.07.2020

73 Titular/es:

SEQUENOM, INC. (100.0%)
3595 John Hopkins Court
San Diego, CA 92121, US

72 Inventor/es:

DECIU, COSMIN y
KIM, SUNG, K.

74 Agente/Representante:

ILLESCAS TABOADA, Manuel

ES 2 772 029 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

Métodos y procesos para la evaluación no invasiva de variaciones genéticas

5 **Campo**

La presente invención está definida por las reivindicaciones. En el presente documento se describe, en líneas generales, tecnología que se relaciona, en parte, con los métodos, procesos y máquinas para la evaluación no invasiva de variaciones genéticas.

10

Antecedentes

La información genética de los organismos vivos (por ejemplo, animales, plantas y microorganismos) y otras formas de información genética de replicación (por ejemplo, virus) se codifica en el ácido desoxirribonucleico (ADN) o en el ácido ribonucleico (ARN). La información genética es una sucesión de nucleótidos o nucleótidos modificados que representan la estructura primaria de los ácidos nucleicos químicos o hipotéticos. En los seres humanos, el genoma completo contiene aproximadamente 30.000 genes ubicados en veinticuatro (24) cromosomas (véase The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992). Cada gen codifica una proteína específica, que después de la expresión mediante transcripción y traducción, cumple una función bioquímica específica dentro de una célula viva.

15

20

Muchas afecciones médicas son causadas por una o más variaciones genéticas. Ciertas variaciones genéticas causan afecciones médicas que incluyen, por ejemplo, hemofilia, talasemia, Distrofia Muscular de Duchenne (DMD), Enfermedad de Huntington (EH), Enfermedad de Alzheimer y fibrosis quística (FQ) (Human Genome Mutations, D. N. Cooper y M. Krawczak, BIOS Publishers, 1993). Dichas enfermedades genéticas pueden ser el resultado de una adición, sustitución o delección de un solo nucleótido en el ADN de un gen particular. Ciertos defectos de nacimiento están causados por una anomalía cromosómica, también conocida como aneuploidía, tal como Trisomía 21 (Síndrome de Down), Trisomía 13 (Síndrome de Patau), Trisomía 18 (Síndrome de Edward), Monosomía X (Síndrome de Turner) y ciertas aneuploidías de los cromosomas sexuales como el Síndrome de Klinefelter (XXY), por ejemplo. Otra variación genética es el sexo fetal, que a menudo se puede determinar basándose en los cromosomas sexuales X e Y. Algunas variaciones genéticas pueden predisponer a un individuo a, o causar, una serie de enfermedades tales como, por ejemplo, diabetes, arteriosclerosis, obesidad, diversas enfermedades autoinmunitarias y cáncer (p. ej., colorrectal, de mama, de ovario, de pulmón).

25

30

La identificación de una o más variaciones o varianzas genéticas puede llevar a un diagnóstico de, o determinar la predisposición a, una afección médica particular. La identificación de una varianza genética puede dar lugar a facilitar una decisión médica y/o emplear un procedimiento médico útil. En determinados casos, la identificación de una o más variaciones o varianzas genéticas implica el análisis de ADN libre de células. El ADN libre de células (ADNcf) está compuesto de fragmentos de ADN que se originan a partir de la muerte celular y circulan en la sangre periférica. Las altas concentraciones de ADNcf pueden ser indicativas de ciertas afecciones clínicas como el cáncer, traumatismo, quemaduras, infarto de miocardio, ictus, sepsis, infección y otras dolencias. Adicionalmente, el ADN fetal libre de células (ADN-CCF) puede detectarse en el torrente sanguíneo materno y usarse para varios diagnósticos prenatales no invasivos. En el documento WO2013/055817 se describen métodos, procesos y aparatos para la evaluación no invasiva de variaciones genéticas. En el documento WO2011/091063 se describen métodos de detección definidos por partición, entre otros, para determinar la aneuploidía fetal. En el documento WO2010/065470 se describen composiciones y métodos para distinguir el ADN masculino fetal del no fetal en una muestra de sangre entera materna humana. En el documento WO2012/118745, se describe la detección de regiones de ácido nucleico seleccionadas de regiones pseudoautosómicas para identificar aneuploidía cromosómica sexual y para determinar el sexo fetal. Kitzman JO, et al., Sci Transl Med. 6 de junio de 2012; 4 (137): 137ra76, han descrito la secuenciación no invasiva del genoma completo. Tabor HK, et al., Am J Med Genet A. 2012 Oct; 158A (10): 2382-4, han analizado la secuenciación no invasiva del genoma fetal.

35

40

45

50

La presencia de ácido nucleico fetal en el plasma materno permite el diagnóstico prenatal no invasivo a través del análisis de una muestra de sangre materna. Por ejemplo, las anomalías cuantitativas del ADN fetal en el plasma materno pueden asociarse con varios trastornos asociados con el embarazo, incluyendo preeclampsia, parto prematuro, hemorragia antes del parto, placentación invasiva, síndrome de Down fetal y otras aneuploidías cromosómicas fetales. Por lo tanto, el análisis del ácido nucleico fetal en plasma materno puede ser un mecanismo útil para el control del bienestar feto-materno.

55

Sumario

60

La presente invención está definida por las reivindicaciones. En particular, la presente invención se refiere a un método para determinar la presencia o ausencia de una aneuploidía cromosómica, que comprende: (a) obtener recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son: (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácido nucleico que son más cortos que 160 bases, en donde la obtención de recuentos comprende eliminar lecturas de secuencias de nucleótidos que son de fragmentos más largos de 160 bases; (b)

65

normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas; y (c) determinar la presencia o ausencia de una aneuploidía cromosómica de acuerdo con los recuentos normalizados.

5 En el presente documento también se proporciona, en determinados aspectos, un método para determinar la presencia o ausencia de una variación genética que comprende (a) mapear lecturas de secuencias de nucleótidos en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas en las que se mapea una cantidad significativa de lecturas de fragmentos de
10 ácido nucleico más cortos que aproximadamente 150 bases, (b) determinar la cantidad de lecturas de secuencias mapeadas en el subconjunto de secciones genómicas y (c) determinar la presencia o ausencia de una variación genética de acuerdo con la cantidad determinada en (b).

15 En el presente documento también se proporciona un sistema o una máquina que comprende uno o más procesadores y una memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores cuyas instrucciones ejecutables por el uno o más procesadores están configuradas para (a) acceder a lecturas de secuencias de nucleótidos mapeadas en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas en las que se mapea una cantidad significativa
20 de lecturas de fragmentos de ácido nucleico más cortos que aproximadamente 150 bases, (b) determinar la cantidad de lecturas de secuencias mapeadas en el subconjunto de secciones genómicas; y (c) determinar la presencia o ausencia de una variación genética de acuerdo con la cantidad determinada en (b).

25 Ciertos aspectos de la tecnología se describen con más detalle en la siguiente descripción, ejemplos, reivindicaciones y dibujos.

Breve descripción de los dibujos

30 Los dibujos ilustran la tecnología descrita en el presente documento y no son limitantes. Para mayor claridad y facilidad de ilustración, los dibujos no están hechos a escala y, en algunos casos, pueden mostrarse varios aspectos exagerados o agrandados para facilitar la comprensión de realizaciones particulares.

35 La FIG. 1 muestra una comparación por pares de FRS (estadística de relación fetal, por sus siglas en inglés) (eje vertical izquierdo, histograma superior) y el número de exones por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 13. Las porciones se muestran en la parte inferior, eje X horizontal.

40 La FIG. 2 muestra una comparación por pares de FRS (eje vertical izquierdo, histograma superior) y el de contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 13. Las porciones se muestran en la parte inferior, eje X horizontal.

45 La FIG. 3 muestra una comparación por pares del número de exones por porción de 50 kb (eje vertical izquierdo, histograma superior) y el contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 13. Las porciones se muestran en la parte inferior, eje X horizontal.

50 La FIG. 4 muestra una comparación por pares de FRS (eje vertical izquierdo, histograma superior) y el número de exones por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 18. Las porciones se muestran en la parte inferior, eje X horizontal.

55 La FIG. 5 muestra una comparación por pares de FRS (eje vertical izquierdo, histograma superior) y el de contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 18. Las porciones se muestran en la parte inferior, eje X horizontal.

60 La FIG. 6 muestra una comparación por pares del número de exones por porción de 50 kb (eje vertical izquierdo, histograma superior) y el contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 18. Las porciones se muestran en la parte inferior, eje X horizontal.

65 La FIG. 7 muestra una comparación por pares de FRS (eje vertical izquierdo, histograma superior) y el número de exones por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 21. Las porciones se muestran en la parte inferior, eje X horizontal.

La FIG. 8 muestra una comparación por pares de FRS (eje vertical izquierdo, histograma superior) y el de contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el cromosoma 21. Las porciones se muestran en la parte inferior, eje X horizontal.

La FIG. 9 muestra una comparación por pares del número de exones por porción de 50 kb (eje vertical izquierdo, histograma superior) y el contenido de GC por porción de 50 kb (eje vertical derecho, histograma inferior) para el

cromosoma 21. Las porciones se muestran en la parte inferior, eje X horizontal.

La FIG. 10 muestra un ajuste PERUN con puntuaciones Z de regresión local LOESS (eje X) frente a un ajuste PERUN con puntuaciones Z de regresión local LOESS basadas en porciones "fetales no enriquecidas" (eje Y) para el cromosoma 21. Los cuatro cuadrantes representan concordancia y discordancia. Las líneas del cuadrante se dibujan en $Z = 3$. Los cuadrantes superior derecho e inferior izquierdo están separados por una línea diagonal gris discontinua. La línea de puntos es una línea de regresión solo para muestras que no son T21. La línea de guiones y puntos es una línea de regresión para muestras de T21 basadas en porciones con FRS elevadas.

La FIG. 11 muestra un ajuste PERUN con puntuaciones Z de regresión local LOESS (eje X) frente a un ajuste PERUN con puntuaciones Z de regresión local LOESS basadas en porciones "fetales enriquecidas" (es decir, porciones con FRS elevadas) (eje Y) para el cromosoma 21. Los cuatro cuadrantes representan concordancia y discordancia. Las líneas del cuadrante se dibujan en $Z = 3$. Los cuadrantes superior derecho e inferior izquierdo están separados por una línea diagonal gris discontinua. La línea de puntos es una línea de regresión solo para muestras que no son T21. La línea de guiones y puntos es una línea de regresión para muestras de T21 basadas en porciones con FRS elevadas.

La FIG. 12 muestra un método para determinar la longitud de las porciones de ácidos nucleicos, que incluye las etapas de 1) hibridación de la sonda (P; línea de puntos) con el fragmento (línea continua), 2) corte de la sonda y 3) medición de la longitud de la sonda. La determinación del tamaño del fragmento se muestra para un fragmento derivado fetal (F) y un fragmento derivado materno (M).

La FIG. 13 muestra una distribución de longitudes de fragmentos para tres métodos diferentes de preparación de bibliotecas. Incluyen enzimático con limpieza automatizada de perlas, enzimático sin limpieza automatizada de perlas y TRUSEQ con limpieza automatizada de perlas. Las líneas verticales representan tamaños de fragmentos de 143 bases y 166 bases.

La FIG. 14 muestra la representación del cromosoma 13 sin un filtro de tamaño de fragmento.

La FIG. 15 muestra la representación del cromosoma 13 con un filtro de tamaño de fragmento a 150 bases.

La FIG. 16 muestra la representación del cromosoma 18 sin un filtro de tamaño de fragmento.

La FIG. 17 muestra la representación del cromosoma 18 con un filtro de tamaño de fragmento a 150 bases.

La FIG. 18 muestra la representación del cromosoma 21 sin un filtro de tamaño de fragmento.

La FIG. 19 muestra la representación del cromosoma 21 con un filtro de tamaño de fragmento a 150 bases.

La FIG. 20 muestra la representación del cromosoma 13 (Ajuste PERUN con LOESS) con filtros de tamaño de fragmento variable.

La FIG. 21 muestra la representación del cromosoma 18 (Ajuste PERUN con LOESS) con filtros de tamaño de fragmento variable.

La FIG. 22 muestra la representación del cromosoma 21 (Ajuste PERUN con LOESS) con filtros de tamaño de fragmento variable.

La FIG. 23 muestra una tabla que presenta una descripción de los datos utilizados para ciertos análisis.

La FIG. 24 muestra un ejemplo ilustrativo de un sistema en donde se pueden implementar ciertos casos de la tecnología.

Descripción detallada

En el presente documento se proporcionan métodos para analizar polinucleótidos en una mezcla de ácidos nucleicos que incluyen, por ejemplo, métodos para determinar la presencia o ausencia de una variación genética. La evaluación de una variación genética, tal como, por ejemplo, una aneuploidía fetal, de una muestra materna generalmente implica la secuenciación del ácido nucleico presente en la muestra, mapear las lecturas de secuencia a ciertas regiones en el genoma, cuantificar las lecturas de secuencia para la muestra y analizar la cuantificación. Dichos métodos a menudo analizan directamente el ácido nucleico en la muestra y obtienen lecturas de la secuencia de nucleótidos para la totalidad o sustancialmente todo el ácido nucleico en la muestra, lo que puede ser costoso y puede generar datos superfluos y/o irrelevantes. Sin embargo, ciertas aproximaciones de separación basadas en la secuencia y/o basadas en la longitud combinadas con ciertos análisis basados en la secuencia y/o basados en la longitud, pueden generar información específica sobre las regiones genómicas selectivas, tal como, por ejemplo, un cromosoma específico, y

en algunos casos, pueden diferenciar los orígenes de los fragmentos de ácido nucleico, como orígenes maternos frente a fetales. Ciertos métodos pueden incluir el uso de métodos de secuenciación, técnicas de enriquecimiento y análisis basados en la longitud. Ciertos métodos descritos en el presente documento pueden realizarse sin determinar secuencias de nucleótidos de los fragmentos de ácido nucleico. En el presente documento se proporcionan métodos para analizar polinucleótidos en una mezcla de ácidos nucleicos (por ejemplo, determinar la presencia o ausencia de una aneuploidía fetal) utilizando una combinación de aproximaciones de separación y análisis basadas en la secuencia y/o basadas en la longitud.

También se proporcionan métodos, procesos y máquinas útiles para identificar una variación genética. Identificar una variación genética a veces comprende detectar una variación en el número de copias y/o en ocasiones comprende ajustar un nivel que comprende una variación en el número de copias. En algunos casos, se ajusta un nivel proporcionando una identificación de una o más variaciones o varianzas genéticas con una probabilidad reducida de un diagnóstico falso positivo o falso negativo. En algunos casos, la identificación de una variación genética mediante un método descrito en el presente documento puede llevar a un diagnóstico de, o determinar una predisposición a, una afección médica particular. La identificación de una varianza genética puede dar lugar a facilitar una decisión médica y/o emplear un procedimiento médico útil.

También se proporcionan, en el presente documento, sistemas, máquinas y módulos que, en algunos casos, pueden realizar los métodos descritos en el presente documento.

Muestras

En el presente documento se proporcionan métodos y composiciones para analizar ácido nucleico. En algunos casos se analizan fragmentos de ácido nucleico en una mezcla de fragmentos de ácidos nucleicos. Una mezcla de ácidos nucleicos puede comprender dos o más especies de fragmentos de ácido nucleico que tienen diferentes secuencias de nucleótidos, diferentes longitudes de fragmentos, diferentes orígenes (p. ej., orígenes genómicos, orígenes fetales frente a orígenes maternos, orígenes celulares o tisulares, orígenes de muestras, orígenes de sujetos, y similares), o combinaciones de los mismos.

El ácido nucleico o una mezcla de ácido nucleico utilizada en los métodos y máquinas descritos en el presente documento a menudo se aísla de una muestra obtenida de un sujeto. Un sujeto puede ser cualquier organismo vivo o no vivo, incluyendo, pero sin limitación, un ser humano, un animal no humano, una planta, una bacteria, un hongo o un protista. Se puede seleccionar cualquier animal humano o no humano, incluyendo, pero sin limitación, mamífero, reptil, ave, anfibio, peces, ungulado, rumiante, bovino (p. ej., una vaca), equino (p. ej., un caballo), caprino y ovino (por ejemplo, oveja, cabra), porcino (por ejemplo, cerdo), camélido (por ejemplo, camello, llama, alpaca), mono, simio (por ejemplo, gorila, chimpancé), úrsido (por ejemplo, oso), ave de corral, perro, gato, ratón, rata, peces, delfín, ballena y tiburón. El sujeto puede ser macho o hembra (por ejemplo, una mujer, una gestante). El sujeto puede tener cualquier edad (por ejemplo, puede ser un embrión, un feto, un bebé, un niño, un adulto).

El ácido nucleico puede aislarse de cualquier tipo de muestra biológica adecuada o muestra (por ejemplo, una muestra de ensayo). Una muestra o muestra de ensayo puede ser cualquier muestra aislada u obtenida de un sujeto o parte del mismo (por ejemplo, un sujeto humano, una gestante, un feto). Ejemplos no limitantes de muestras incluyen líquido o tejido de un sujeto, incluyendo, sin limitación, sangre o un producto sanguíneo (por ejemplo, suero, plasma o similares), sangre de cordón umbilical, vellosidades coriónicas, líquido amniótico, líquido cefalorraquídeo, líquido espinal, líquido de lavado (por ejemplo, broncoalveolar, gástrico, peritoneal, ductal, ótico, artroscópico), muestra de biopsia (por ejemplo, de embrión preimplantado), muestra de celocentesis, células (células sanguíneas, células placentarias, células embrionarias o fetales, células nucleadas fetales o restos de células fetales) o partes de las mismas (por ejemplo, mitocondria, núcleo, extractos o similares), lavados del aparato reproductor femenino, orina, heces, esputo, saliva, mucosa nasal, líquido prostático, lavado, semen, líquido linfático, bilis, lágrimas, sudor, leche materna, líquido mamario, similares o combinaciones de los mismos. En algunos casos, una muestra biológica es un frotis cervical de un sujeto. En algunos casos, una muestra biológica puede ser sangre y, a veces, plasma o suero. El término "sangre", como se usa en el presente documento, se refiere a una muestra o preparación de sangre de una mujer gestante o una mujer que está siendo examinada para un posible embarazo. El término abarca sangre completa, producto sanguíneo o cualquier fracción de sangre, tal como suero, plasma, capa leucocitaria o similares, como se define convencionalmente. La sangre o sus fracciones a menudo comprenden nucleosomas (por ejemplo, nucleosomas maternos y/o fetales). Los nucleosomas comprenden ácidos nucleicos y algunas veces están libres de células o son intracelulares. La sangre también comprende capas leucocitarias. Las capas leucocitarias a veces se aíslan usando un gradiente de Ficoll. Las capas leucocitarias pueden comprender glóbulos blancos (por ejemplo, leucocitos, linfocitos T, células B, plaquetas y similares). En ciertos casos, las capas leucocitarias comprenden ácido nucleico materno y/o fetal. El plasma sanguíneo se refiere a la fracción de sangre completa resultante de la centrifugación de la sangre tratada con anticoagulantes. El suero sanguíneo se refiere a la porción acuosa de líquido que queda después de que una muestra de sangre se haya coagulado. Las muestras de líquido o tejido a menudo se recolectan de acuerdo con protocolos convencionales que generalmente siguen los hospitales o clínicas. Para la sangre, a menudo se recolecta una cantidad apropiada de sangre periférica (por ejemplo, entre 3 y 40 mililitros) y se puede almacenar de acuerdo con procedimientos convencionales antes o después de la preparación. Una muestra de líquido o tejido de la que se extrae el ácido nucleico puede ser acelular (por ejemplo, sin células). En algunos casos,

una muestra de líquido o tejido puede contener elementos celulares o restos celulares. En algunos casos, se pueden incluir células fetales o células cancerosas en la muestra.

5 Una muestra a menudo es heterogénea, lo que significa que está presente en la muestra más de un tipo de especie de ácido nucleico. Por ejemplo, el ácido nucleico heterogéneo puede incluir, pero sin limitación, (i) ácido nucleico derivado del feto y derivado de la madre, (ii) ácido nucleico de cáncer o no de cáncer, (iii) ácido nucleico de patógeno y de hospedador, y más generalmente, (iv) ácido nucleico mutado y de tipo nativo. Una muestra puede ser heterogénea porque está presente más de un tipo celular, tal como una célula fetal y una célula materna, una célula cancerosa y no cancerosa, o una célula patógena y hospedadora. En algunos casos hay una especie minoritaria de ácido nucleico y una especie mayoritaria de ácido nucleico.

15 Para las aplicaciones prenatales de la tecnología descrita en el presente documento, se puede obtener una muestra de tejido o líquido de una mujer en una edad gestacional adecuada para la prueba, o de una mujer a la que se le está haciendo una prueba para un posible embarazo. La edad gestacional adecuada puede variar según la prueba prenatal que se esté realizando. En ciertos casos, una gestante está a veces en el primer trimestre de gestación, otras veces en el segundo trimestre de gestación o, a veces, en el tercer trimestre de gestación. En ciertos casos, se recoge un líquido o tejido de una gestante entre aproximadamente 1 y aproximadamente 45 semanas de gestación fetal (por ejemplo, a 1-4, 4-8, 8-12, 12-16, 16-20, 20-24, 24-28, 28-32, 32-36, 36-40 o 40-44 semanas de gestación fetal), y a veces entre aproximadamente 5 y aproximadamente 28 semanas de gestación fetal (por ejemplo, a las 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 o 27 semanas de gestación fetal). En ciertos casos, se recoge una muestra de líquido o tejido de una gestante durante el parto o inmediatamente después del parto (por ejemplo, de 0 a 72 horas después del parto) (por ejemplo, parto vaginal o no vaginal (por ejemplo, parto quirúrgico)).

Adquisición de Muestras de Sangre y Extracción de ADN

25 Los métodos en el presente documento a menudo incluyen la separación, enriquecimiento y análisis del ADN fetal encontrado en la sangre materna como un medio no invasivo para detectar la presencia o ausencia de una variación genética materna y/o fetal y/o para monitorizar la salud de un feto y/o gestante durante y, a veces, después del embarazo. Por lo tanto, las primeras etapas para practicar ciertos métodos del presente documento a menudo incluyen obtener una muestra de sangre de una gestante y extraer ADN de una muestra.

Adquisición de Muestras de Sangre

35 Se puede obtener una muestra de sangre de una gestante en una edad gestacional adecuada para la prueba utilizando un método de la tecnología actual. Una edad gestacional adecuada puede variar según el trastorno examinado, como se analiza a continuación. La recogida de sangre de una mujer a menudo se realiza de acuerdo con el protocolo convencional que generalmente siguen los hospitales o clínicas. Una cantidad apropiada de sangre periférica, por ejemplo, generalmente entre 5-50 ml, a menudo se recolecta y puede almacenarse de acuerdo con el procedimiento convencional antes de una preparación adicional. Se pueden recoger muestras de sangre, almacenadas o transportadas de una manera que minimice la degradación o la calidad del ácido nucleico presente en la muestra.

Preparación de Muestras de Sangre

45 Se puede realizar un análisis de ADN fetal encontrado en sangre materna usando, por ejemplo, sangre completa, suero o plasma. Se conocen métodos para preparar suero o plasma a partir de sangre materna. Por ejemplo, la sangre de una gestante se puede colocar en un tubo que contiene EDTA o un producto comercial especializado como Vacutainer SST (Becton Dickinson, Franklin Lakes, N.J.) para evitar la coagulación de la sangre, y el plasma se puede obtener, a continuación, de la sangre completa mediante centrifugación. El suero se puede obtener con o sin centrifugación después de la coagulación sanguínea. Si se usa la centrifugación, entonces generalmente, aunque no de forma exclusiva, se realiza a una velocidad apropiada, por ejemplo, 1500-3000 veces g. El plasma o el suero pueden someterse a etapas de centrifugación adicionales antes de ser transferidos a un tubo nuevo para la extracción de ADN.

55 Además de la porción acelular de la sangre completa, el ADN también puede recuperarse de la fracción celular, enriquecido en la porción de capa leucocitaria, que se puede obtener después de la centrifugación de una muestra de sangre completa de la mujer y la eliminación del plasma.

Extracción de ADN

60 Existen numerosos métodos conocidos para extraer ADN de una muestra biológica, incluida la sangre. Se pueden seguir los métodos generales de preparación de ADN (por ejemplo, descritos por Sambrook y Russell, "Molecular Cloning: A Laboratory Manual 3ª ed., 2001); también se pueden usar diversos reactivos o kits disponibles comercialmente, como el kit QIAamp Circulating Nucleic Acid de QIAgen, el mini kit QiaAmp DNA o el mini kit QiaAmp DNA Blood (Qiagen, Hilden, Alemania), el kit de aislamiento de ADN en sangre GenomicPrep™ (Promega, Madison, Wisconsin) y el kit de purificación de ADN genómico en sangre GFX™ (Amersham, Piscataway, NJ) para obtener ADN de una muestra de sangre de una gestante. También pueden usarse combinaciones de más de uno de estos

métodos.

En algunos casos, la muestra puede primero enriquecerse o enriquecerse relativamente para el ácido nucleico fetal mediante uno o más métodos. Por ejemplo, la discriminación del ADN fetal y materno se puede realizar utilizando las composiciones y procesos de la tecnología desvelada en el presente documento, en solitario o en combinación con otros factores discriminantes. Ejemplos de estos factores incluyen, pero sin limitación, diferencias de un solo nucleótido entre los cromosomas X e Y, secuencias específicas del cromosoma Y, polimorfismos localizados en otras partes del genoma, diferencias de tamaño entre el ADN fetal y materno y diferencias en el patrón de metilación entre los tejidos maternos y fetales.

Se describen otros métodos para enriquecer una muestra para una especie particular de ácido nucleico en la solicitud de patente PCT número PCT/US07/69991, presentada el 30 de mayo de 2007, solicitud de patente PCT n.º PCT/US2007/071232, presentada el 15 de junio de 2007, solicitudes provisionales de EE. UU. números 60/968.876 y 60/968.878 (asignadas al solicitante), (solicitud de patente PCT número PCT/EP05/012707, presentada el 28 de noviembre de 2005). En algunos casos, el ácido nucleico materno se elimina de manera selectiva (ya sea parcial, sustancialmente, casi completamente o completamente) de la muestra.

Los términos "ácido nucleico" y "molécula de ácido nucleico" se pueden usar indistintamente a lo largo de la divulgación. Los términos se refieren a los ácidos nucleicos de cualquier composición, tal como ADN (por ejemplo, ADN complementario (ADNc), ADN genómico (ADNg) y similares), ARN (por ejemplo, ARN mensajero (ARNm), ARN inhibidor corto (ARNsi), ARN ribosómico (ARNr), ARNt, microARN, ARN altamente expresado por el feto o placenta, y similares), y/o análogos de ADN o ARN (por ejemplo, que contienen análogos de bases, análogos de azúcar y/o una cadena principal no natural y similares), híbridos de ARN/ADN y ácidos nucleicos de poliamida (ANP), todos los cuales pueden estar en forma monocatenaria o bicatenaria, y, a menos que se limite, pueden abarcar análogos conocidos de nucleótidos naturales que pueden funcionar de una manera similar a los nucleótidos de origen natural. Un ácido nucleico puede ser o puede ser de, un plásmido, fago, secuencia de replicación autónoma (SRA), centrómero, cromosoma artificial, cromosoma u otro ácido nucleico capaz de replicarse o replicarse *in vitro* o en una célula hospedadora, una célula, un núcleo celular o citoplasma de una célula en algunos casos. En algunos casos, un ácido nucleico molde puede ser de un solo cromosoma (por ejemplo, una muestra de ácido nucleico puede ser de un cromosoma de una muestra obtenida de un organismo diploide). A menos que se limite específicamente, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares a las del ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos de origen natural. A menos que se indique de otro modo, una secuencia particular de ácido nucleico también abarca implícitamente variantes modificadas conservativamente de la misma (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, polimorfismos de un solo nucleótido (SNP) y secuencias complementarias, así como la secuencia explícitamente indicada. Específicamente, las sustituciones de codones degenerados pueden lograrse generando secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) está sustituida con restos de base mixta y/o desoxiinosina. El término ácido nucleico se usa indistintamente con locus, gen, ADNc y ARNm codificado por un gen. El término también puede incluir, como equivalentes, derivados, variantes y análogos de ARN o ADN sintetizados a partir de análogos de nucleótidos, de cadena sencilla ("sentido" o "antisentido", cadenas "más" o cadenas "menos", marco de lectura "directo" o marco de lectura "inverso") y polinucleótidos bicatenarios. El término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica; incluye regiones que preceden y siguen a la región codificante (líder y tráiler) involucrada en la transcripción/traducción del producto génico y la regulación de la transcripción/traducción, así como secuencias intermedias (intrones) entre segmentos de codificación individuales (exones).

Los desoxirribonucleótidos incluyen desoxiadenosina, desoxicitidina, desoxiguanosina y desoxitimidina. Para el ARN, la base de citosina se reemplaza con uracilo. Se puede preparar un ácido nucleico molde utilizando un ácido nucleico obtenido de un sujeto como molde.

50 *Aislamiento y procesamiento de Ácidos Nucleicos*

El ácido nucleico puede derivar de una o más fuentes (por ejemplo, células, suero, plasma, capa leucocitaria, líquido linfático, piel, suelo y similares) por métodos conocidos en la técnica. Se puede usar cualquier método adecuado para el aislamiento, extracción y/o purificación de ADN de una muestra biológica (por ejemplo, de sangre o un producto sanguíneo), cuyos ejemplos no limitantes incluyen métodos de preparación de ADN (por ejemplo, descritos por Sambrook y Russell, "Molecular Cloning: A Laboratory Manual 3ª ed., 2001), varios reactivos o kits disponibles comercialmente, tal como Kit QIAamp Circulating Nucleic Acid de Qiagen, el mini kit QiaAmp DNA o el mini kit QiaAmp DNA Blood (Qiagen, Hilden, Alemania), el kit de aislamiento de ADN en sangre GenomicPrep™ (Promega, Madison, Wis) y el kit de purificación de ADN genómico en sangre GFX™ (Amersham, Piscataway, N.J.), similares o combinaciones de los mismos.

Los procedimientos y reactivos de lisis celular son conocidos en la técnica y generalmente se pueden realizar mediante métodos químicos (por ejemplo, detergente, soluciones hipotónicas, procedimientos enzimáticos y similares, o una combinación de los mismos), físicos (por ejemplo, prensa francesa, sonicación y similares) o de lisis electrolítica. Se puede utilizar cualquier procedimiento de lisis adecuado. Por ejemplo, los métodos químicos generalmente emplean agentes de lisis para romper las células y extraer los ácidos nucleicos de las células, seguido de un tratamiento con

sales caotrópicas. También son útiles los métodos físicos tales como congelación/descongelación seguidos de trituración, el uso de prensas de células y similares. Los procedimientos de lisis de sal elevada también se utilizan comúnmente. Por ejemplo, se puede utilizar cualquier procedimiento de lisis alcalina. El procedimiento anterior tradicionalmente incorpora el uso de soluciones de fenol-cloroformo, y se puede utilizar un procedimiento alternativo libre de fenol-cloroformo que implique tres soluciones. En los últimos procedimientos, una solución puede contener Tris 15 mM, pH 8,0; EDTA 10 mM y 100 ug/ml de RNasa A; una segunda solución puede contener NaOH 0,2 N y SDS al 1%; y una tercera solución puede contener KOAc 3 M, pH 5,5. Estos procedimientos se pueden encontrar en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989).

El ácido nucleico se puede aislar en un momento diferente en comparación con otro ácido nucleico, donde cada una de las muestras proviene de la misma fuente o de una diferente. Un ácido nucleico puede ser de una biblioteca de ácidos nucleicos, como una biblioteca de ADNc o ARN, por ejemplo. Un ácido nucleico puede ser un resultado de la purificación de ácido nucleico o el aislamiento y/o la amplificación de las moléculas de ácido nucleico de la muestra. El ácido nucleico proporcionado para los procesos descritos en el presente documento puede contener ácido nucleico de una muestra o de dos o más muestras (por ejemplo, de 1 o más, 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 11 o más, 12 o más, 13 o más, 14 o más, 15 o más, 16 o más, 17 o más, 18 o más, 19 o más, o 20 o más muestras).

En ciertos casos, los ácidos nucleicos pueden incluir ácido nucleico extracelular. La expresión "ácido nucleico extracelular", como se usa en el presente documento, puede referirse a ácido nucleico aislado de una fuente que sustancialmente no contiene células y también se conoce como ácido nucleico "libre de células", "ácido nucleico libre de células circulante" (por ejemplo, fragmentos CCF) y/o "ácido nucleico circulante libre de células". El ácido nucleico extracelular puede estar presente y obtenerse de la sangre (por ejemplo, de la sangre de una gestante). El ácido nucleico extracelular a menudo incluye células no detectables y puede contener elementos celulares o restos celulares. Ejemplos no limitantes de fuentes acelulares para el ácido nucleico extracelular son sangre, plasma en sangre, suero sanguíneo y orina. Como se usa en el presente documento, la expresión "obtener ácido nucleico de muestra circulante libre de células" incluye obtener una muestra directamente (por ejemplo, recoger una muestra, por ejemplo, una muestra de ensayo) u obtener una muestra de otra persona que haya recolectado una muestra. Sin quedar ligado a teoría alguna, el ácido nucleico extracelular puede ser un producto de la apoptosis celular y la degradación celular, que proporciona la base para el ácido nucleico extracelular que a menudo tiene una serie de longitudes a través de un espectro (por ejemplo, una "escala").

El ácido nucleico extracelular puede incluir diferentes especies de ácido nucleico y, por lo tanto, en el presente documento se denomina "heterogéneo". Por ejemplo, el suero o plasma sanguíneo de una persona que tiene cáncer puede incluir ácido nucleico de células cancerosas y ácido nucleico de células no cancerosas. En otro ejemplo, el suero o plasma sanguíneo de una gestante puede incluir ácido nucleico materno y ácido nucleico fetal. En algunos casos, el ácido nucleico fetal a veces es de aproximadamente el 5% a aproximadamente el 50% del ácido nucleico total (por ejemplo, aproximadamente 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, o 49 % del total del ácido nucleico es ácido nucleico fetal). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 500 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 500 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 250 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 250 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 200 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % del ácido nucleico fetal tiene una longitud de aproximadamente 200 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 150 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 150 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 100 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 100 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 50 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 50 pares de bases o menos). En algunos casos, la mayoría del ácido nucleico fetal en el ácido nucleico tiene una longitud de aproximadamente 25 pares de bases o menos (por ejemplo, aproximadamente el 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100% del ácido nucleico fetal tiene una longitud de aproximadamente 25 pares de bases o menos).

En ciertos casos, puede proporcionarse ácido nucleico para llevar a cabo los métodos descritos en el presente documento sin procesar la(s) muestra(s) que contiene(n) el ácido nucleico. En algunos casos, puede proporcionarse ácido nucleico para realizar los métodos descritos en el presente documento después del procesamiento de la(s) muestra(s) que contiene(n) el ácido nucleico. Por ejemplo, un ácido nucleico se puede extraer, aislar, purificar, purificar parcialmente o amplificar a partir de la(s) muestra(s). El término "aislado", como se usa en el presente documento, se refiere al ácido nucleico eliminado de su entorno original (por ejemplo, el entorno natural si es de origen natural, o una

célula hospedadora si se expresa de forma exógena), y por lo tanto se altera por la intervención humana (por ejemplo, "por la mano del hombre ") de su entorno original. La expresión "ácido nucleico aislado" como se usa en el presente documento puede referirse a un ácido nucleico eliminado de un sujeto (por ejemplo, un sujeto humano). Se puede proporcionar un ácido nucleico aislado con menos componentes no-ácido nucleico (por ejemplo, proteína, lípido) que la cantidad de componentes presentes en una muestra fuente. Una composición que comprende ácido nucleico aislado puede ser aproximadamente del 50% a mayor que el 99% libre de componentes no-ácido nucleico. Una composición que comprende ácido nucleico aislado puede ser de aproximadamente el 90%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99% o mayor que el 99% libre de componentes no-ácido nucleico. El término "purificado" como se usa en el presente documento puede referirse a un ácido nucleico siempre que contenga menos componentes no-ácido nucleico (por ejemplo, proteína, lípido, carbohidrato) que la cantidad de componentes no-ácido nucleico presentes antes de someter el ácido nucleico a un procedimiento de purificación. Una composición que comprende ácido nucleico purificado puede ser de aproximadamente el 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99% o mayor que el 99% libre de otros componentes no-ácido nucleico. El término "purificado" como se usa en el presente documento puede referirse a un ácido nucleico siempre que contenga menos especies de ácido nucleico que en la fuente de muestra de la que deriva el ácido nucleico. Una composición que comprende ácido nucleico purificado puede ser de aproximadamente el 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99% o mayor que el 99 % libre de otras especies de ácidos nucleicos. Por ejemplo, el ácido nucleico fetal se puede purificar a partir de una mezcla que comprende ácido nucleico materno y fetal. En algunos ejemplos, los nucleosomas que comprenden pequeños fragmentos de ácido nucleico fetal pueden purificarse a partir de una mezcla de complejos de nucleosomas más grandes que comprenden fragmentos más grandes de ácido nucleico materno.

En algunos casos, los ácidos nucleicos se fragmentan o escinden antes, durante o después de un método descrito en el presente documento. El ácido nucleico fragmentado o escindido puede tener una longitud nominal, promedio o media de aproximadamente 5 a aproximadamente 10.000 pares de bases, aproximadamente 100 a aproximadamente 1.000 pares de bases, aproximadamente 100 a aproximadamente 500 pares de bases, o aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 o 9000 pares de bases. Los fragmentos pueden generarse mediante un método adecuado conocido en la técnica, y la longitud promedio, media o nominal de los fragmentos de ácido nucleico puede controlarse seleccionando un procedimiento apropiado de generación de fragmentos.

Los fragmentos de ácido nucleico pueden contener secuencias de nucleótidos solapantes, y tales secuencias solapantes pueden facilitar la construcción de una secuencia de nucleótidos del ácido nucleico homólogo no fragmentado, o un segmento del mismo. Por ejemplo, un fragmento puede tener subsecuencias x e y y otro fragmento puede tener subsecuencias y y z, donde x, y y z son secuencias de nucleótidos que pueden tener una longitud de 5 nucleótidos o mayor. En ciertos casos, la secuencia de solapamiento y puede utilizarse para facilitar la construcción de la secuencia de nucleótidos x-y-z en el ácido nucleico a partir de una muestra. En ciertos casos, el ácido nucleico puede estar parcialmente fragmentado (por ejemplo, a partir de una reacción de escisión específica incompleta o terminada) o completamente fragmentado.

En algunos casos, el ácido nucleico se fragmenta o escinde mediante un método adecuado, cuyos ejemplos no limitantes incluyen métodos físicos (p.ej., ruptura mecánica, por ejemplo, sonicación, prensa francesa, calor, irradiación UV, similares), procesos enzimáticos (por ejemplo, agentes de escisión enzimática (p.ej., una nucleasa adecuada, una enzima de restricción adecuada, una enzima de restricción sensible a la metilación adecuada)), métodos químicos (por ejemplo, alquilación, DMS, piperidina, hidrólisis ácida, hidrólisis básica, calor, similares o combinaciones de los mismos), procesos descritos en la Publicación de Solicitud de patente de EE. UU. N.º 20050112590, similares o combinaciones de los mismos.

Como se usa en el presente documento, "fragmentación" o "escisión" se refiere a un procedimiento o condiciones en las que una molécula de ácido nucleico, tal como una molécula de gen molde de ácido nucleico o un producto amplificado de la misma, puede dividirse en dos o más moléculas de ácido nucleico más pequeñas. Dicha fragmentación o escisión puede ser específica de secuencia, específica de base o inespecífica, y puede lograrse mediante cualquiera de una variedad de métodos, reactivos o condiciones, incluyendo, por ejemplo, fragmentación química, enzimática o física.

Como se usa en el presente documento, "fragmentos", "productos de escisión", "productos escindidos" o variantes gramaticales de los mismos, se refieren a moléculas de ácido nucleico resultantes de una fragmentación o escisión de una molécula de gen del molde de ácido nucleico o producto amplificado de la misma. Mientras que tales fragmentos o productos escindidos pueden referirse a todas las moléculas de ácido nucleico resultantes de una reacción de escisión, generalmente tales fragmentos o productos escindidos se refieren solo a moléculas de ácido nucleico resultantes de una fragmentación o escisión de una molécula de gen del molde de ácido nucleico o el segmento de un producto amplificado de la misma que contiene la secuencia de nucleótidos correspondiente de una molécula de gen del molde de ácido nucleico. El término "amplificado" como se usa en el presente documento se refiere a someter un ácido nucleico diana en una muestra a un proceso que genera lineal o exponencialmente ácidos nucleicos de amplificación que tienen la misma o sustancialmente la misma secuencia de nucleótidos que el ácido nucleico diana, o segmento de los mismos. En ciertos casos, el término "amplificado" se refiere a un método que comprende

- una reacción en cadena de la polimerasa (PCR). Por ejemplo, un producto amplificado puede contener uno o más nucleótidos más que la región de nucleótidos amplificada de una secuencia del molde de ácido nucleico (por ejemplo, un cebador puede contener nucleótidos "extra", como una secuencia de iniciación transcripcional, además de nucleótidos complementarios a una molécula de gen del molde de ácido nucleico, dando como resultado un producto amplificado que contiene nucleótidos "extra" o nucleótidos que no corresponden a la región de nucleótidos amplificada de la molécula de gen del molde de ácido nucleico). Por consiguiente, los fragmentos pueden incluir fragmentos que surgen de segmentos o partes de moléculas de ácido nucleico amplificadas que contienen, al menos en parte, la información de la secuencia de nucleótidos de o basada en la molécula de molde de ácido nucleico representativa.
- 5
- 10 Como se usa en el presente documento, la expresión "reacciones de escisión complementarias" se refiere a las reacciones de escisión que se llevan a cabo en el mismo ácido nucleico usando diferentes reactivos de escisión o alterando la especificidad de escisión del mismo reactivo de escisión de tal manera que se generan patrones de escisión alternativos de la misma diana o ácido nucleico de referencia o proteína. En ciertos casos, el ácido nucleico puede tratarse con uno o más agentes de escisión específicos (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más agentes de escisión específicos) en uno o más vasos de reacción (por ejemplo, el ácido se trata con cada agente de escisión específico en un vaso separado). La expresión "agente de escisión específico", como se usa en el presente documento, se refiere a un agente, a veces un producto químico o una enzima que puede escindir un ácido nucleico en uno o más sitios específicos.
- 15
- 20 El ácido nucleico también puede exponerse a un proceso que modifica ciertos nucleótidos en el ácido nucleico antes de proporcionar ácido nucleico para un método descrito en el presente documento. Puede aplicarse, al ácido nucleico, un proceso que modifica selectivamente el ácido nucleico basándose en el estado de metilación de los nucleótidos en el mismo, por ejemplo. Además, condiciones tales como alta temperatura, radiación ultravioleta, radiación X, pueden inducir cambios en la secuencia de una molécula de ácido nucleico. El ácido nucleico se puede proporcionar en cualquier forma adecuada útil para realizar un análisis de secuencia adecuado.
- 25

El ácido nucleico puede ser monocatenario o bicatenario. El ADN monocatenario, por ejemplo, puede generarse desnaturizando ADN bicatenario mediante calentamiento o por tratamiento con álcali, por ejemplo. En ciertos casos, el ácido nucleico está en una estructura de bucle D, formada por la invasión de la cadena de una molécula de ADN dúplex por un oligonucleótido o una molécula similar a ADN, como el ácido nucleico peptídico (ANP). La formación del bucle D se puede facilitar mediante la adición de la proteína RecA de *E. Coli* y/o mediante la alteración de la concentración de sal, por ejemplo, utilizando métodos conocidos en la técnica.

30

Dianas genómicas

35

- En algunos casos, los ácidos nucleicos diana, también denominados en el presente documento fragmentos diana, incluyen fragmentos de polinucleótidos de una región genómica particular o una pluralidad de regiones genómicas (por ejemplo, cromosoma único, conjunto de cromosomas y/o ciertas regiones cromosómicas). En algunos casos, dichas regiones genómicas pueden asociarse con anomalías genéticas fetales (p. ej., Aneuploidía), así como con otras variaciones genéticas que incluyen, pero sin limitación, mutaciones (p. ej., mutaciones puntuales), inserciones, adiciones, deleciones, translocaciones, trastornos de repetición de trinucleótidos y/o polimorfismos de un solo nucleótido (SNP). En algunos casos, los ácidos nucleicos de referencia, también denominados en el presente documento fragmentos de referencia, incluyen fragmentos de polinucleótidos de una región genómica particular o una pluralidad de regiones genómicas no asociadas con anomalías genéticas fetales. En algunos casos, los ácidos nucleicos diana y/o de referencia (es decir, fragmentos diana y/o fragmentos de referencia) comprenden secuencias de nucleótidos que son sustancialmente únicas para el cromosoma de interés o cromosoma de referencia (p. ej., secuencias de nucleótidos idénticas o secuencias de nucleótidos sustancialmente similares que no se encuentran en ninguna otra parte en el genoma).
- 40
- 45
- 50 En algunos casos, se analizan fragmentos de una pluralidad de regiones genómicas. En algunos casos, se analizan los fragmentos diana y los fragmentos de referencia de una pluralidad de regiones genómicas. En algunos casos, los fragmentos de una pluralidad de regiones genómicas se analizan para determinar la presencia, ausencia, cantidad (por ejemplo, cantidad relativa) o proporción de un cromosoma de interés, por ejemplo. En algunos casos, un cromosoma de interés es un cromosoma sospechoso de ser aneuploide y se puede denominar en el presente documento un "cromosoma de ensayo". En algunos casos, se analizan los fragmentos de una pluralidad de regiones genómicas para un supuesto cromosoma euploide. Un cromosoma de este tipo puede denominarse en el presente documento un "cromosoma de referencia". En algunos casos, puede analizarse una pluralidad de cromosomas de ensayo. En algunos casos, los cromosomas de ensayo se seleccionan entre el cromosoma 13, el cromosoma 18 y el cromosoma 21. En algunos casos, los cromosomas de referencia se seleccionan entre los cromosomas 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X e Y, y algunas veces, los cromosomas de referencia se seleccionan de autosomas (es decir, no X ni Y). En algunos casos, el cromosoma 20 se selecciona como un cromosoma de referencia. En algunos casos, el cromosoma 14 se selecciona como un cromosoma de referencia. En algunos casos, el cromosoma 9 se selecciona como un cromosoma de referencia. En algunos casos, un cromosoma de ensayo y un cromosoma de referencia son del mismo individuo. En algunos casos, un cromosoma de ensayo y un cromosoma de referencia son de diferentes individuos.
- 55
- 60
- 65

En algunos casos, los fragmentos de al menos una región genómica se analizan para un cromosoma de ensayo y/o de referencia. En algunos casos, los fragmentos de al menos 10 regiones genómicas (por ejemplo, aproximadamente 20, 30, 40, 50, 60, 70, 80 o 90 regiones genómicas) se analizan para un cromosoma de ensayo y/o un cromosoma de referencia. En algunos casos, los fragmentos de al menos 100 regiones genómicas (por ejemplo, aproximadamente 200, 300, 400, 500, 600, 700, 800 o 900 regiones genómicas) se analizan para un cromosoma de ensayo y/o un cromosoma de referencia. En algunos casos, los fragmentos de al menos 1.000 regiones genómicas (por ejemplo, aproximadamente 2000, 3000, 4000, 5000, 6000, 7000, 8000 o 9000 regiones genómicas) se analizan para un cromosoma de ensayo y/o un cromosoma de referencia. En algunos casos, los fragmentos de al menos 10.000 regiones genómicas (por ejemplo, aproximadamente 20.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000 o 90.000 regiones genómicas) se analizan para un cromosoma de ensayo y/o un cromosoma de referencia. En algunos casos, los fragmentos de al menos 100.000 regiones genómicas (por ejemplo, aproximadamente 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000 o 900.000 regiones genómicas) se analizan para un cromosoma de ensayo y/o un cromosoma de referencia.

15 *Determinación del contenido de ácido nucleico fetal*

En algunos casos, se determina la cantidad de ácido nucleico fetal (por ejemplo, concentración, cantidad relativa, cantidad absoluta, número de copias y similares) en ácido nucleico. En ciertos casos, la cantidad de ácido nucleico fetal en una muestra se denomina "fracción fetal". En algunos casos, "fracción fetal" se refiere a la fracción de ácido nucleico fetal en el ácido nucleico circulante libre de células en una muestra (por ejemplo, una muestra de sangre, una muestra de suero, una muestra de plasma) obtenida de una mujer gestante. En algunos casos, un método en donde se determina una variación genética también puede comprender determinar la fracción fetal. La determinación de la fracción fetal se puede realizar de una manera adecuada, cuyos ejemplos no limitativos incluyen los métodos que se describen a continuación.

En algunos casos, la fracción fetal se puede determinar, usando los métodos descritos en este documento para determinar la longitud del fragmento. Los fragmentos de ácido nucleico fetal libre de células generalmente son más cortos que los fragmentos de ácido nucleico derivados de la madre (véase, por ejemplo, Chan et al. (2004) Clin. Chem. 50: 88-92; Lo et al. (2010) Sci. Transl. Med 2:61 ra91). Por lo tanto, la fracción fetal se puede determinar, en algunos casos, contando fragmentos bajo un umbral de longitud particular y comparando los recuentos con la cantidad de ácido nucleico total en la muestra. Los métodos para contar fragmentos de ácido nucleico de una longitud particular se describen con más detalle a continuación.

En ciertos casos, la cantidad de ácido nucleico fetal se determina según los marcadores específicos de un feto masculino (por ejemplo, marcadores STR del cromosoma Y (por ejemplo, marcadores DYS 19, DYS 385, DYS 392); marcador RhD en mujeres RhD negativas), relaciones alélicas de secuencias polimórficas, o de acuerdo con uno o más marcadores específicos de ácido nucleico fetal y no de ácido nucleico materno (por ejemplo, biomarcadores epigenéticos diferenciales (por ejemplo, metilación; descritos con más detalle a continuación) entre la madre y el feto, o marcadores de ARN fetal en plasma sanguíneo materno (véase, por ejemplo, Lo, 2005, Journal of Histochemistry and Cytochemistry 53 (3): 293-296)).

La determinación del contenido de ácido nucleico fetal (por ejemplo, fracción fetal) a veces se realiza usando un ensayo de cuantificación fetal (FQA) como se describe, por ejemplo, en la publicación de solicitud de patente de Estados Unidos nº 2010/0105049. Este tipo de ensayo permite la detección y cuantificación de ácido nucleico fetal en una muestra materna en función del estado de metilación del ácido nucleico en la muestra. En ciertos casos, la cantidad de ácido nucleico fetal de una muestra materna se puede determinar en relación con la cantidad total de ácido nucleico presente, proporcionando así el porcentaje de ácido nucleico fetal en la muestra. En ciertos casos, el número de copias de ácido nucleico fetal se puede determinar en una muestra materna. En ciertos casos, la cantidad de ácido nucleico fetal se puede determinar de una manera específica de secuencia (o de una porción específica) y, a veces, con suficiente sensibilidad para permitir un análisis de dosis cromosómico preciso (por ejemplo, para detectar la presencia o ausencia de una aneuploidía fetal).

Se puede realizar un ensayo de cuantificación fetal (FQA) junto con cualquiera de los métodos descritos en el presente documento. Dicho ensayo se puede realizar mediante cualquier método conocido en la técnica y/o descrito en la publicación de solicitud de patente de Estados Unidos nº 2010/0105049, tal como, por ejemplo, mediante un método que pueda distinguir entre el ADN materno y fetal en función del estado de metilación diferencial y cuantificar (es decir, determinar la cantidad de) el ADN fetal. Métodos para diferenciar el ácido nucleico basado en el estado de metilación incluye, pero no se limita a, captura sensible a la metilación, por ejemplo, utilizando un fragmento MBD2-Fc en donde el dominio de unión a metilo de MBD2 se fusiona con el fragmento Fc de un anticuerpo (MBD-Fc) (Gebhard et al. (2006) Cancer Res. 66 (12): 6118-28); anticuerpos específicos de metilación; métodos de conversión de bisulfito, por ejemplo, MSP (PCR sensible a metilación), COBRA, extensión de cebador de un solo nucleótido sensible a metilación (Ms-SNuPE) o tecnología Sequenom MassCLEAVE™; y el uso de enzimas de restricción sensibles a metilación (por ejemplo, digestión de ADN materno en una muestra materna usando una o más enzimas de restricción sensibles a metilación, enriqueciendo así el ADN fetal). Las enzimas sensibles a metilo también pueden usarse para diferenciar el ácido nucleico basándose en el estado de metilación, que, por ejemplo, pueden escindir o digerir preferencial o sustancialmente en su secuencia de reconocimiento de ADN si este último no está metilado. Por lo tanto, una muestra

de ADN no metilada se cortará en fragmentos más pequeños que una muestra de ADN metilada y una muestra de ADN hipermetilada no se escindirán. Excepto donde se indique explícitamente, cualquier método para diferenciar el ácido nucleico basándose en el estado de metilación se puede usar con las composiciones y métodos de la tecnología en el presente documento. La cantidad de ADN fetal se puede determinar, por ejemplo, introduciendo uno o más competidores a concentraciones conocidas durante una reacción de amplificación. La determinación de la cantidad de ADN fetal también puede realizarse, por ejemplo, por RT-PCR, extensión de cebadores, secuenciación y/o recuento. En ciertos casos, la cantidad de ácido nucleico se puede determinar usando la tecnología BEAMing como se describe en la publicación de solicitud de patente de Estados Unidos n° 2007/0065823. En ciertos casos, la eficacia de restricción se puede determinar y la tasa de eficacia se utiliza para determinar adicionalmente la cantidad de ADN fetal.

En ciertos casos, se puede usar un ensayo de cuantificación fetal (FQA) para determinar la concentración de ADN fetal en una muestra materna, por ejemplo, mediante el siguiente método: a) determinar la cantidad total de ADN presente en una muestra materna; b) digerir selectivamente el ADN materno en una muestra materna utilizando una o más enzimas de restricción sensibles a metilación, enriqueciendo así el ADN fetal; c) determinar la cantidad de ADN fetal de la etapa b); y d) comparar la cantidad de ADN fetal de la etapa c) con la cantidad total de ADN de la etapa a), determinando así la concentración de ADN fetal en la muestra materna. En ciertos casos, el número absoluto de copias de ácido nucleico fetal en una muestra materna puede determinarse, por ejemplo, usando espectrometría de masas y/o un sistema que utiliza un enfoque competitivo de PCR para mediciones de números absolutos de copias. Véanse, por ejemplo, Ding y Cantor (2003) PNAS USA 100: 3059-3064, y la publicación de solicitud de patente de Estados Unidos n° 2004/0081993.

En ciertos casos, la fracción fetal se puede determinar basándose en las relaciones alélicas de las secuencias polimórficas (por ejemplo, polimorfismos de un solo nucleótido (SNP)), tal como, por ejemplo, el uso de un método descrito en la publicación de solicitud de patente de Estados Unidos n° 2011/0224087. En dicho método, las lecturas de secuencia de nucleótidos se obtienen para una muestra materna y la fracción fetal se determina comparando el número total de lecturas de secuencias de nucleótidos que mapean en un primer alelo y el número total de lecturas de secuencias de nucleótidos que mapean en un segundo alelo en un sitio polimórfico informativo (por ejemplo, SNP) en un genoma de referencia. En ciertos casos, los alelos fetales se identifican, por ejemplo, por su contribución relativamente menor a la mezcla de ácidos nucleicos fetales y maternos en la muestra en comparación con la contribución principal a la mezcla de los ácidos nucleicos maternos. Por consiguiente, la abundancia relativa de ácido nucleico fetal en una muestra materna se puede determinar como un parámetro del número total de lecturas de secuencias únicas mapeadas en una secuencia de ácido nucleico diana en un genoma de referencia para cada uno de los dos alelos de un sitio polimórfico.

La cantidad de ácido nucleico fetal en ácido nucleico extracelular puede cuantificarse y usarse junto con un método proporcionado en este documento. Por lo tanto, en ciertos casos, los métodos de la tecnología descritos en el presente documento comprenden una etapa adicional para determinar la cantidad de ácido nucleico fetal. La cantidad de ácido nucleico fetal se puede determinar en una muestra de ácido nucleico de un sujeto antes o después del procesamiento para preparar una muestra de ácido nucleico. En ciertos casos, la cantidad de ácido nucleico fetal se determina en una muestra después de que el ácido nucleico de la muestra se procese y prepare, cantidad que se utiliza para una evaluación adicional. En algunos casos, un resultado comprende factorizar la fracción de ácido nucleico fetal en el ácido nucleico de la muestra (por ejemplo, ajustar los recuentos, eliminar muestras, hacer o no una selección).

La etapa de determinación puede realizarse antes, durante, en cualquier punto de un método descrito en este documento, o después de ciertos métodos (por ejemplo, detección de aneuploidía, determinación de género fetal) descritos en este documento. Por ejemplo, para lograr un método de determinación de género o aneuploidía fetal con una sensibilidad o especificidad dada, se puede implementar un método de cuantificación de ácido nucleico fetal antes, durante o después de la determinación de sexo o aneuploidía fetal para identificar esas muestras con más de aproximadamente 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17 %, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25% o más de ácido nucleico fetal. En algunos casos, las muestras en las que se determina que tienen una cierta cantidad umbral de ácido nucleico fetal (por ejemplo, aproximadamente 15% o más de ácido nucleico fetal; aproximadamente 4% o más de ácido nucleico fetal) se analizan adicionalmente para determinar el género fetal o la aneuploidía, o la presencia o ausencia de aneuploidía o variación genética, por ejemplo. En ciertos casos, las determinaciones de, por ejemplo, el género fetal o la presencia o ausencia de aneuploidía se seleccionan (por ejemplo, se seleccionan y se comunican a un paciente) solo para muestras que tienen una cierta cantidad umbral de ácido nucleico fetal (por ejemplo, aproximadamente 15% o más de ácido nucleico fetal; aproximadamente 4% o más de ácido nucleico fetal).

En algunos casos, la determinación de la fracción fetal o la determinación de la cantidad de ácido nucleico fetal no es necesaria o no se requiere para identificar la presencia o la ausencia de una aneuploidía cromosómica. En algunos casos, identificar la presencia o ausencia de una aneuploidía cromosómica no requiere la diferenciación de secuencia del ADN fetal frente al materno. En ciertos casos, esto se debe a que se analiza la suma de la contribución de las secuencias materna y fetal en un determinado cromosoma, la porción cromosómica o segmento del mismo. En algunos casos, identificar la presencia o ausencia de una aneuploidía cromosómica no se basa en información de secuencia a priori que distinga el ADN fetal del ADN materno.

Enriquecimiento y separación de subpoblaciones de ácido nucleico

En algunos casos, el ácido nucleico (por ejemplo, el ácido nucleico extracelular) está enriquecido o relativamente enriquecido para una subpoblación o especie de ácido nucleico. Las subpoblaciones de ácido nucleico pueden incluir, por ejemplo, ácido nucleico fetal, ácido nucleico materno, ácido nucleico que comprende fragmentos de una longitud o intervalo de longitudes particulares, o ácido nucleico de una región del genoma particular (por ejemplo, cromosoma único, conjunto de cromosomas y/o ciertas regiones cromosómicas). Dichas muestras enriquecidas se pueden usar junto con un método que se proporciona en el presente documento. Por lo tanto, en ciertos casos, los métodos de la tecnología comprenden una etapa adicional de enriquecimiento para una subpoblación de ácido nucleico en una muestra, tal como, por ejemplo, ácido nucleico fetal. En ciertos casos, un método para determinar la fracción fetal descrito anteriormente también se puede usar para enriquecer el ácido nucleico fetal. En ciertos casos, el ácido nucleico materno se elimina de manera selectiva (parcial, sustancialmente, casi completamente o completamente) de la muestra. En ciertos casos, el enriquecimiento para un ácido nucleico de especies con un bajo número de copias particular (por ejemplo, ácido nucleico fetal) puede mejorar la sensibilidad cuantitativa. Los métodos para enriquecer una muestra para una especie particular de ácido nucleico se describen, por ejemplo, en la Patente de Estados Unidos n.º 6.927.028, la Publicación de Solicitud de Patente Internacional N.º WO2007/140417, la Publicación de Solicitud de Patente Internacional N.º WO2007/147063, la Publicación de Solicitud de Patente Internacional N.º WO2009/032779, la Publicación de Solicitud de Patente Internacional N.º WO2009/032781, la Publicación de Solicitud de Patente Internacional N.º WO2010/033639, la Publicación de Solicitud de Patente Internacional N.º WO2011/034631, la Publicación de Solicitud de Patente Internacional N.º WO2006/056480 y la Publicación de Solicitud de Patente Internacional N.º WO2011/143659.

En algunos casos, el ácido nucleico se enriquece para ciertas especies de fragmentos diana y/o especies de fragmentos de referencia. En ciertos casos, el ácido nucleico se enriquece para una longitud de fragmentos de ácidos nucleicos o un intervalo de longitudes de fragmentos específicos utilizando uno o más métodos de separación basados en la longitud descritos a continuación. En ciertos casos, el ácido nucleico se enriquece para fragmentos de una región genómica seleccionada (por ejemplo, un cromosoma) utilizando uno o más métodos de separación basados en secuencias descritos en el presente documento y/o conocidos en la técnica. Ciertos métodos de enriquecimiento para una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) en una muestra se describen en detalle a continuación.

Algunos métodos de enriquecimiento para una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que se pueden usar con un método descrito en el presente documento incluyen métodos que aprovechan las diferencias epigenéticas entre el ácido nucleico materno y fetal. Por ejemplo, el ácido nucleico fetal se puede diferenciar y separar del ácido nucleico materno en función de las diferencias de metilación. Los métodos de enriquecimiento de ácido nucleico fetal basados en la metilación se describen en la Publicación de Solicitud de Patente de Estados Unidos número 2010/0105049. Tales métodos a veces implican la unión de un ácido nucleico de muestra a un agente de unión específico de la metilación (proteína de unión a metil-CpG (MBD), anticuerpos específicos de metilación y similares) y la separación del ácido nucleico unido del ácido nucleico no unido en función del estado de metilación diferencial. Tales métodos también pueden incluir el uso de enzimas de restricción sensibles a la metilación (como se describe anteriormente; por ejemplo, HhaI y HpaII), que permiten el enriquecimiento de las regiones de ácido nucleico fetal en una muestra materna mediante la digestión selectiva del ácido nucleico de la muestra materna con un enzima que digiere selectiva y completamente o sustancialmente el ácido nucleico materno para enriquecer la muestra en al menos una región de ácido nucleico fetal.

Otro método de enriquecimiento para una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que se puede usar con un método descrito en el presente documento es una aproximación de secuencia polimórfica mejorada por endonucleasas de restricción, tal como un método descrito en la Publicación de Solicitud de Patente de Estados Unidos N.º 2009/0317818. Tales métodos incluyen la escisión del ácido nucleico que comprende un alelo no diana con una endonucleasa de restricción que reconoce el ácido nucleico que comprende el alelo no-diana pero no el alelo diana; y la amplificación del ácido nucleico no-escindido pero no el ácido nucleico escindido, donde el ácido nucleico amplificado no-escindido representa ácido nucleico diana enriquecido (por ejemplo, ácido nucleico fetal) en relación con el ácido nucleico no-diana (por ejemplo, ácido nucleico materno). En ciertos casos, el ácido nucleico puede seleccionarse de manera que comprenda un alelo que tenga un sitio polimórfico que sea susceptible de digestión selectiva por un agente de escisión, por ejemplo.

Algunos métodos de enriquecimiento para una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que se pueden usar con un método descrito en el presente documento incluye aproximaciones de degradación enzimática selectiva. Dichos métodos implican proteger las secuencias diana de la digestión por exonucleasas, facilitando, de este modo, la eliminación de secuencias no deseadas en una muestra (por ejemplo, ADN materno). Por ejemplo, en una aproximación, el ácido nucleico de la muestra se desnaturaliza para generar ácido nucleico monocatenario, el ácido nucleico monocatenario se pone en contacto con al menos un par de cebadores específicos de diana en condiciones de hibridación adecuadas, los cebadores hibridados se extienden por polimerización de nucleótidos generando secuencias diana bicatenarias, y digiriendo el ácido nucleico monocatenario utilizando una nucleasa que digiere ácido nucleico monocatenario (es decir, no-diana). En ciertos casos, el método se puede repetir durante al menos un ciclo adicional. En ciertos casos, se usa el mismo par de cebadores específicos de diana para cebar cada

uno del primer y segundo ciclos de extensión y, en ciertos casos, se usan diferentes pares de cebadores específicos de diana para el primer y segundo ciclos.

En algunos casos, el ácido nucleico se enriquece para fragmentos de una región genómica seleccionada (por ejemplo, un cromosoma) utilizando uno o más métodos de separación basados en secuencias descritos en el presente documento. En algunos casos, el ácido nucleico se enriquece para una longitud o un intervalo de longitudes de fragmento de polinucleótido específico y para fragmentos de una región genómica seleccionada (por ejemplo, un cromosoma) usando una combinación de métodos de separación basados en la longitud y basados en la secuencia. Dichos métodos de separación basados en la longitud y en la secuencia se describen con más detalle a continuación.

Algunos métodos de enriquecimiento para una subpoblación de ácidos nucleicos (por ejemplo, ácido nucleico fetal) que se pueden usar con un método descrito en el presente documento, incluyen aproximaciones de secuenciación de firma masiva en paralelo (MPSS). La MPSS generalmente es un método de fase sólida que utiliza la unión del adaptador (es decir, marcador), seguido de la decodificación del adaptador y la lectura de la secuencia de ácido nucleico en pequeños incrementos. Los productos de PCR marcados se amplifican, generalmente, de manera que cada ácido nucleico genera un producto de PCR con un marcador único. Los marcadores se utilizan a menudo para unir los productos de PCR a microperlas. Después de varias rondas de determinación de secuencia basada en la unión, por ejemplo, se puede identificar una firma de secuencia de cada perla. Se analiza cada secuencia de firma (marcador MPSS) en un conjunto de datos MPSS, en comparación con todas las demás firmas, y se cuentan todas las firmas idénticas.

En ciertos casos, ciertos métodos de enriquecimiento (por ejemplo, ciertos métodos de enriquecimiento basados en MPS y/o MPSS) pueden incluir aproximaciones basadas en la amplificación (por ejemplo, PCR). En ciertos casos, se pueden usar métodos de amplificación específicos de loci (por ejemplo, utilizando cebadores de amplificación específicos de loci). En ciertos casos, se puede utilizar una aproximación de PCR de alelo de SNP multiplex. En ciertos casos, se puede utilizar una aproximación de PCR de alelo de SNP multiplex en combinación con la secuenciación uniplex. Por ejemplo, dicha aproximación puede implicar el uso de PCR multiplex (por ejemplo, el sistema MASSARRAY) y la incorporación de secuencias de sonda de captura en los amplicones, seguido de una secuenciación utilizando, por ejemplo, el sistema MPSS de Illumina. En ciertos casos, se puede utilizar una aproximación de PCR de alelo SNP multiplex en combinación con un sistema de tres cebadores y secuenciación indexada. Por ejemplo, tal aproximación puede implicar el uso de PCR multiplex (por ejemplo, el sistema MASSARRAY) con cebadores que tienen una primera sonda de captura incorporada en ciertos cebadores de PCR directos específicos de loci y secuencias adaptadoras incorporadas en cebadores de PCR inversos específicos de loci, para generar, de este modo, amplicones, seguida por una PCR secundaria para incorporar secuencias de captura inversa y códigos de barras de índice molecular para la secuenciación utilizando, por ejemplo, el sistema MPSS de Illumina. En ciertos casos, se puede utilizar una aproximación de PCR de alelo SNP multiplex en combinación con un sistema de cuatro cebadores y secuenciación indexada. Por ejemplo, tal aproximación puede implicar el uso de PCR multiplex (por ejemplo, el sistema MASSARRAY) con cebadores que tienen secuencias adaptadoras incorporadas en los cebadores de PCR tanto específicos de loci directos como específicos de loci inversos, seguida por una PCR secundaria para incorporar secuencias de captura tanto directa como inversa y códigos de barras de índice molecular para la secuenciación utilizando, por ejemplo, el sistema MPSS de Illumina. En ciertos casos, se puede utilizar una aproximación de microfluidos. En ciertos casos, se puede utilizar una aproximación de microfluidos basada en matrices. Por ejemplo, dicha aproximación puede implicar el uso de una matriz de microfluidos (por ejemplo, Fluidigm) para la amplificación con baja complejidad y la incorporación de sondas de índice y captura, seguido de secuenciación. En ciertos casos, se puede utilizar una aproximación de microfluidos en emulsión, tal como, por ejemplo, PCR digital en gotitas.

En ciertos casos, se pueden usar métodos de amplificación universal (por ejemplo, utilizando cebadores de amplificación universales o no específicos de loci). En ciertos casos, los métodos de amplificación universal se pueden utilizar en combinación con las aproximaciones descendentes (*pull down*). En ciertos casos, un método puede incluir el análisis descendente de ultrámero biotinilado (por ejemplo, ensayos descendentes de Agilent o IDT) a partir de una biblioteca de secuenciación amplificada universalmente. Por ejemplo, tal aproximación puede implicar la preparación de una biblioteca convencional, el enriquecimiento para regiones seleccionadas mediante un ensayo descendente y una etapa de amplificación universal secundaria. En ciertos casos, las aproximaciones descendentes se pueden utilizar en combinación con métodos basados en la unión. En ciertos casos, un método puede incluir un ensayo descendente de ultrámeros biotinilados con unión al adaptador específico de secuencia (por ejemplo, PCR HALOPLEX, Halo Genomics). Por ejemplo, dicha aproximación puede implicar el uso de sondas de selección para capturar fragmentos digeridos con enzimas de restricción, seguido de la unión de los productos capturados a un adaptador, y la amplificación universal seguida de la secuenciación. En ciertos casos, las aproximaciones descendentes se pueden utilizar en combinación con métodos basados en la extensión y unión. En ciertos casos, un método puede incluir la extensión y la unión de la sonda de inversión molecular (MIP). Por ejemplo, tal aproximación puede implicar el uso de sondas de inversión molecular en combinación con adaptadores de secuencia seguidos de amplificación y secuenciación universales. En ciertos casos, el ADN complementario se puede sintetizar y secuenciar sin amplificación.

En ciertos casos, las aproximaciones de extensión y unión se pueden realizar sin un componente descendente. En ciertos casos, un método puede incluir la hibridación, extensión y unión de cebadores directos e inversos específicos

de loci. Dichos métodos pueden incluir además amplificación universal o síntesis de ADN complementario sin amplificación, seguido de secuenciación. En ciertos casos, tales métodos pueden reducir o excluir secuencias de fondo durante el análisis.

5 En ciertos casos, se pueden usar aproximaciones descendentes con un componente de amplificación opcional o sin componente de amplificación. En ciertos casos, un método puede incluir un ensayo descendente modificado y unión con incorporación completa de sondas de captura sin amplificación universal. Por ejemplo, dicha aproximación puede implicar el uso de sondas de selección modificadas para capturar fragmentos digeridos con enzimas de restricción, seguido de la unión de los productos capturados a un adaptador, amplificación opcional y secuenciación. En ciertos
10 casos, un método puede incluir un ensayo descendente biotinilado con extensión y unión de la secuencia adaptadora en combinación con la unión circular de una sola hebra. Por ejemplo, dicha aproximación puede implicar el uso de sondas de selección para capturar regiones de interés (es decir, secuencias diana), extensión de las sondas, unión del adaptador, unión circular de una sola hebra, amplificación opcional y secuenciación. En ciertos casos, el análisis del resultado de la secuenciación puede separar las secuencias diana del fondo.

15 En algunos casos, el ácido nucleico se enriquece para fragmentos de una región genómica seleccionada (por ejemplo, un cromosoma) utilizando uno o más métodos de separación basados en secuencias descritos en el presente documento. La separación basada en secuencias generalmente se basa en secuencias de nucleótidos presentes en los fragmentos de interés (por ejemplo, fragmentos diana y/o de referencia) y no presentes, sustancialmente, en otros
20 fragmentos de la muestra o están presentes en una cantidad insustancial de los otros fragmentos (por ejemplo, 5 % o menos). En algunos casos, la separación basada en la secuencia puede generar fragmentos diana separados y/o fragmentos de referencia separados. En ciertos casos, los fragmentos diana separados y/o los fragmentos de referencia separados a menudo se aíslan de los fragmentos restantes en la muestra de ácido nucleico. En ciertos casos, los fragmentos diana separados y los fragmentos de referencia separados también se aíslan unos de otros (por
25 ejemplo, aislarse en compartimentos de ensayo separados). En ciertos casos, los fragmentos diana separados y los fragmentos de referencia separados se aíslan juntos (por ejemplo, se aíslan en el mismo compartimento de ensayo). En algunos casos, los fragmentos sin unir se pueden eliminar, degradar o digerir de forma diferencial.

30 En algunos casos, se usa un proceso selectivo de captura de ácidos nucleicos para separar fragmentos diana y/o de referencia de la muestra de ácido nucleico. Los sistemas de captura de ácidos nucleicos disponibles comercialmente incluyen, por ejemplo, Sistema de captura de secuencias Nimblegen (Roche NimbleGen, Madison, WI); plataforma BEADARRAY de Illumina (Illumina, San Diego, CA); plataforma GENECHIP de Affymetrix (Affymetrix, Santa Clara, CA); Sistema de Enriquecimiento de Dianas SureSelect de Agilent (Agilent Technologies, Santa Clara, CA); y plataformas relacionadas. Tales métodos generalmente implican la hibridación de un oligonucleótido de captura a un
35 segmento o la totalidad de la secuencia de nucleótidos de un fragmento diana o de referencia y puede incluir el uso de una fase sólida (por ejemplo, una matriz de fase sólida) y/o una plataforma basada en disoluciones. Los oligonucleótidos de captura (a veces denominados "cebos") pueden seleccionarse o diseñarse de modo que hibriden preferentemente con fragmentos de ácido nucleico de regiones o loci genómicos seleccionados (por ejemplo, uno de los cromosomas 21, 18, 13, X o Y, o un cromosoma de referencia). En ciertos casos, se puede usar un método basado
40 en hibridación (por ejemplo, utilizando matrices de oligonucleótidos) para enriquecer las secuencias de ácido nucleico de ciertos cromosomas (por ejemplo, un cromosoma potencialmente aneuploide, cromosoma de referencia u otro cromosoma de interés) o segmentos de interés del mismo.

45 Los oligonucleótidos de captura generalmente comprenden una secuencia de nucleótidos capaz de hibridar o emparejar con un fragmento de ácido nucleico de interés (por ejemplo, fragmento diana, fragmento de referencia) o una porción del mismo. Un oligonucleótido de captura puede ser de origen natural o sintético y puede estar basado en ADN o ARN. Los oligonucleótidos de captura pueden permitir la separación específica de, por ejemplo, un fragmento diana y/o de referencia de otros fragmentos en una muestra de ácido nucleico. El término "específico" o "especificidad", como se usa en el presente documento, se refiere a la unión o hibridación de una molécula a otra molécula, tal como
50 un oligonucleótido para un polinucleótido diana. "Específico" o "especificidad" se refiere al reconocimiento, contacto y formación de un complejo estable entre dos moléculas, en comparación con el reconocimiento, contacto o formación de complejos sustancialmente menor de cualquiera de esas dos moléculas con otras moléculas. Como se usa en el presente documento, el término "emparejar" se refiere a la formación de un complejo estable entre dos moléculas. Las expresiones "oligonucleótido de captura", "oligo de captura", "oligo" u "oligonucleótido" se pueden usar indistintamente
55 en todo el documento, cuando se hace referencia a oligonucleótidos de captura. Las siguientes características de los oligonucleótidos se pueden aplicar a los cebadores y otros oligonucleótidos, tales como las sondas proporcionadas en el presente documento.

60 Un oligonucleótido de captura puede diseñarse y sintetizarse usando un proceso adecuado, y puede ser de cualquier longitud adecuada para hibridar con una secuencia de nucleótidos de interés y realizar procesos de separación y/o análisis descritos en el presente documento. Los oligonucleótidos pueden diseñarse basándose en una secuencia de nucleótidos de interés (por ejemplo, secuencia de fragmentos diana, secuencia de fragmentos de referencia). En algunos casos, un oligonucleótido puede ser de aproximadamente 10 a aproximadamente 300 nucleótidos, de aproximadamente 10 a aproximadamente 100 nucleótidos, de aproximadamente 10 a aproximadamente 70
65 nucleótidos, de aproximadamente 10 a aproximadamente 50 nucleótidos, de aproximadamente 15 a aproximadamente 30 nucleótidos, o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50,

55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 nucleótidos de longitud. Un oligonucleótido puede estar compuesto de nucleótidos de origen naturales y/o de origen no natural (por ejemplo, nucleótidos etiquetados), o una mezcla de los mismos. Los oligonucleótidos adecuados para su uso con los casos descritos en el presente documento, pueden sintetizarse y etiquetarse usando técnicas conocidas. Los oligonucleótidos se pueden sintetizar químicamente de acuerdo con el método del triéster de fosoramidita en fase sólida descrito por primera vez por Beaucage y Caruthers (1981) *Tetrahedron Letts.* 22:1859-1862, utilizando un sintetizador automatizado, y/o como se describe en Needham-VanDevanter et al. (1984) *Nucleic Acids Res.* 12: 6159-6168. La purificación de oligonucleótidos puede efectuarse mediante electroforesis en gel de acrilamida natural o mediante cromatografía líquida de intercambio aniónico de alto rendimiento (HPLC), por ejemplo, como se describe en Pearson y Regnier (1983) *J. Chrom.* 255: 137-149.

En algunos casos, toda o una porción de una secuencia de oligonucleótidos (de origen natural o sintética) puede ser sustancialmente complementaria a una secuencia de fragmentos diana y/o de referencia o a una porción de la misma. Como se cita en el presente documento, "sustancialmente complementario" con respecto a las secuencias se refiere a secuencias de nucleótidos que hibridarán entre sí. La rigurosidad de las condiciones de hibridación se puede alterar para tolerar cantidades variables de desapareamientos de secuencia. Se incluyen las secuencias diana/referencia y oligonucleótidos que son un 55% o más, 56 % o más, 57 % o más, 58 % o más, 59 % o más, 60 % o más, 61 % o más, 62 % o más, 63 % o más, 64 % o más, 65 % o más, 66 % o más, 67 % o más, 68 % o más, 69 % o más, 70 % o más, 71 % o más, 72 % o más, 73 % o más, 74 % o más, 75 % o más, 76 % o más, 77 % o más, 78 % o más, 79 % o más, 80 % o más, 81 % o más, 82 % o más, 83 % o más, 84 % o más, 85 % o más, 86 % o más, 87 % o más, 88 % o más, 89 % o más, 90 % o más, 91 % o más, 92 % o más, 93 % o más, 94 % o más, 95 % o más, 96 % o más, 97 % o más, 98 % o más o 99 % o más complementarias entre sí.

Los oligonucleótidos que son sustancialmente complementarios a una secuencia de ácido nucleico de interés (por ejemplo, secuencia de fragmentos diana, secuencia de fragmentos de referencia) o porción de los mismos también son sustancialmente similares al complemento de la secuencia de ácido nucleico diana o una porción relevante de los mismos (por ejemplo, sustancialmente similares a la cadena no codificante del ácido nucleico). Una prueba para determinar si dos secuencias de nucleótidos son sustancialmente similares es determinar el porcentaje de secuencias de nucleótidos idénticas compartidas. Como se cita en el presente documento, "sustancialmente similar" con respecto a las secuencias se refiere a secuencias de nucleótidos que son un 55% o más, 56 % o más, 57 % o más, 58 % o más, 59 % o más, 60 % o más, 61 % o más, 62 % o más, 63 % o más, 64 % o más, 65 % o más, 66 % o más, 67 % o más, 68 % o más, 69 % o más, 70 % o más, 71 % o más, 72 % o más, 73 % o más, 74 % o más, 75 % o más, 76 % o más, 77 % o más, 78 % o más, 79 % o más, 80 % o más, 81 % o más, 82 % o más, 83 % o más, 84 % o más, 85 % o más, 86 % o más, 87 % o más, 88 % o más, 89 % o más, 90 % o más, 91 % o más, 92 % o más, 93 % o más, 94 % o más, 95 % o más, 96 % o más, 97 % o más, 98 % o más o 99 % o más idénticas entre sí.

Las condiciones de emparejamiento (por ejemplo, condiciones de hibridación) se pueden determinar y/o ajustar, dependiendo de las características de los oligonucleótidos utilizados en un ensayo. La secuencia y/o la longitud de los oligonucleótidos pueden afectar, a veces, a la hibridación a una secuencia de ácido nucleico de interés. Dependiendo del grado de desapareamiento entre un oligonucleótido y un ácido nucleico de interés, se pueden usar condiciones de rigurosidad baja, media o alta para efectuar el emparejamiento. Como se usa en el presente documento, la expresión "condiciones rigurosas" se refiere a las condiciones para la hibridación y el lavado. Los métodos para la optimización de las condiciones de temperatura de reacción de hibridación son conocidos en la técnica y se pueden encontrar en *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Los métodos acuosos y no acuosos se describen en esa referencia y se puede usar cualquiera de los dos. Ejemplos no limitantes de condiciones de hibridación rigurosas son la hibridación en 6X de cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45 °C, seguida de uno o más lavados en 0,2 X de SSC, SDS al 0,1% a 50 °C. Otro ejemplo de condiciones de hibridación rigurosas es la hibridación en 6X de cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45 °C, seguida de uno o más lavados en 0,2 X de SSC, SDS al 0,1% a 55 °C. Un ejemplo adicional de condiciones de hibridación rigurosas es la hibridación en 6X de cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45 °C, seguida de uno o más lavados en 0,2 X de SSC, SDS al 0,1% a 60 °C. A menudo, las condiciones de hibridación rigurosas son la hibridación en 6X de cloruro de sodio /citrato de sodio (SSC) a aproximadamente 45 °C, seguida de uno o más lavados en 0,2 X de SSC, SDS al 0,1% a 65 °C. Más a menudo, las condiciones de rigurosidad son fosfato de sodio 0,5 M, SDS al 7% a 65 °C, seguido de uno o más lavados a 0,2X de SSC, SDS al 1% a 65 °C. Las temperaturas de hibridación rigurosas también se pueden alterar (es decir, disminuir) con la adición de ciertos disolventes orgánicos, formamida, por ejemplo. Los disolventes orgánicos, como la formamida, reducen la estabilidad térmica de los polinucleótidos bicatenarios, de modo que la hibridación se puede realizar a temperaturas más bajas, mientras se mantienen las condiciones rigurosas y se prolonga la vida útil de los ácidos nucleicos que pueden ser lábiles al calor.

Como se usa en el presente documento, la frase "hibridación" o variaciones gramaticales de la misma, se refiere al emparejamiento de una primera molécula de ácido nucleico a una segunda molécula de ácido nucleico en condiciones de rigurosidad baja, media o alta, o en condiciones de síntesis de ácidos nucleicos. La hibridación puede incluir casos en los que una primera molécula de ácido nucleico se empareja a una segunda molécula de ácido nucleico, donde la primera y la segunda molécula de ácido nucleico son complementarias. Como se usa en el presente documento, "hibrida específicamente" se refiere a la hibridación preferencial en condiciones de síntesis de ácidos nucleicos de un oligonucleótido a una molécula de ácido nucleico que tiene una secuencia complementaria al oligonucleótido en comparación con la hibridación a una molécula de ácido nucleico que no tiene una secuencia complementaria. Por

ejemplo, la hibridación específica incluye la hibridación de un oligonucleótido de captura con una secuencia de fragmentos diana que es complementaria al oligonucleótido.

5 En algunos casos, uno o más oligonucleótidos de captura se asocian con un ligando de afinidad, como un miembro de un par de unión (por ejemplo, biotina) o un antígeno que puede unirse a un agente de captura, como avidina, estreptavidina, un anticuerpo o un receptor. Por ejemplo, un oligonucleótido de captura puede ser biotinilado de manera que pueda capturar en una perla recubierta con estreptavidina.

10 En algunos casos, uno o más oligonucleótidos de captura y/o agentes de captura se enlazan eficazmente a un soporte sólido o sustrato. Un soporte sólido o sustrato puede ser cualquier sólido físicamente separable al que un oligonucleótido de captura pueda unirse directa o indirectamente, incluyendo, pero sin limitación, superficies proporcionadas por micromatrices y pocillos, y partículas tales como perlas (por ejemplo, perlas paramagnéticas, perlas magnéticas, micropérlas o nanopérlas), micropartículas y nanopartículas. Los soportes sólidos pueden incluir, por ejemplo, chips, columnas, fibras ópticas, toallitas, filtros (por ejemplo, filtros de superficie plana), uno o más capilares, vidrio y vidrio modificado o funcionalizado (por ejemplo, vidrio de poro controlado (CPG)), cuarzo, mica, membranas diazotizadas (papel o nailon), poliformaldehído, celulosa, acetato de celulosa, papel, cerámicas, metales, metaloides, materiales semiconductores, puntos cuánticos, perlas recubiertas o partículas, otros materiales cromatográficos, partículas metálicas; plásticos (incluyendo acrílicos, poliestireno, copolímeros de estireno y otros materiales, polibutileno, poliuretanos, TEFLON™, polietileno, polipropileno, poliamida, poliéster, difluoruro de polivinilideno (PVDF), y similares), polisacáridos, nailon o nitrocelulosa, resinas, sílice o materiales a base de sílice, incluidos sílice, gel de sílice y sílice modificado, Sephadex®, Sepharose®, carbono, metales (p.ej., acero, oro, plata, aluminio, silicio y cobre), vidrios inorgánicos, polímeros conductores (incluidos los polímeros tales como polipirrol y poliiindol); superficies micro o nanoestructuradas tales como matrices de colocación de ácidos nucleicos, superficies decoradas con nanotubos, nanocables o nanopartículas; o superficies porosas o geles tales como metacrilatos, acrilamidas, polímeros de azúcar, celulosa, silicatos y otros polímeros fibrosos o en cadena. En algunos casos, el soporte sólido o el sustrato pueden recubrirse utilizando recubrimientos pasivos o derivados químicamente con cualquier número de materiales, incluidos polímeros, como los dextranos, acrilamidas, gelatinas o agarosa. Las perlas y/o partículas pueden estar libres o en conexión unas con otras (por ejemplo, sinterizadas). En algunos casos, la fase sólida puede ser una colección de partículas. En algunos casos, las partículas pueden comprender sílice, y la sílice puede comprender dióxido de sílice. En algunos casos, la sílice puede ser porosa, y en algunos casos, la sílice puede ser no porosa. En algunos casos, las partículas comprenden además un agente que confiere una propiedad paramagnética a las partículas. En ciertos casos, el agente comprende un metal, y en ciertos casos, el agente es un óxido de metal, (por ejemplo, hierro u óxidos de hierro, donde el óxido de hierro contiene una mezcla de Fe²⁺ y Fe³⁺). Los oligonucleótidos pueden unirse al soporte sólido mediante enlaces covalentes o mediante interacciones no covalentes y pueden unirse al soporte sólido directa o indirectamente (por ejemplo, a través de un agente intermediario, como una molécula espaciadora o biotina). Una sonda puede estar unida al soporte sólido antes, durante o después de la captura del ácido nucleico.

40 En algunos casos, el ácido nucleico se enriquece para una longitud o intervalo de longitudes particular, o longitudes por debajo o por encima de un umbral o límite particular de fragmentos de ácidos nucleicos, utilizando uno o más métodos de separación basados en longitud. La longitud del fragmento de ácido nucleico generalmente se refiere al número de nucleótidos en el fragmento. También se hace referencia a la longitud del fragmento de ácido nucleico como tamaño de fragmento de ácido nucleico. En algunos casos, se realiza un método de separación basado en longitud sin medir longitudes de fragmentos individuales. En algunos casos, se realiza un método de separación basado en longitud junto con un método para determinar la longitud de fragmentos individuales. En algunos casos, la separación basada en la longitud se refiere a un procedimiento de fraccionamiento por tamaño en donde se puede aislar la totalidad o parte del conjunto fraccionado (por ejemplo, retener) y/o analizar. Los procedimientos de fraccionamiento por tamaño son conocidos en la técnica (por ejemplo, separación en una matriz, separación por un tamiz molecular, separación por electroforesis en gel, separación por cromatografía en columna (por ejemplo, columnas de exclusión por tamaño) y aproximaciones basadas en microfluidos). En ciertos casos, las aproximaciones de separación basadas en la longitud pueden incluir circularización de fragmentos, tratamiento químico (por ejemplo, formaldehído, polietilenglicol (PEG)), espectrometría de masas y/o amplificación de ácidos nucleicos de tamaño específico, por ejemplo.

55 En algunos casos, se separan de la muestra los fragmentos de ácido nucleico de una cierta longitud, intervalo de longitudes o longitudes por debajo o por encima de un umbral o límite particular. En algunos casos, fragmentos que tienen una longitud por debajo de un umbral o límite particular (por ejemplo, 500 pb, 400 pb, 300 pb, 200 pb, 150 pb, 100 pb) se denominan fragmentos "cortos" y fragmentos que tienen una longitud superior a un umbral o límite particular (por ejemplo, 500 pb, 400 pb, 300 pb, 200 pb, 150 pb, 100 pb) se denominan fragmentos "largos". En algunos casos, se retienen los fragmentos de una cierta longitud, intervalo de longitudes o longitudes por debajo o por encima de un umbral o límite particular, para su análisis mientras que no se retienen los fragmentos de una longitud o intervalo de longitudes diferentes, o longitudes por encima o por debajo del umbral o límite para el análisis. En algunos casos, se retienen fragmentos que tienen menos de aproximadamente 500 pb. En algunos casos, se retienen fragmentos que tienen menos de aproximadamente 400 pb. En algunos casos, se retienen fragmentos que tienen menos de aproximadamente 300 pb. En algunos casos, se retienen fragmentos que tienen menos de aproximadamente 200 pb. En algunos casos, se retienen fragmentos que tienen menos de aproximadamente 150 pb. Por ejemplo, se retienen

los fragmentos que tienen menos de aproximadamente 190 pb, 180 pb, 170 pb, 160 pb, 150 pb, 140 pb, 130 pb, 120 pb, 110 pb o 100 pb. En algunos casos, se retienen fragmentos que tienen de aproximadamente 100 pb a aproximadamente 200 pb. Por ejemplo, se retienen fragmentos que son de aproximadamente 190 pb, 180 pb, 170 pb, 160 pb, 150 pb, 140 pb, 130 pb, 120 pb o 110 pb. En algunos casos, se retienen los fragmentos que pueden estar en el intervalo de aproximadamente 100 pb a aproximadamente 200 pb. Por ejemplo, se retienen fragmentos que están en el intervalo de aproximadamente 110 pb a aproximadamente 190 pb, 130 pb a aproximadamente 180 pb, 140 pb a aproximadamente 170 pb, 140 pb a aproximadamente 150 pb, 150 pb a aproximadamente 160 pb, o 145 pb a aproximadamente 155 pb. En algunos casos, se retienen los fragmentos que son de aproximadamente 10 pb a aproximadamente 30 pb más cortos que otros fragmentos de una determinada longitud o intervalo de longitudes. En algunos casos, se retienen los fragmentos que son de aproximadamente 10 pb a aproximadamente 20 pb más cortos que otros fragmentos de una determinada longitud o intervalo de longitudes. En algunos casos, se retienen los fragmentos que son de aproximadamente 10 pb a aproximadamente 15 pb más cortos que otros fragmentos de una determinada longitud o intervalo de longitudes.

En algunos casos, el ácido nucleico se enriquece para una longitud, intervalo de longitudes particular o longitudes por debajo o por encima de un umbral o límite particular de fragmento de ácido nucleico, usando uno o más métodos basados en bioinformática (por ejemplo, *in silico*). Por ejemplo, se pueden obtener lecturas de secuencias de nucleótidos para fragmentos de ácidos nucleicos utilizando un proceso de secuenciación de nucleótidos adecuado. En algunos casos, tal como cuando se usa un método de secuenciación de extremo pareado, la longitud de un fragmento particular puede determinarse en función de las posiciones de las lecturas de secuencia mapeadas obtenidas de cada extremo del fragmento. Las lecturas de secuencia utilizadas para un análisis particular (por ejemplo, la determinación de la presencia o ausencia de una variación genética) pueden enriquecerse o filtrarse de acuerdo con una o más longitudes de fragmentos o valores umbrales de longitud de fragmentos seleccionadas de fragmentos correspondientes, como se describe con más detalle en el presente documento.

Ciertos métodos de separación basados en la longitud que se pueden usar con los métodos descritos en el presente documento a veces emplean una aproximación de marcado de secuencias selectivo, por ejemplo. La expresión "marcado de secuencias" se refiere a la incorporación de una secuencia reconocible y distinta en un ácido nucleico o población de ácidos nucleicos. La expresión "marcado de secuencias" como se usa en el presente documento tiene un significado diferente a la expresión "marcador de secuencias" descrito más adelante en el presente documento. En tales métodos de marcado de secuencias, los ácidos nucleicos de una especie de tamaño de fragmento (por ejemplo, fragmentos cortos) se someten a un marcado de secuencias selectivo en una muestra que incluye ácidos nucleicos largos y cortos. Tales métodos generalmente implican realizar una reacción de amplificación de ácidos nucleicos usando un conjunto de cebadores anidados que incluyen cebadores internos y cebadores externos. En ciertos casos, uno o ambos de los cebadores internos se pueden marcar para introducir de este modo un marcador en el producto de amplificación diana. Los cebadores externos generalmente no se emparejan con los fragmentos cortos que portan la secuencia diana (interna). Los cebadores internos pueden emparejarse a los fragmentos cortos y generar un producto de amplificación que lleva un marcador y la secuencia diana. Generalmente, el marcado de los fragmentos largos se inhibe a través de una combinación de mecanismos que incluyen, por ejemplo, la extensión bloqueada de los cebadores internos mediante el emparejamiento previo y la extensión de los cebadores externos. El enriquecimiento para los fragmentos marcados se puede lograr mediante cualquiera de una variedad de métodos, incluyendo, por ejemplo, la digestión con exonucleasa de ácido nucleico monocatenario y la amplificación de los fragmentos marcados utilizando cebadores de amplificación específicos para al menos un marcador.

Otro método de separación basado en la longitud que se puede usar con los métodos descritos en el presente documento implica someter una muestra de ácidos nucleicos a precipitación con polietilenglicol (PEG). Los ejemplos de métodos incluyen los descritos en las Publicaciones de Solicitud de Patente Internacional N° WO2007/140417 y WO2010/115016. Este método en general implica poner en contacto una muestra de ácidos nucleicos con PEG en presencia de una o más sales monovalentes en condiciones suficientes para precipitar sustancialmente ácidos nucleicos grandes sin precipitar sustancialmente ácidos nucleicos pequeños (por ejemplo, menos de 300 nucleótidos).

Otro método de enriquecimiento basado en el tamaño que se puede usar con los métodos descritos en el presente documento implica la circularización por unión, por ejemplo, usando circligase. Los fragmentos de ácidos nucleicos cortos generalmente pueden circularizarse con mayor eficacia que los fragmentos largos. Las secuencias no circularizadas se pueden separar de las secuencias circularizadas, y los fragmentos cortos enriquecidos se pueden usar para un análisis adicional.

Determinación de la longitud del fragmento

En algunos casos, la longitud se determina para uno o más fragmentos de ácidos nucleicos. En algunos casos, la longitud se determina para uno o más fragmentos diana, identificando de este modo, una o más especies de tamaño de fragmento diana. La longitud se puede determinar para uno o más fragmentos diana y uno o más fragmentos de referencia, identificando, de este modo, una o más especies de longitud de fragmento diana y una o más especies de longitud de fragmento de referencia. En algunos casos, la longitud del fragmento se determina midiendo la longitud de una sonda que hibrida con el fragmento, lo que se explica con mayor detalle a continuación. La longitud del fragmento de ácido nucleico o de la sonda se puede determinar utilizando cualquier método en la técnica adecuado para

determinar la longitud de fragmentos de ácidos nucleicos, tal como, por ejemplo, un proceso sensible a la masa (por ejemplo, espectrometría de masas (por ejemplo, espectrometría de masas por ionización por desorción mediante láser asistida por matriz (MALDI) y espectroscopía de masas por electropulverización (ES)), electroforesis (por ejemplo, electroforesis capilar), microscopía (microscopía de efecto túnel, microscopía de fuerza atómica), midiendo la longitud usando un nanoporo, y la determinación de la longitud basada en la secuencia (por ejemplo, secuenciación de extremo pareado). En algunos casos, la longitud del fragmento o sonda se puede determinar sin utilizar un método de separación basado en la carga del fragmento. En algunos casos, la longitud del fragmento o sonda se puede determinar sin utilizar un proceso de electroforesis. En algunos casos, la longitud del fragmento o sonda se puede determinar sin utilizar un proceso de secuenciación de nucleótidos.

Espectrometría de masas

En algunos casos, se usa la espectrometría de masas para determinar la longitud del fragmento de ácido nucleico. Los métodos de espectrometría de masas se usan generalmente para determinar la masa de una molécula, tal como un fragmento de ácido nucleico. En algunos casos, la longitud del fragmento de ácido nucleico se puede extrapolar de la masa del fragmento. En algunos casos un intervalo predicho de longitudes de fragmento de ácido nucleico se puede extrapolar de la masa del fragmento. En algunos casos la longitud del fragmento de ácido nucleico se puede extrapolar de la masa de una sonda que hibrida con el fragmento, que se describe con más detalle a continuación. En algunos casos, la presencia de un ácido nucleico diana y/o de referencia de una longitud determinada se puede verificar comparando la masa de la señal detectada con la masa esperada del fragmento diana y/o de referencia. La intensidad de señal relativa, por ejemplo, el máximo de masa en un espectro, para un fragmento de ácido nucleico y/o la longitud del fragmento particular, puede indicar a veces la población relativa de las especies de fragmentos entre otros ácidos nucleicos en la muestra (véase, por ejemplo, Jurinke et al. (2004) Mol. Biotechnol. 26, 147-164).

La espectrometría de masas generalmente funciona ionizando compuestos químicos para generar moléculas o fragmentos de moléculas cargadas y midiendo sus relaciones masa a carga. Un procedimiento típico de espectrometría de masas implica varias etapas, que incluyen (1) cargar una muestra en el instrumento de espectrometría de masas seguido de vaporización, (2) ionización de los componentes de la muestra por cualquiera de una variedad de métodos (p. ej., impactando con un haz de electrones), dando como resultado partículas cargadas (iones), (3) separación de iones según su relación masa a carga en un analizador por campos electromagnéticos, (4) detección de iones (por ejemplo, mediante un método cuantitativo) y (5) procesamiento de la señal iónica en espectros de masas.

Los métodos de espectrometría de masas son bien conocidos en la técnica (véase, por ejemplo, Burlingame et al. Anal. Chem. 70:647R-716R (1998)), e incluyen, por ejemplo, espectrometría de masas cuadrupolar, espectrometría de masas con trampa de iones, espectrometría de masas de tiempo de vuelo, espectrometría de masas con cromatografía de gases y espectrometría de masas en tándem que se pueden utilizar con los métodos descritos en el presente documento. Los procesos básicos asociados con un método de espectrometría de masas son la generación de iones de fase gaseosa derivados de la muestra y la medición de su masa. El movimiento de los iones de fase gaseosa se puede controlar con precisión utilizando campos electromagnéticos generados en el espectrómetro de masas. El movimiento de iones en estos campos electromagnéticos es proporcional a la m/z (relación masa a carga) del ion y esto constituye la base para medir la m/z y, por lo tanto, la masa de una muestra. El movimiento de iones en estos campos electromagnéticos permite la contención y el enfoque de los iones, lo que explica la alta sensibilidad de la espectrometría de masas. Durante el curso de la medición de m/z , los iones se transmiten con alta eficacia a detectores de partículas que registran la llegada de estos iones. La cantidad de iones en cada m/z se demuestra mediante máximos en un gráfico donde el eje x es m/z y el eje y es la abundancia relativa. Los diferentes espectrómetros de masas tienen diferentes niveles de resolución, es decir, la capacidad de resolver máximos entre iones estrechamente relacionados en masa. La resolución se define como $R = m/\Delta m$, donde m es la masa de iones y Δm es la diferencia de masa entre dos máximos en un espectro de masas. Por ejemplo, un espectrómetro de masas con una resolución de 1000 puede resolver un ion con una m/z de 100,0 de un ion con una m/z de 100,1. Ciertos métodos de espectrometría de masas pueden utilizar varias combinaciones de fuentes de iones y analizadores de masas, lo que permite flexibilidad en el diseño de protocolos de detección personalizados. En algunos casos, los espectrómetros de masas se pueden programar para transmitir todos los iones desde la fuente de iones al espectrómetro de masas de forma secuencial o al mismo tiempo. Se puede programar un espectrómetro de masas para seleccionar iones de una masa particular para transmitirlos al espectrómetro de masas mientras se bloquean otros iones.

Varios tipos de espectrómetros de masas están disponibles o se pueden producir con diversas configuraciones. En general, un espectrómetro de masas tiene los siguientes componentes principales: una entrada de muestra, una fuente de iones, un analizador de masas, un detector, un sistema de vacío, un sistema de control de instrumentos y un sistema de datos. La diferencia en la entrada de muestra, la fuente de iones y el analizador de masas generalmente definen el tipo de instrumento y sus capacidades. Por ejemplo, una entrada puede ser una fuente de cromatografía líquida de columna capilar o puede ser una sonda o etapa directa como la utilizada en la desorción con láser asistida por matriz. Las fuentes de iones comunes son, por ejemplo, la electropulverización, incluida la nanopulverización y la micropulverización o la desorción con láser asistida por matriz. Los analizadores de masa incluyen, por ejemplo, un filtro de masa cuadrupolar, un analizador de masas con trampa de iones y un analizador de masas de tiempo de vuelo.

El proceso de formación de iones es un punto de partida para el análisis del espectro de masas. Hay varios métodos de ionización disponibles y la elección del método de ionización depende de la muestra utilizada para el análisis. Por ejemplo, para el análisis de polipéptidos puede ser deseable un procedimiento de ionización relativamente suave como la ionización por electropulverización (ESI). Para ESI, se pasa una solución que contiene la muestra a través de una
5 aguja fina con un alto potencial, lo que crea un campo eléctrico fuerte que produce una pulverización fina de gotas altamente cargadas que se dirigen hacia el espectrómetro de masas. Otros procedimientos de ionización incluyen, por ejemplo, bombardeo por átomos rápidos (FAB), que utiliza un haz de alta energía de átomos neutros para golpear una muestra sólida que causa desorción e ionización. La ionización por desorción mediante láser asistida por matriz (MALDI, por sus siglas en inglés) es un método en donde se usa un pulso láser para golpear una muestra que se ha
10 cristalizado en una matriz de compuesto absorbente de UV (por ejemplo, ácido 2,5-dihidroxibenzoico, ácido alfa-ciano-4-hidroxicinámico, ácido 3-hidroxipicolínico (3-HPA), citrato de di-amonio (DAC) y combinaciones de los mismos). Otros procedimientos de ionización bien conocidos en la técnica incluyen, por ejemplo, descarga de plasma y luz incandescente, ionización por desorción de plasma, ionización por resonancia e ionización secundaria.

15 Existe una variedad de analizadores de masas disponibles que pueden combinarse con diferentes fuentes de iones. Diferentes analizadores de masas tienen diferentes ventajas como se conoce en la técnica y como se describe en el presente documento. El espectrómetro de masas y los métodos elegidos para la detección dependen del ensayo particular, por ejemplo, se puede usar un analizador de masas más sensible cuando se genera una pequeña cantidad de iones para la detección. A continuación, se describen varios tipos de analizadores de masas y métodos de
20 espectrometría de masas.

La espectrometría de masa de movilidad iónica (IM) es un método de separación en fase gaseosa. IM separa los iones en fase gaseosa en función de su sección transversal de colisión y se puede acoplar a la espectrometría de masas de tiempo de vuelo (TOF). La IM-MS se discute con más detalle por Verbeck et al. en el Journal of Biomolecular
25 Techniques (Vol. 13, Issue 2, 56-61).

La espectrometría de masas cuadrupolar utiliza un analizador o filtro de masa cuadrupolar. Este tipo de analizador de masas está compuesto por cuatro vástagos dispuestos como dos conjuntos de dos vástagos conectados eléctricamente. Se aplica una combinación de voltajes de rf y dc a cada par de vástagos que producen campos que
30 causan un movimiento oscilante de los iones a medida que se mueven desde el principio del filtro de masas hasta el final. El resultado de estos campos es la producción de un filtro de masas de paso alto en un par de vástagos y un filtro de paso bajo en el otro par de vástagos. La superposición entre el filtro de paso alto y paso bajo deja una m/z definida que puede pasar ambos filtros y atravesar la longitud del cuadrupolar. Esta m/z se selecciona y permanece estable en el filtro de masa cuadrupolar mientras que todos las demás m/z tienen trayectorias inestables y no
35 permanecen en el filtro de masas. Se obtiene un espectro de masas al rampear los campos aplicados de manera que se selecciona una m/z creciente para pasar a través del filtro de masas y alcanzar el detector. Además, los cuadrupolos también se pueden configurar para contener y transmitir iones de todas las m/z aplicando un campo de solo rf. Esto permite que los cuadrupolos funcionen como lentes o sistemas de enfoque en regiones del espectrómetro de masas donde se necesita la transmisión de iones sin filtrado de masas.

40 Un analizador de masas cuadrupolar, así como los otros analizadores de masas descritos en el presente documento, pueden programarse para analizar una m/z o intervalo de masas definido. Debido a que el intervalo de masas deseado del fragmento de ácido nucleico es conocido, en algunos casos, un espectrómetro de masas puede programarse para transmitir iones del intervalo de masas correcto proyectado mientras se excluyen iones de un intervalo de masas mayor o menor. La capacidad de seleccionar un intervalo de masas puede disminuir el ruido de fondo en el ensayo y, por lo tanto, aumentar la relación señal-ruido. Por lo tanto, en algunos casos, un espectrómetro de masas puede realizar una
45 etapa de separación, así como la detección e identificación de ciertos fragmentos de ácido nucleico distinguibles en masa.

50 La espectrometría de masas con trampa de iones utiliza un analizador de masas con trampa de iones. Generalmente, los campos se aplican de manera que los iones de todas las m/z se atrapan inicialmente y oscilan en el analizador de masas. Los iones ingresan a la trampa de iones desde la fuente de iones a través de un dispositivo de enfoque, como un sistema de lente octopolar. La captura de iones se lleva a cabo en la región de captura antes de la excitación y la expulsión a través de un electrodo al detector. El análisis de masas se puede realizar mediante la aplicación secuencial
55 de voltajes que aumentan la amplitud de las oscilaciones de una manera que expulsa iones de m/z en aumento fuera de la trampa y dentro del detector. En contraste con la espectrometría de masas cuadrupolar, todos los iones se retienen en los campos del analizador de masas, excepto aquellos con la m/z seleccionada. El control del número de iones se puede lograr variando el tiempo durante el cual los iones se inyectan en la trampa.

60 La espectrometría de masas de tiempo de vuelo utiliza un analizador de masas de tiempo de vuelo. Generalmente, primero se le da a un ion una cantidad fija de energía cinética mediante la aceleración en un campo eléctrico (generado por alto voltaje). Tras la aceleración, el ion entra en una región libre de campo o "deriva" donde viaja a una velocidad que es inversamente proporcional a su m/z . Por lo tanto, los iones con m/z baja viajan más rápidamente que los iones con m/z alta. El tiempo requerido para que los iones viajen a lo largo de la región libre de campo se mide y se usa para
65 calcular la m/z del ion.

La espectrometría de masas con cromatografía de gases a menudo puede ser un objetivo en tiempo real. La porción de cromatografía de gases (GC) del sistema separa la mezcla química en pulsos de analito y el espectrómetro de masas (MS) identifica y cuantifica el analito.

5 La espectrometría de masas en tándem puede utilizar combinaciones de los analizadores de masas descritos anteriormente. Los espectrómetros de masas en tándem pueden usar un primer analizador de masas para separar los iones de acuerdo con su m/z para aislar un ion de interés para un análisis adicional. A continuación, el ion de interés aislado se divide en iones de fragmentos (llamado disociación activada por colisión o disociación inducida por colisión) y los iones de fragmentos se analizan mediante el segundo analizador de masas. Estos tipos de sistemas de 10 espectrómetro de masas en tándem se denominan en tándem en sistemas espaciales porque los dos analizadores de masas están separados en el espacio, generalmente por una celda de colisión. Los sistemas de espectrómetros de masas en tándem también incluyen sistemas de tándem en el tiempo donde se usa un analizador de masas, sin embargo, el analizador de masas se usa secuencialmente para aislar un ion, inducir la fragmentación y a continuación realizar el análisis de masas.

15 Los espectrómetros de masas en la categoría de tándem en el espacio tienen más de un analizador de masas. Por ejemplo, un sistema de espectrómetro de masas de cuadrupolo en tándem puede tener un primer filtro de masas de cuadrupolo, seguido de una celda de colisión, seguido de un segundo filtro de masas de cuadrupolo y a continuación el detector. Otra disposición es utilizar un filtro de masa cuadrupolo para el primer analizador de masas y un analizador 20 de masas de tiempo de vuelo para el segundo analizador de masas con una celda de colisión que separa los dos analizadores de masas. Otros sistemas en tándem son conocidos en la técnica, incluyendo espectrometría de masas de tiempo de vuelo con reflectrón, sector en tándem y sector-cuadrupolo.

25 Los espectrómetros de masas en la categoría de tándem en el tiempo tienen un analizador de masas que realiza diferentes funciones en diferentes momentos. Por ejemplo, se puede usar un espectrómetro de masas con trampa de iones para atrapar iones de todas las m/z . Se aplican una serie de funciones de escaneo rf que expulsan iones de todas las m/z de la trampa, excepto la m/z de los iones de interés. Después de aislar la m/z de interés, se aplica un pulso rf para producir colisiones con moléculas de gas en la trampa para inducir la fragmentación de los iones. A continuación, los valores m/z de los iones fragmentados se miden con el analizador de masas. Los instrumentos de 30 resonancia ciclotrónica de iones, también conocidos como espectrómetros de masas de transformada de Fourier, son un ejemplo de sistemas en tándem en el tiempo.

Se pueden realizar varios tipos de experimentos de espectrometría de masas en tándem controlando los iones que se seleccionan en cada etapa del experimento. Los diferentes tipos de experimentos utilizan diferentes modos de 35 operación, a veces llamados "exploraciones", de los analizadores de masas. En un primer ejemplo, llamado una exploración de espectro de masas, el primer analizador de masas y la celda de colisión transmiten todos los iones para el análisis de masas al segundo analizador de masas. En un segundo ejemplo, llamado una exploración de iones del producto, los iones de interés se seleccionan en función de la masa en el primer analizador de masas y luego se fragmentan en la celda de colisión. Los iones formados se analizan, a continuación, las masas explorando el segundo 40 analizador de masas. En un tercer ejemplo, llamado rastreo de iones del precursor, el primer analizador de masas se explora para transmitir secuencialmente los iones de masas analizadas en la celda de colisión para la fragmentación. El segundo analizador de masas selecciona en función de la masa el ion del producto de interés para transmitirlo al detector. Por lo tanto, la señal del detector es el resultado de todos los iones de precursores que pueden fragmentarse en un ion de producto común. Otros formatos experimentales incluyen exploraciones de pérdida neutra donde una 45 diferencia de masa constante se tiene en cuenta en las exploraciones de masa.

Para la cuantificación, se pueden usar controles que pueden proporcionar una señal en relación con la cantidad de fragmento de ácido nucleico, por ejemplo, que está presente o se introduce. Un control para permitir la conversión de 50 señales de masa relativa en cantidades absolutas se puede lograr mediante la adición de una cantidad conocida de un marcador de masa o etiqueta de masa a cada muestra antes de la detección de los fragmentos de ácido nucleico. Véase, por ejemplo, Ding y Cantor (2003) PNAS EEUU. 18 de marzo; 100(6):3059-64. Cualquier marcador de masa que no interfiera con la detección de los fragmentos puede usarse para normalizar la señal de masa. Tales patrones generalmente tienen propiedades de separación que son diferentes de las de cualquiera de los marcadores moleculares en la muestra, y podrían tener la misma o diferentes firmas de masa.

55 En ocasiones, se puede usar una etapa de separación para eliminar sales, enzimas u otros componentes del tampón de la muestra de ácido nucleico. Se pueden utilizar varios métodos bien conocidos en la técnica, tales como cromatografía, electroforesis en gel o precipitación para limpiar la muestra. Por ejemplo, se puede usar cromatografía de exclusión por tamaño o cromatografía de afinidad para eliminar la sal de una muestra. La elección del método de 60 separación puede depender de la cantidad de una muestra. Por ejemplo, cuando se dispone de pequeñas cantidades de muestra o se utiliza una máquina miniaturizada, se puede usar una etapa de separación por cromatografía de afinidad. Además, si se desea una etapa de separación, la elección del método de separación, puede depender del método de detección utilizado. Las sales a veces pueden absorber la energía del láser en la desorción/ionización mediante láser asistida por matriz y dar como resultado una menor eficacia de ionización. Por lo tanto, la eficacia de 65 la desorción/ionización mediante láser asistida por matriz y la ionización por electropulverización a veces se puede mejorar eliminando sales de una muestra.

Electroforesis

En algunos casos, se usa electroforesis para determinar la longitud del fragmento de ácido nucleico. En algunos casos, no se usa electroforesis para determinar la longitud del fragmento de ácido nucleico. En algunos casos se determina la longitud de una sonda correspondiente (por ejemplo, una sonda recortada correspondiente descrita en el presente documento) utilizando electroforesis. En algunos casos, la electroforesis también se puede usar como un método de separación basado en la longitud como se describe en el presente documento. Se puede usar cualquier método de electroforesis conocido en la técnica, por el que los ácidos nucleicos se separan por longitud, junto con los métodos proporcionados en el presente documento, que incluyen, pero sin limitación, técnicas electroforéticas convencionales y técnicas electroforéticas especializadas, tal como, por ejemplo, electroforesis capilar. En la técnica se pueden encontrar ejemplos de métodos para separar el ácido nucleico y medir la longitud del fragmento de ácido nucleico usando técnicas electroforéticas convencionales. Se expone, en el presente documento, un ejemplo no limitante. Después de ejecutar una muestra de ácido nucleico en un gel de agarosa o poliacrilamida, el gel puede etiquetarse (p. ej., teñirse) con bromuro de etidio (véase, Sambrook y Russell, "Molecular Cloning: A Laboratory Manual 3ª ed., 2001). La presencia de una banda del mismo tamaño que un control convencional es una indicación de la presencia de una longitud de secuencia de ácido nucleico particular, cuya cantidad se puede comparar con el control en función de la intensidad de la banda, detectando y cuantificando, de este modo, la longitud de la secuencia de ácido nucleico de interés.

En algunos casos, se usa electroforesis capilar para separar, identificar y, a veces, cuantificar fragmentos de ácido nucleico. La electroforesis capilar (CE) abarca una familia de técnicas de separación relacionadas que utilizan capilares de sílice fundida de calibre angosto para separar una matriz compleja de moléculas grandes y pequeñas, tal como, por ejemplo, ácidos nucleicos de longitud variable. Se pueden utilizar altas intensidades de campo eléctrico para separar las moléculas de ácidos nucleicos en función de las diferencias en la carga, el tamaño y la hidrofobicidad. La introducción de la muestra se realiza sumergiendo el extremo del capilar en un vial de muestra y aplicando presión, vacío o voltaje. Dependiendo de los tipos de capilares y electrolitos utilizados, la tecnología de CE se puede segmentar en varias técnicas de separación, cualquiera de las cuales se puede adaptar a los métodos proporcionados en el presente documento. Ejemplos no limitantes de estos incluyen la Electroforesis Capilar en Zona (CZE), también conocida como CE en solución libre (FSCE), Enfoque Isoeléctrico Capilar (CIEF), Isotacoforesis (ITP), Cromatografía Electrocinética (EKC), Cromatografía Electrocinética Capilar Micelar (MECC o MEKC), Cromatografía Electrocinética Microemulsionada (MEEKC), Electroforesis Capilar no acuosa (NACE) y Electro cromatografía Capilar (CEC).

Cualquier dispositivo, instrumento o máquina capaz de realizar una electroforesis capilar puede usarse junto con los métodos que se proporcionan en el presente documento. En general, los componentes principales del sistema de electroforesis capilar son un vial de muestra, viales de origen y destino, un capilar, electrodos, una fuente de alimentación de alto voltaje, un detector y un dispositivo de manejo y salida de datos. El vial de origen, el vial de destino y el capilar se llenan con un electrolito, como una solución tampón acuosa. Para introducir la muestra, la entrada del capilar se coloca en un vial que contiene la muestra y a continuación, se devuelve al vial de origen (la muestra se introduce en el capilar a través de acción capilar, presión o por un sifón). La migración de los analitos (es decir, ácidos nucleicos) se inicia, a continuación, mediante un campo eléctrico que se aplica entre los viales de origen y de destino y se suministra a los electrodos mediante la fuente de alimentación de alto voltaje. Los iones, positivos o negativos, son atraídos a través del capilar en la misma dirección por el flujo electroosmótico. Los analitos (es decir, los ácidos nucleicos) se separan a medida que migran debido a su movilidad electroforética y se detectan cerca del extremo de salida del capilar. La salida del detector se envía a un dispositivo de manejo y salida de datos, como un integrador o un ordenador. Los datos se muestran, a continuación, como un electroferograma, que puede indicar la respuesta del detector en función del tiempo. Los ácidos nucleicos separados pueden aparecer como máximos con diferentes tiempos de migración en un electroferograma.

La separación por electroforesis capilar puede detectarse por varios dispositivos de detección. La mayoría de los sistemas comerciales utilizan la absorbancia UV o UV-Vis como su principal modo de detección. En estos sistemas, una sección del propio capilar se utiliza como la celda de detección. El uso de la detección en tubo permite la detección de analitos separados sin pérdida de resolución. En general, los capilares utilizados en la electroforesis capilar se pueden recubrir con un polímero para aumentar la estabilidad. La porción del capilar utilizado para la detección de rayos UV suele ser ópticamente transparente. La longitud del recorrido de la celda de detección en la electroforesis capilar (~ 50 micrómetros) es mucho menor que la de una celda UV tradicional (~ 1 cm). De acuerdo con la ley de Beer-Lambert, la sensibilidad del detector es proporcional a la longitud del recorrido de la celda. Para mejorar la sensibilidad, se puede aumentar la longitud del recorrido, aunque esto puede dar como resultado una pérdida de resolución. El propio tubo capilar se puede expandir en el punto de detección, creando una "celda burbuja" con una longitud de recorrido más larga o se pueden agregar tubos adicionales en el punto de detección. Ambos métodos, sin embargo, pueden disminuir la resolución de la separación.

La detección de fluorescencia también se puede usar en la electroforesis capilar para muestras que presentan fluorescencia natural o se modifican químicamente para contener marcadores fluorescentes, tal como, por ejemplo, fragmentos de ácido nucleico o sondas etiquetados descritos en el presente documento. Este modo de detección ofrece una alta sensibilidad y una selectividad mejorada para estas muestras. El método requiere que el haz de luz se enfoque en el capilar. La fluorescencia inducida por láser se puede usar en sistemas de CE con límites de detección

tan bajos como 10^{-18} a 10^{-21} mol. La sensibilidad de la técnica se atribuye a la alta intensidad de la luz incidente y la capacidad de enfocar con precisión la luz en el capilar.

Varias máquinas de electroforesis capilar son conocidas en la técnica y pueden usarse junto con los métodos proporcionados en el presente documento. Estas incluyen, pero sin limitación, CALIPER LAB CHIP GX (Caliper Life Sciences, Mountain View, CA), P/ACE 2000 Series (Beckman Coulter, Brea, CA), CE HP G1600A (Hewlett-Packard, Palo Alto, CA), CE AGILENT 7100 (Agilent Technologies, Santa Clara, CA), y Analizador Genético ABI PRISM (Applied Biosystems, Carlsbad, CA).

10 *Microscopía*

En algunos casos, la longitud de los fragmentos de ácidos nucleicos se determina utilizando un método basado en imágenes, como un método de microscopía. En algunos casos, se determina la longitud de una sonda correspondiente (por ejemplo, una sonda recortada correspondiente descrita en el presente documento) utilizando un método basado en imágenes. En algunos casos, la longitud de los fragmentos puede determinarse por visualización microscópica de fragmentos de ácidos nucleicos únicos (véase, por ejemplo, la patente de EE.UU. n.º 5.720.928). En algunos casos, los fragmentos de ácidos nucleicos se fijan a una superficie (por ejemplo, una superficie de vidrio modificada) en un estado alargado, teñirse y visualizarse microscópicamente. Las imágenes de los fragmentos se pueden recopilar y procesar (por ejemplo, medir su longitud). En algunos casos, las etapas de obtención de imágenes y de análisis de imágenes se pueden automatizar. Los métodos para visualizar directamente los fragmentos de ácidos nucleicos utilizando microscopía son conocidos en la técnica (véase, por ejemplo, Lai et al. (1999) Nat. Genet. 23(3):309-13; Aston et al. (1999) Trends Biotechnol. 17(7):297-302; Aston et al. (1999) Methods Enzymol. 303:55-73; O'Toole, et al. (1998) Proc Natl Acad Sci USA. 95(14):8046-51; y la Patente de Estados Unidos n.º 5.720.928). Otros métodos de microscopía que se pueden usar con los métodos descritos en el presente documento, incluyen, sin limitación, microscopía de barrido de efecto túnel (STM), microscopía de fuerza atómica (AFM), microscopía de fuerza de barrido (SFM), microscopía de barrido de fotones (PSTM), potenciometría de barrido de efecto túnel (STP), microscopía de fuerza magnética (MFM), microscopía de barrido de sonda, microscopía de voltaje de barrido, microscopía de fuerza atómica fotoconductora, microscopía de barrido electroquímico de efecto túnel, microscopía electrónica, microscopía de barrido de efecto túnel de espín polarizado (SPSTM), microscopía de barrido térmico, microscopía de expansión de barrido de julios, microespectroscopía fototérmica y similares.

En algunos casos, se puede usar microscopía de barrido de efecto túnel (STM) para determinar la longitud de los fragmentos de ácidos nucleicos. A menudo, los métodos STM pueden generar imágenes de moléculas a nivel atómico, como los fragmentos de ácidos nucleicos. La STM se puede llevar a cabo, por ejemplo, en aire, agua, vacío ultra -alto, otros ambientes diversos de líquidos o gases, y se puede realizar a temperaturas que oscilan desde cerca de cero Kelvin a unos pocos cientos de grados centígrados, por ejemplo. Los componentes de un sistema STM incluyen generalmente la punta de barrido, la altura piezoeléctrica controlada y el dispositivo de barrido x, y, el control bruto de muestra a punta, el sistema de aislamiento por vibración y el ordenador. Los métodos de STM se basan generalmente en el concepto del efecto túnel cuántico. Por ejemplo, cuando una punta conductora se acerca a la superficie de una molécula (por ejemplo, un fragmento de ácido nucleico), una desviación (es decir, una diferencia de voltaje) aplicada entre las dos puede permitir que los electrones formen un túnel a través del vacío entre ellas. La corriente de efecto túnel resultante es una función de la posición de la punta, el voltaje aplicado y la densidad local de estados (LDOS) de la muestra. La información se obtiene al monitorizar la corriente a medida que se barre la posición de la punta a través de la superficie y se puede mostrar en forma de imagen. Si la punta se mueve a través de la muestra en el plano x-y, los cambios en la altura de la superficie y la densidad de los estados causan cambios en la corriente. Estos cambios se pueden mapear en imágenes. El cambio en la corriente con respecto a la posición a veces se puede medir por sí mismo, o se puede medir la altura, z, de la punta correspondiente a una corriente constante. Estos dos modos a menudo se denominan modo de altura constante y modo de corriente constante, respectivamente.

En algunos casos, se puede usar microscopía de fuerza atómica (AFM) para determinar la longitud de los fragmentos de ácidos nucleicos. La AFM generalmente es un tipo de microscopía a nanoescala de alta resolución. La información sobre un objeto (por ejemplo, un fragmento de ácido nucleico) normalmente se recopila al "sentir" la superficie con una sonda mecánica. Los elementos piezoeléctricos que facilitan movimientos pequeños pero precisos en el comando electrónico pueden facilitar el barrido muy preciso. En algunas variaciones, los potenciales eléctricos se pueden barrer utilizando soportes de conducción. Los componentes de un sistema AFM generalmente incluyen un soporte con una punta afilada (es decir, una sonda) en su extremo que se usa para barrer la superficie de una muestra (por ejemplo, un fragmento de ácido nucleico). El soporte es generalmente de silicio o nitruro de silicio con un radio de curvatura de la punta del orden de nanómetros. Cuando la punta se acerca a la superficie de una muestra, las fuerzas entre la punta y la muestra conducen a una desviación del soporte según la ley de Hooke. Dependiendo de la situación, las fuerzas que se miden en AFM incluyen, por ejemplo, fuerza de contacto mecánico, fuerzas de van der Waals, fuerzas capilares, enlace químico, fuerzas electrostáticas, fuerzas magnéticas, fuerzas de Casimir, fuerzas de solvatación y similares. Generalmente, la desviación se mide utilizando un punto láser reflejado desde la superficie superior del soporte en una serie de fotodiodos. Otros métodos que se utilizan incluyen interferometría óptica, detección capacitiva o soportes piezorresistivos de AFM.

65

Nanoporos

En algunos casos, la longitud de los fragmentos de ácidos nucleicos se puede determinar utilizando un nanoporo. En algunos casos, la longitud de una sonda correspondiente (por ejemplo, una sonda recortada correspondiente descrita en el presente documento) se determina utilizando un nanoporo. Un nanoporo es un orificio o canal pequeño, generalmente del orden de 1 nanómetro de diámetro. Ciertas proteínas celulares transmembrana pueden actuar como nanoporos (por ejemplo, alfa-hemolisina). En algunos casos, los nanoporos se pueden sintetizar (por ejemplo, utilizando una plataforma de silicio). La inmersión de un nanoporo en un fluido conductor y la aplicación de un potencial a través de él da como resultado una ligera corriente eléctrica debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño del nanoporo. Cuando un fragmento de ácido nucleico pasa a través de un nanoporo, la molécula de ácido nucleico obstruye el nanoporo en cierto grado y genera un cambio en la corriente. Puede medirse la duración del cambio de corriente a medida que el fragmento de ácido nucleico pasa a través del nanoporo. En algunos casos, longitud del fragmento de ácido nucleico puede determinarse basándose en esta medición.

En algunos casos, la longitud del fragmento de ácido nucleico puede determinarse en función del tiempo. Los fragmentos de ácidos nucleicos más largos a veces pueden necesitar relativamente más tiempo para pasar a través de un nanoporo y los fragmentos de ácidos nucleicos más cortos a veces pueden necesitar relativamente menos tiempo para pasar a través de un nanoporo. Por lo tanto, en algunos casos, se puede determinar la longitud relativa de un fragmento en función del tiempo de tránsito a través del nanoporo. En algunos casos, se puede determinar la longitud aproximada o absoluta de los fragmentos comparando el tiempo de tránsito a través de los nanoporos de los fragmentos diana y/o los fragmentos de referencia con los tiempos de tránsito para un conjunto de patrones (es decir, con longitudes conocidas).

Sondas

En algunos casos, la longitud de los fragmentos se determina utilizando una o más sondas. En algunos casos, las sondas pueden diseñarse de modo que cada una hibride con un ácido nucleico de interés en una muestra. Por ejemplo, una sonda puede comprender una secuencia de polinucleótidos que es complementaria a un ácido nucleico de interés o puede comprender una serie de monómeros que pueden unirse a un ácido nucleico de interés. Las sondas pueden ser de cualquier longitud adecuada para hibridar (por ejemplo, hibridar completamente) con uno o más fragmentos de ácidos nucleicos de interés. Por ejemplo, las sondas pueden tener cualquier longitud que abarque o se extienda más allá de la longitud de un fragmento de ácido nucleico con el que hibrida. Las sondas pueden tener una longitud de aproximadamente 100 pb o más. Por ejemplo, las sondas pueden ser de al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900 o 1000 pb de longitud.

En algunos casos, las sondas pueden comprender una secuencia de polinucleótidos que es complementaria a un ácido nucleico de interés y una o más secuencias de polinucleótidos que no son complementarias a un ácido nucleico de interés (es decir, secuencias no complementarias). Las secuencias no complementarias pueden encontrarse, por ejemplo, en el extremo 5' y/o 3' de una sonda. En algunos casos, las secuencias no complementarias pueden comprender secuencias de nucleótidos que no existen en el organismo de interés y/o secuencias que no son capaces de hibridar con ninguna secuencia en el genoma humano. Por ejemplo, las secuencias no complementarias pueden derivar de cualquier genoma no humano conocido en la técnica, tal como, por ejemplo, genomas de animales no mamíferos, genomas de plantas, genomas de hongos, genomas bacterianos o genomas virales. En algunos casos, una secuencia no complementaria es del genoma de PhiX 174. En algunos casos, una secuencia no complementaria puede comprender nucleótidos modificados o sintéticos que no son capaces de hibridar con un nucleótido complementario.

Las sondas pueden diseñarse y sintetizarse de acuerdo con métodos conocidos en la técnica y descritos en el presente documento para oligonucleótidos (por ejemplo, oligonucleótidos de captura). Las sondas también pueden incluir cualquiera de las propiedades conocidas en la técnica y descritas en el presente documento, para oligonucleótidos. Las sondas, en el presente documento, pueden diseñarse de manera que comprendan nucleótidos (por ejemplo, adenina (A), timina (T), citosina (C), guanina (G) y uracilo (U)), nucleótidos modificados (p. ej., pseudouridina, dihidouridina, inosina (I), y 7-metilguanosina), nucleótidos sintéticos, bases degeneradas (p. ej., 6H,8H-3,4-dihidropirimido [4,5-c] [1,2] oxazin-7-ona (P), 2-amino-6-metoxiaminopurina (K), N6-metoxiadenina (Z) e hipoxantina (I)), bases universales y/o monómeros distintos de los nucleótidos, nucleótidos modificados o nucleótidos sintéticos, o combinaciones de los mismos, y generalmente se diseñan de tal manera que inicialmente tienen longitudes más largas que los fragmentos con los que hibridan.

En algunos casos, una sonda comprende una pluralidad de monómeros que son capaces de hibridar con cualquiera de las versiones de origen natural o modificadas de nucleótidos como la adenina (A), timina (T), citosina (C), guanina (G) y uracilo (U)). En algunos casos, una sonda comprende una pluralidad de monómeros que son capaces de hibridar con al menos tres de adenina, timina, citosina y guanina. Por ejemplo, una sonda puede incluir una especie de monómero que sea capaz de hibridar con A, T y C; A, T y G; G, C y T; o G, C y A. En algunos casos, una sonda comprende una pluralidad de monómeros que son capaces de hibridar con todos los nucleótidos de adenina, timina, citosina y guanina. Por ejemplo, una sonda puede incluir una especie de monómero que sea capaz de hibridar con A,

T, C y G. En algunos casos, las condiciones de hibridación (por ejemplo, la rigurosidad) se pueden ajustar de acuerdo con los métodos descritos en el presente documento, por ejemplo, para facilitar la hibridación de ciertas especies de monómeros con varias especies de nucleótidos. En algunos casos, los monómeros incluyen nucleótidos. En algunos casos, los monómeros incluyen nucleótidos de origen natural. En algunos casos, los monómeros incluyen nucleótidos modificados.

En algunos casos, los monómeros de una sonda incluyen inosina. La inosina es un nucleótido que se encuentra comúnmente en los ARNt y es capaz, en algunos casos, de hibridar con A, T y C. El ejemplo 9, en el presente documento, describe un método que utiliza sondas de poliinosina para la determinación del tamaño de los fragmentos de ácidos nucleicos. En algunos casos, las sondas de poliinosina se hibridan con los fragmentos de ácido nucleico en condiciones de hibridación poco rigurosas o no rigurosas (tales como, por ejemplo, baja temperatura y/o elevada sal en comparación con las condiciones de hibridación rigurosas descritas en el presente documento). En algunos casos, los fragmentos de ácido nucleico se tratan con bisulfito de sodio, lo que provoca la desaminación de los restos de citosina no metilados en los fragmentos para formar restos de uracilo. En algunos casos, los fragmentos de ácidos nucleicos tratados con bisulfito de sodio se amplifican (por ejemplo, se amplifican por PCR) antes del tratamiento con bisulfito de sodio. En algunos casos, los fragmentos de ácidos nucleicos se unen a una secuencia que comprende un sitio de cebador de amplificación universal que no tiene residuos de citosina. A continuación, se puede generar una segunda hebra complementaria, por ejemplo, usando un cebador de amplificación universal y una reacción de extensión. Generalmente, los restos de uracilo en la primera hebra generan restos de adenina complementarios en la segunda hebra. Por lo tanto, se puede generar una segunda hebra que no tiene restos de guanina. Dichas segundas hebras complementarias libres de guanina, en algunos casos, pueden hibridar con sondas de poliinosina en condiciones de hibridación rigurosas.

En algunos casos, los monómeros de una sonda incluyen monómeros de base universal. Los monómeros de base universal típicamente son análogos de nucleobase o monómeros sintéticos que pueden hibridar de forma no selectiva con cada una de las bases naturales (por ejemplo, A, G, C, T). Por lo tanto, una sonda que comprende monómeros de base universal a veces puede hibridar con un fragmento de ácido nucleico independientemente de la secuencia de nucleótidos. Las bases universales pueden incluir, sin limitación, 3-nitropirrol, 4-nitroindol, 5-nitroindol, 6-nitroindol, 3-metil 7-propinil isocarboestirilo (PIM), 3-metil isocarboestirilo (MICS) y 5-metil isocarboestirilo (5MICS) (véase, por ejemplo, Nichols et al. (1994) *Nature* 369, 492-493; Bergstrom et al. (1995) *J. Am. Chem. Soc.* 117, 1201-1209; Loakes y Brown (1994) *Nucleic Acids Res.* 22, 4039-4043; Lin y Brown (1992) *Nucleic Acids Res.* 20, 5149-5152; Lin y Brown (1989) *Nucleic Acids Res.* 17, 10383; Brown y Lin (1991) *Carbohydrate Research* 216, 129-139; Berger et al. (2000) *Nucleic Acids Res.* 28(15): 2911-2914).

En algunos casos, los monómeros de una sonda incluyen monómeros no nucleotídicos. En algunos casos, los monómeros incluyen subunidades de un polímero sintético. En algunos casos, los monómeros incluyen pirrolidona. La pirrolidona es un monómero de la polipirrolidona polimérica sintética y es capaz, en algunos casos, de hibridar con A, T, G y C.

En algunos casos, un método para determinar la longitud de los fragmentos incluye la etapa de poner en contacto, en condiciones de emparejamiento, fragmentos de ácido nucleico (p. ej., fragmentos diana y/o de referencia) con una pluralidad de sondas que pueden emparejarse con los fragmentos, generando, de este modo, especies fragmentos-sondas como, por ejemplo, especies sonda-diana y especies sonda-referencia. Las sondas y/o las condiciones de hibridación (por ejemplo, rigurosidad) pueden optimizarse para favorecer la unión de fragmentos completa o sustancialmente completa (por ejemplo, alta rigurosidad). Las hibridaciones de fragmentos-sonda completas o sustancialmente completas generalmente incluyen dúplex donde el fragmento no comprende porciones no hibridadas y la sonda puede comprender porciones no hibridadas, como se describe con más detalle a continuación.

En algunos casos, tales como cuando la longitud de la sonda es más larga que la longitud del fragmento, la especie sonda-diana y/o la especie sonda-referencia pueden comprender, cada una, porciones de la sonda no hibridadas (es decir, porciones de sonda monocatenaria; véase, por ejemplo, la FIG. 12). Las porciones de la sonda no hibridadas pueden estar en cualquiera de los extremos de la sonda (por ejemplo, el extremo 3' o 5' de una sonda) o en ambos extremos de la sonda (es decir, los extremos 3' y 5' de una sonda) y pueden comprender cualquier número de monómeros. En algunos casos, las porciones de sonda no hibridadas pueden comprender de aproximadamente 1 a aproximadamente 500 monómeros. Por ejemplo, las porciones de sonda no hibridadas pueden comprender aproximadamente 5, 10, 20, 30, 40, 50, 100, 200, 300 o 400 monómeros.

En algunos casos, las porciones de sonda no hibridadas pueden eliminarse de la especie sonda-diana y/o sonda-referencia, generando de este modo, sondas recortadas. La eliminación de las porciones de sonda no hibridadas se puede lograr mediante cualquier método conocido en la técnica para escindir y/o digerir un polímero, tal como, por ejemplo, un método para escindir o digerir un ácido nucleico monocatenario. Las porciones de la sonda sin hibridar se pueden eliminar del extremo 5' de la sonda y/o del extremo 3' de la sonda. Tales métodos pueden comprender el uso de escisión o digestión química y/o enzimática. En algunos casos, se usa una enzima capaz de escindir los enlaces fosfodiéster entre subunidades de nucleótidos de un ácido nucleico para eliminar las porciones de sonda no hibridadas. Tales enzimas pueden incluir, sin limitación, nucleasas (p.ej., DNasa I, RNasa I), endonucleasas (por ejemplo, nucleasa de frijol mung, nucleasa S1 y similares), nucleasas de restricción, exonucleasas (por ejemplo, Exonucleasa

I, Exonucleasa III, Exonucleasa T, Exonucleasa T7, Exonucleasa Lambda y similares), fosfodiesterasas (p. ej., Fosfodiesterasa II, fosfodiesterasa de bazo de ternera, fosfodiesterasa de veneno de serpiente y similares), desoxirribonucleasas (DNAsa), ribonucleasas (RNAsa), endonucleasas Flap, 5' nucleasas, 3' nucleasas, 3'-5' exonucleasas, 5'-3' exonucleasas y similares, o combinaciones de las mismas. Las sondas recortadas generalmente tienen la misma o sustancialmente la misma longitud que el fragmento con el que hibridan. Por lo tanto, la determinación de la longitud de una sonda recortada en el presente documento puede proporcionar una medida de la longitud del fragmento de ácido nucleico correspondiente. La longitud de la sonda recortada se puede medir utilizando cualquiera de los métodos conocidos en la técnica o descritos en el presente documento, para determinar la longitud de fragmentos de ácidos nucleicos. En algunos casos, las sondas pueden contener una molécula o entidad detectable para facilitar la detección y/o la determinación de la longitud (por ejemplo, un fluoróforo, radioisótopo, agente colorimétrico, partículas, enzima y similares). La longitud de la sonda recortada se puede evaluar con o sin separar los productos de las porciones no hibridadas después de eliminarlos.

En algunos casos, las sondas recortadas se disocian (es decir, se separan) de sus fragmentos de ácidos nucleicos correspondientes. Las sondas pueden separarse de sus correspondientes fragmentos de ácidos nucleicos utilizando cualquier método conocido en la técnica, incluyendo, pero sin limitación, desnaturalización térmica. Las sondas recortadas se pueden distinguir de los fragmentos de ácidos nucleicos correspondientes mediante un método conocido en la técnica o descrito en el presente documento para etiquetar y/o aislar una especie de molécula en una mezcla. Por ejemplo, una sonda y/o un fragmento de ácido nucleico pueden comprender una propiedad detectable de modo que una sonda se pueda distinguir del ácido nucleico con el que hibrida. Ejemplos no limitantes de propiedades detectables incluyen propiedades ópticas, propiedades eléctricas, propiedades magnéticas, propiedades químicas y tiempo y/o velocidad a través de una abertura de tamaño conocido. En algunos casos, las sondas y los fragmentos de ácidos nucleicos de muestra están físicamente separados entre sí. La separación se puede lograr, por ejemplo, utilizando ligandos de captura, como biotina u otros ligandos de afinidad, y agentes de captura, como avidina, estreptavidina, un anticuerpo o un receptor. Una sonda o fragmento de ácido nucleico puede contener un ligando de captura que tiene actividad de unión específica para un agente de captura. Por ejemplo, los fragmentos de una muestra de ácido nucleico se pueden biotinar o unir a un ligando de afinidad usando métodos bien conocidos en la técnica y separar de las sondas usando un ensayo descendente con perlas recubiertas con estreptavidina, por ejemplo. En algunos casos, se puede usar un ligando de captura y un agente de captura o cualquier otro resto (por ejemplo, marcador de masa) para agregar masa a los fragmentos de ácidos nucleicos de manera que puedan excluirse del intervalo de masas de las sondas detectadas en un espectrómetro de masas. En algunos casos, la masa se agrega a las sondas, por medio de los propios monómeros y/o la adición de un marcador de masa, para cambiar el intervalo de masas fuera del intervalo de masas para los fragmentos de ácidos nucleicos.

35 *Biblioteca de ácidos nucleicos*

En algunos casos, una biblioteca de ácidos nucleicos es una pluralidad de moléculas de polinucleótidos (por ejemplo, una muestra de ácidos nucleicos) que se preparan, ensamblan y/o modifican para un proceso específico, cuyos ejemplos no limitantes incluyen la inmovilización en una fase sólida (por ejemplo, un soporte sólido, por ejemplo, una celda de flujo, una perla), enriquecimiento, amplificación, clonación, detección y/o secuenciación de ácidos nucleicos. En ciertos casos, se prepara una biblioteca de ácidos nucleicos antes o durante un proceso de secuenciación. Se puede preparar una biblioteca de ácidos nucleicos (por ejemplo, una biblioteca de secuenciación) mediante un método adecuado como se conoce en la técnica. Se puede preparar una biblioteca de ácidos nucleicos mediante un proceso de preparación dirigido o no dirigido.

En algunos casos, una biblioteca de ácidos nucleicos se modifica para que comprenda un resto químico (por ejemplo, un grupo funcional) configurado para la inmovilización de ácidos nucleicos en un soporte sólido. En algunos casos, una biblioteca de ácidos nucleicos se modifica para que comprenda una biomolécula (por ejemplo, un grupo funcional) y/o un miembro de un par de unión configurado para la inmovilización de la biblioteca en un soporte sólido, cuyos ejemplos no limitantes incluyen globulina de unión a tiroxina, proteínas de unión a esteroides, anticuerpos, antígenos, haptenos, enzimas, lectinas, ácidos nucleicos, represores, proteína A, proteína G, avidina, estreptavidina, biotina, componentes del complemento C1q, proteínas de unión a ácidos nucleicos, receptores, hidratos de carbono, oligonucleótidos, polinucleótidos, secuencias de ácidos nucleicos complementarias, similares y combinaciones de los mismos. Algunos ejemplos de pares de unión específicos incluyen, sin limitación: un resto de avidina y un resto de biotina; un epítipo antigénico y un anticuerpo o fragmento inmunológicamente reactivo del mismo; un anticuerpo y un hapteno; un resto de digoxígeno y un anticuerpo anti-digoxígeno; un resto de fluoresceína y un anticuerpo anti-fluoresceína; un operador y un represor; una nucleasa y un nucleótido; una lectina y un polisacárido; un esteroide y una proteína de unión a esteroides; un compuesto activo y un receptor de compuesto activo; una hormona y un receptor de hormonas; una enzima y un sustrato; una inmunoglobulina y proteína A; oligonucleótido o polinucleótido y su complemento correspondiente; similares o combinaciones de los mismos.

En algunos casos, una biblioteca de ácidos nucleicos se modifica para que comprenda uno o más polinucleótidos de composición conocida, cuyos ejemplos no limitantes incluyen un identificador (por ejemplo, un marcador, un marcador de indexación), una secuencia de captura, una etiqueta, un adaptador, un sitio de enzimas de restricción, un promotor, un potenciador, un origen de replicación, un tallo lazo, una secuencia complementaria (por ejemplo, un sitio de unión a cebador, un sitio de emparejamiento), un sitio de integración adecuado (por ejemplo, un transposón, un sitio de

integración viral), un nucleótido modificado, similares o combinaciones de los mismos. Los polinucleótidos de secuencia conocida se pueden añadir en una posición adecuada, por ejemplo, en el extremo 5', extremo 3' o dentro de una secuencia de ácido nucleico. Los polinucleótidos de secuencia conocida pueden ser secuencias iguales o diferentes. En algunos casos, un polinucleótido de secuencia conocida se configura para hibridar con uno o más oligonucleótidos inmovilizados en una superficie (por ejemplo, una superficie en una celda de flujo). Por ejemplo, una molécula de ácido nucleico que comprende una secuencia 5' conocida puede hibridar con una primera pluralidad de oligonucleótidos, mientras que la secuencia 3' conocida puede hibridar con una segunda pluralidad de oligonucleótidos. En algunos casos, una biblioteca de ácidos nucleicos puede comprender marcadores específicos de cromosoma, secuencias de captura, etiquetas y/o adaptadores. En algunos casos, una biblioteca de ácidos nucleicos comprende una o más etiquetas detectables. En algunos casos, se pueden incorporar una o más etiquetas detectables en una biblioteca de ácidos nucleicos en un extremo 5', en un extremo 3' y/o en cualquier posición de nucleótido dentro de un ácido nucleico en la biblioteca. En algunos casos, una biblioteca de ácidos nucleicos comprende oligonucleótidos hibridados. En ciertos casos, los oligonucleótidos hibridados son sondas etiquetadas. En algunos casos, una biblioteca de ácidos nucleicos comprende sondas de oligonucleótidos hibridadas antes de la inmovilización en una fase sólida.

En algunos casos, un polinucleótido de secuencia conocida comprende una secuencia universal. Una secuencia universal es una secuencia de nucleótidos específica que se integra en dos o más moléculas de ácido nucleico o dos o más subconjuntos de moléculas de ácido nucleico donde la secuencia universal es la misma para todas las moléculas o subconjuntos de moléculas en las que está integrada. Una secuencia universal a menudo se diseña para hibridar y/o amplificar una pluralidad de secuencias diferentes utilizando un único cebador universal que es complementario de una secuencia universal. En algunos casos, se usan dos (por ejemplo, un par) o más secuencias universales y/o cebadores universales. Un cebador universal a menudo comprende una secuencia universal. En algunos casos, los adaptadores (por ejemplo, adaptadores universales) comprenden secuencias universales. En algunos casos, se usan una o más secuencias universales para capturar, identificar y/o detectar múltiples especies o subconjuntos de ácidos nucleicos.

En ciertos casos de preparación de una biblioteca de ácidos nucleicos, (por ejemplo, en ciertos procedimientos de secuenciación por síntesis), los ácidos nucleicos se seleccionan por tamaño y/o se fragmentan en longitudes de varios cientos de pares de bases, o menos (por ejemplo, en la preparación para la generación de bibliotecas). En algunos casos, la preparación de la biblioteca se realiza sin fragmentación (por ejemplo, cuando se utiliza ADNccf).

En ciertos casos, se usa un método de preparación de biblioteca basado en unión (por ejemplo, ILLUMINA TRUSEQ, Illumina, San Diego CA). Los métodos de preparación de bibliotecas basados en la unión, a menudo hacen uso de un diseño de adaptador (por ejemplo, un adaptador metilado) que puede incorporar una secuencia de índice en la etapa de unión inicial y, a menudo, se puede usar para preparar muestras para secuenciación de lectura única, secuenciación de extremo pareado y secuenciación multiplexada. Por ejemplo, a veces los ácidos nucleicos (p. ej., ácidos nucleicos fragmentados o ADNccf) se reparan al final mediante una reacción de ajuste, una reacción con exonucleasa o una combinación de las mismas. En algunos casos, el ácido nucleico reparado en el extremo romo resultante puede extenderse a continuación, por un solo nucleótido, que es complementario a un solo nucleótido saliente en el extremo 3' de un adaptador/cebador. Se puede usar cualquier nucleótido para los nucleótidos de extensión/salientes. En algunos casos, la preparación de la biblioteca de ácidos nucleicos comprende unir un oligonucleótido adaptador. Los oligonucleótidos adaptadores a menudo son complementarios a los anclajes de las celdas de flujo y, a veces, se utilizan para inmovilizar una biblioteca de ácidos nucleicos en un soporte sólido, como la superficie interior de una celda de flujo, por ejemplo. En algunos casos, un oligonucleótido adaptador comprende un identificador, uno o más sitios de hibridación de cebadores de secuenciación (por ejemplo, secuencias complementarias a cebadores de secuenciación universales, cebadores de secuenciación de extremos únicos, cebadores de secuenciación de extremos pareados, cebadores de secuenciación multiplexados y similares), o combinaciones de los mismos (por ejemplo, adaptador/secuenciación, adaptador/identificador, adaptador/identificador/secuenciación).

Un identificador puede ser una etiqueta detectable adecuada incorporada o unida a un ácido nucleico (por ejemplo, un polinucleótido) que permite la detección y/o identificación de ácidos nucleicos que comprenden el identificador. En algunos casos, se incorpora un identificador en, o se une a, un ácido nucleico durante un método de secuenciación (por ejemplo, mediante una polimerasa). Ejemplos no limitantes de identificadores incluyen marcadores de ácidos nucleicos, índices o códigos de barras de ácidos nucleicos, una etiqueta radioactiva (por ejemplo, un isótopo), etiqueta metálica, una etiqueta fluorescente, una etiqueta quimioluminiscente, una etiqueta fosforescente, un desactivador de fluoróforos, un colorante, una proteína (por ejemplo, una enzima, un anticuerpo o parte del mismo, un enlazador, un miembro de un par de unión), similares o combinaciones de los mismos. En algunos casos, un identificador (por ejemplo, un índice o un código de barras de ácidos nucleicos) es una secuencia única, conocida y/o identificable de nucleótidos o análogos de nucleótidos. En algunos casos, los identificadores son seis o más nucleótidos contiguos. Una multitud de fluoróforos están disponibles con una variedad de espectros de excitación y emisión diferentes. Cualquier tipo y/o número adecuado de fluoróforos puede usarse como un identificador. En algunos casos, se utilizan 1 o más, 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 20 o más, 30 o más o 50 o más identificadores diferentes en un método descrito en el presente documento (por ejemplo, un método de detección y/o secuenciación de ácidos nucleicos). En algunos casos, uno o dos tipos de identificadores (por ejemplo, etiquetas fluorescentes) se unen a cada ácido nucleico en una biblioteca. La detección y/o cuantificación de un identificador se puede realizar mediante un método o una máquina adecuados, cuyos ejemplos no limitantes incluyen

citometría de flujo, reacción en cadena de la polimerasa cuantitativa (qPCR), electroforesis en gel, un luminómetro, un fluorómetro, un espectrofotómetro, un análisis de chip genético o de micromatrices adecuado, Western Blot, espectrometría de masas, cromatografía, análisis citofluorométrico, microscopía de fluorescencia, un método adecuado de fluorescencia o imágenes digitales, microscopía de barrido con láser confocal, citometría de barrido láser, cromatografía de afinidad, separación manual por lotes, suspensión de campo eléctrico, un método adecuado de secuenciación de ácidos nucleicos y/o una máquina de secuenciación de ácidos nucleicos, similares y combinaciones de los mismos.

En algunos casos, se usa un método de preparación de bibliotecas basado en transposones (por ejemplo, EPICENTRE NEXTERA, Epicentre, Madison WI). Los métodos basados en transposones suelen utilizar la transposición *in vitro* para fragmentar y marcar simultáneamente el ADN en una reacción de un solo tubo (a menudo permitiendo la incorporación de marcadores específicos de la plataforma y códigos de barras opcionales), y preparar bibliotecas listas para el secuenciador.

En algunos casos, se amplifica una biblioteca de ácidos nucleicos o partes de la misma (por ejemplo, amplificada por un método basado en PCR). En algunos casos, un método de secuenciación comprende la amplificación de una biblioteca de ácidos nucleicos. Una biblioteca de ácidos nucleicos puede amplificarse antes o después de la inmovilización en un soporte sólido (por ejemplo, un soporte sólido en una celda de flujo). La amplificación de ácidos nucleicos incluye el proceso de amplificar o aumentar los números de un molde de ácido nucleico y/o de un complemento de la misma que está presente (por ejemplo, en una biblioteca de ácidos nucleicos), al producir una o más copias del molde y/o su complemento. La amplificación puede llevarse a cabo por un método adecuado. Una biblioteca de ácidos nucleicos puede amplificarse mediante un método de termociclado o mediante un método de amplificación isotérmica. En algunos casos, se usa un método de amplificación por círculo rodante. En algunos casos, la amplificación tiene lugar en un soporte sólido (por ejemplo, dentro de una celda de flujo) donde se inmoviliza una biblioteca de ácidos nucleicos o una porción de la misma. En ciertos métodos de secuenciación, se agrega una biblioteca de ácidos nucleicos a una celda de flujo y se inmoviliza por hibridación a anclajes en condiciones adecuadas. Este tipo de amplificación de ácidos nucleicos a menudo se denomina amplificación en fase sólida. En algunos casos de amplificación en fase sólida, todos o una porción de los productos amplificados se sintetizan mediante una extensión iniciada a partir de un cebador inmovilizado. Las reacciones de amplificación en fase sólida son análogas a las amplificaciones estándar en disolución, excepto que al menos uno de los oligonucleótidos de amplificación (por ejemplo, cebadores) está inmovilizado en un soporte sólido.

En algunos casos, la amplificación en fase sólida comprende una reacción de amplificación de ácidos nucleicos que comprende solo una especie de cebador de oligonucleótidos inmovilizado en una superficie. En ciertos casos, la amplificación en fase sólida comprende una pluralidad de diferentes especies de cebadores de oligonucleótidos inmovilizados. En algunos casos, la amplificación en fase sólida puede comprender una reacción de amplificación de ácidos nucleicos que comprende una especie de cebador de oligonucleótidos inmovilizado en una superficie sólida y una segunda especie de cebador de oligonucleótidos diferente en solución. Se pueden usar múltiples especies diferentes de cebadores inmovilizados o basados en soluciones. Los ejemplos no limitantes de reacciones de amplificación de ácidos nucleicos en fase sólida incluyen amplificación interfacial, amplificación puente, PCR en emulsión, amplificación WildFire (por ejemplo, publicación de patente de EE. UU. US20130012399), similares o combinaciones de las mismas.

Secuenciación

En algunos casos, se pueden secuenciar los ácidos nucleicos (por ejemplo, fragmentos de ácido nucleico, ácido nucleico de muestra, ácido nucleico libre de células). En algunos casos, se obtiene una secuencia completa o sustancialmente completa y, a veces, se obtiene una secuencia parcial. En algunos casos, cuando se realiza un método descrito en el presente documento, un ácido nucleico no está secuenciado, y la secuencia de un ácido nucleico no se determina por un método de secuenciación. En algunos casos, la longitud de los fragmentos se determina utilizando un método de secuenciación. En algunos casos, la longitud del fragmento se determina sin utilizar un método de secuenciación. La secuenciación, el mapeo y los métodos analíticos relacionados se describen en el presente documento y son conocidos en la técnica (por ejemplo, Publicación de Solicitud de Patente de Estados Unidos US2009/0029377). Ciertos aspectos de tales procesos se describen en el presente documento, a continuación.

En algunos casos, la longitud de los fragmentos se determina utilizando un método de secuenciación. En algunos casos, la longitud de los fragmentos se determina utilizando una plataforma de secuenciación de extremos pareados. Dichas plataformas implican la secuenciación de ambos extremos de un fragmento de ácido nucleico. En general, las secuencias correspondientes a ambos extremos del fragmento se pueden mapear a un genoma de referencia (por ejemplo, un genoma humano de referencia). En ciertos casos, ambos extremos se secuencian en una longitud de lectura que sea suficiente para mapear, individualmente para cada extremo del fragmento, a un genoma de referencia. A continuación, se describen ejemplos de longitudes de lecturas de secuencia de extremos pareados. En ciertos casos, todas o una porción de las lecturas de secuencia se pueden mapear a un genoma de referencia sin desapareamiento. En algunos casos, cada lectura se mapea independientemente. En algunos casos, la información de ambas lecturas de secuencia (es decir, de cada extremo) se factoriza en el proceso de mapeo. La longitud de un fragmento se puede determinar, por ejemplo, calculando la diferencia entre las coordenadas genómicas mapeadas en cada lectura de

extremos pareados mapeado.

En algunos casos, la longitud del fragmento se determina utilizando un proceso de secuenciación mediante el cual se obtiene una secuencia de nucleótidos completa o sustancialmente completa para el fragmento. Tales procesos de

5 secuenciación incluyen plataformas que generan longitudes de lectura relativamente largas (por ejemplo, Roche 454, Ion Torrent, molécula única (Pacific Biosciences), tecnología SMRT en tiempo real y similares).

En algunos casos, algunos o todos los ácidos nucleicos en una muestra se enriquecen y/o amplifican (por ejemplo, no específicamente, por ejemplo, mediante un método basado en PCR) antes o durante la secuenciación. En ciertos

10 casos, las porciones o subconjuntos específicos de ácidos nucleicos en una muestra se enriquecen y/o amplifican antes o durante la secuenciación. En algunos casos, una porción o subconjunto de un conjunto preseleccionado de ácidos nucleicos se secuencia aleatoriamente. En algunos casos, los ácidos nucleicos en una muestra no se enriquecen y/o amplifican antes o durante la secuenciación.

Como se usa en el presente documento, "lecturas" (es decir, "una lectura", "una lectura de secuencia") son secuencias de nucleótidos cortas producidas por cualquier proceso de secuenciación descrito en el presente documento o conocido en la técnica. Las lecturas se pueden generar desde un extremo de los fragmentos de ácidos nucleicos ("lecturas de un solo extremo"), y en ocasiones se generan desde ambos extremos de los ácidos nucleicos (por ejemplo, lecturas de extremos pareados, lecturas de dos extremos).

15 20

La longitud de una lectura de secuencia a menudo se asocia con la tecnología de secuenciación en particular. Los métodos de alto rendimiento, por ejemplo, proporcionan lecturas de secuencia que pueden variar en tamaño desde decenas hasta cientos de pares de bases (pb). La secuenciación mediante nanoporos, por ejemplo, puede proporcionar lecturas de secuencia que pueden variar en tamaño desde decenas a cientos hasta miles de pares de bases. En algunos casos, las lecturas de secuencia son de una longitud media, mediana, promedio o absoluta de unos 15 pb a unos 900 pb de largo. En ciertos casos, las lecturas de secuencia son de una longitud media, mediana, promedio o absoluta de unos 1000 pb o más.

25 30

En algunos casos, la longitud nominal, promedio, media o absoluta de las lecturas de un solo extremo a veces es de aproximadamente 1 nucleótido a aproximadamente 500 nucleótidos contiguos, aproximadamente 15 nucleótidos contiguos a aproximadamente 50 nucleótidos contiguos, aproximadamente 30 nucleótidos contiguos a aproximadamente 40 nucleótidos contiguos y, a veces, aproximadamente 35 nucleótidos contiguos o aproximadamente 36 nucleótidos contiguos. En ciertos casos, la longitud promedio, media o absoluta de las lecturas de un solo extremo son de aproximadamente 20 a aproximadamente 30 bases, o de aproximadamente 24 a aproximadamente 28 bases de longitud. En ciertos casos, la longitud nominal, promedio, media o absoluta de las lecturas de un solo extremo es de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 o 49 bases de longitud.

35 40

En ciertos casos, la longitud nominal, promedio, media o absoluta de las lecturas de un solo extremo a veces es de aproximadamente 10 nucleótidos contiguos a aproximadamente 25 nucleótidos contiguos (p.ej., aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 nucleótidos de longitud), aproximadamente 15 nucleótidos contiguos a aproximadamente 20 nucleótidos contiguos y, a veces, aproximadamente 17 nucleótidos contiguos, aproximadamente 18 nucleótidos contiguos, aproximadamente 20 nucleótidos contiguos, aproximadamente 25 nucleótidos contiguos, aproximadamente 36 nucleótidos contiguos a aproximadamente 45 nucleótidos contiguos.

45

Las lecturas en general son representaciones de secuencias de nucleótidos en un ácido nucleico físico. Por ejemplo, en una lectura que contiene una representación ATGC de una secuencia, "A" representa un nucleótido de adenina, "T" representa un nucleótido de timina, "G" representa un nucleótido de guanina y "C" representa un nucleótido de citosina, en un ácido nucleico físico. Las lecturas de secuencia obtenidas de la sangre de una gestante pueden leerse de una mezcla de ácido nucleico materno y fetal. Una mezcla de lecturas relativamente cortas se puede transformar mediante los procesos descritos en el presente documento en una representación de un ácido nucleico genómico presente en la gestante y/o en el feto. Una mezcla de lecturas relativamente cortas se puede transformar en una representación de una variación en el número de copias (por ejemplo, una variación en el número de copias maternas y/o fetales), una variación genética o una aneuploidía, por ejemplo. Las lecturas de una mezcla de ácido nucleico materno y fetal se pueden transformar en una representación de un cromosoma compuesto o un segmento del mismo que comprende características de uno o ambos cromosomas materno y fetal. En ciertos casos, la "obtención" de lecturas de secuencias de ácidos nucleicos de una muestra de un sujeto y/o la "obtención" de lecturas de secuencias de ácidos nucleicos de una muestra biológica de una o más personas de referencia, puede implicar la secuenciación directa del ácido nucleico para obtener la información de la secuencia. En algunos casos, la "obtención" puede implicar recibir información de secuencia obtenida directamente de un ácido nucleico por otro.

50 55 60

En algunos casos, se secuencia una fracción del genoma, que a veces se expresa en la cantidad de genoma cubierta por las secuencias de nucleótidos determinadas (por ejemplo, de cobertura inferior a 1). Cuando un genoma se secuencia con una cobertura de aproximadamente 1X, aproximadamente el 100% de la secuencia de nucleótidos del genoma se encuentra representada mediante lecturas. Un genoma también puede secuenciarse con redundancia,

65

donde una región determinada del genoma puede cubrirse con dos o más lecturas o lecturas solapantes (por ejemplo, superior a 1X). En algunos casos, un genoma se secuencía con una cobertura de aproximadamente 0,01 veces a una cobertura de aproximadamente 100X, cobertura de aproximadamente 0,2X a 20X, o cobertura de aproximadamente 0,2X a aproximadamente 1X (por ejemplo, cobertura de aproximadamente 0,02-, 0,03-, 0,04-, 0,05-, 0,06-, 0,07-, 0,08-, 0,09-, 0,1-, 0,2-, 0,3-, 0,4-, 0,5-, 0,6-, 0,7-, 0,8-, 0,9-, 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 15-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90X).

En algunos casos, la cobertura del genoma o la cobertura de secuencias es proporcional al recuento general de lecturas de secuencias. Por ejemplo, los ensayos que generan y/o analizan mayores cantidades de recuentos de lecturas de secuencias generalmente están asociados con niveles más altos de cobertura de secuencias. Los ensayos que generan y/o analizan menores recuentos de lecturas de secuencias generalmente se asocian con niveles más bajos de cobertura de secuencias. En algunos casos, la cobertura de secuencias y/o el recuento de lecturas de secuencias pueden reducirse sin disminuir significativamente la precisión (por ejemplo, sensibilidad y/o especificidad) de un método descrito en el presente documento. Una disminución significativa en la precisión puede ser una disminución en la precisión de aproximadamente un 1% a aproximadamente un 20% en comparación con un método que no utiliza un recuento de lecturas de secuencias reducido. Por ejemplo, una disminución significativa en la precisión puede ser de aproximadamente un 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15% o más disminución. En algunos casos, la cobertura de secuencias y/o el recuento de lecturas de secuencias se reduce en aproximadamente un 50% o más. Por ejemplo, la cobertura de secuencias y/o el recuento de lecturas de secuencias se puede reducir en aproximadamente un 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% o más. En algunos casos, la cobertura de secuencias y/o el recuento de lecturas de secuencias se reduce en aproximadamente un 60% a aproximadamente un 85%. Por ejemplo, la cobertura de secuencias y/o el recuento de lecturas de secuencias se puede reducir en aproximadamente un 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83% o 84%. En algunos casos, la cobertura de secuencias y/o el recuento de lecturas de secuencias se puede reducir eliminando ciertas lecturas de secuencias. En algunos casos, se eliminan las lecturas de secuencias de fragmentos más largos de una longitud particular (por ejemplo, fragmentos más largos de aproximadamente 160 bases).

En algunos casos, se selecciona un subconjunto de lecturas para el análisis y, a veces, se elimina del análisis una porción determinada de las lecturas. La selección de un subconjunto de lecturas puede, en determinados casos, enriquecer para una especie de ácido nucleico (por ejemplo, ácido nucleico fetal). El enriquecimiento de las lecturas del ácido nucleico fetal, por ejemplo, a menudo aumenta la precisión de un método descrito en el presente documento (por ejemplo, la detección de aneuploidía fetal). Sin embargo, la selección y eliminación de lecturas de un análisis a menudo disminuye la precisión de un método descrito en el presente documento (por ejemplo, debido a una mayor varianza). Por lo tanto, sin quedar ligado a teoría alguna, en general, existe una relación de compromiso entre mayor precisión asociada con el enriquecimiento de la lectura fetal y menor precisión asociada con una cantidad reducida de lecturas en métodos que comprenden la selección y/o eliminación de lecturas (por ejemplo, de fragmentos en un intervalo de tamaño particular). En algunos casos, un método comprende seleccionar un subconjunto de lecturas enriquecidas para lecturas del ácido nucleico fetal sin disminuir significativamente la precisión del método. A pesar de este aparente compromiso, se ha determinado, como se describe en el presente documento, que la utilización de un subconjunto de lecturas de secuencias de nucleótidos (por ejemplo, lecturas de fragmentos relativamente cortos), puede mejorar o mantener la precisión de los análisis genéticos fetales. Por ejemplo, en ciertos casos, alrededor del 80% o más de las lecturas de secuencias de nucleótidos se pueden descartar mientras se mantienen los valores de sensibilidad y especificidad que son similares a los valores de un método comparable que no descarta tales lecturas de secuencias de nucleótidos.

En ciertos casos, se selecciona un subconjunto de fragmentos de ácidos nucleicos antes de la secuenciación. En ciertos casos, se pueden usar técnicas basadas en hibridación (por ejemplo, usando matrices de oligonucleótidos) para seleccionar primero las secuencias de ácidos nucleicos de ciertos cromosomas (por ejemplo, cromosomas sexuales y/o un cromosoma potencialmente aneuploide y otros cromosomas no involucrados en la prueba de aneuploidía). En algunos casos, el ácido nucleico se puede fraccionar por tamaño (por ejemplo, por electroforesis en gel, por cromatografía de exclusión por tamaño o por aproximación basada en microfluidos) y en ciertos casos, el ácido nucleico fetal puede enriquecerse seleccionando el ácido nucleico que tiene un peso molecular más bajo (por ejemplo, menos de 300 pares de bases, menos de 200 pares de bases, menos de 150 pares de bases, menos de 100 pares de bases). En algunos casos, puede enriquecerse el ácido nucleico fetal suprimiendo el ácido nucleico materno de fondo, como por ejemplo mediante la adición de formaldehído. En algunos casos, una porción o subconjunto de un conjunto preseleccionado de fragmentos de ácido nucleico se secuencian aleatoriamente. En algunos casos, el ácido nucleico se amplifica antes de la secuenciación. En algunos casos, una porción o subconjunto del ácido nucleico se amplifica antes de la secuenciación.

En algunos casos, se secuencian una muestra de ácido nucleico de un individuo. En ciertos casos, los ácidos nucleicos de cada una de dos o más muestras se secuencian, donde las muestras son de un individuo o de individuos diferentes. En ciertos casos, muestras de ácidos nucleicos de dos o más muestras biológicas se agrupan, donde cada muestra biológica es de un individuo o dos o más individuos y el conjunto se secuencian. En los últimos casos, una muestra de ácido nucleico de cada muestra biológica a menudo se identifica mediante uno o más identificadores únicos o marcadores de identificación.

En algunos casos, un método de secuenciación utiliza identificadores que permiten la multiplexación de las reacciones de secuencia en un proceso de secuenciación. Cuanto mayor sea el número de identificadores únicos, mayor será el número de muestras y/o cromosomas para la detección, por ejemplo, que se pueden multiplexar en un proceso de secuenciación. Se puede realizar un proceso de secuenciación utilizando cualquier número adecuado de
5 identificadores únicos (por ejemplo, 4, 8, 12, 24, 48, 96 o más).

Un proceso de secuenciación a veces hace uso de una fase sólida, y a veces la fase sólida comprende una celda de flujo en donde se puede unir el ácido nucleico de una biblioteca y los reactivos pueden fluir y ponerse en contacto con el ácido nucleico unido. Una celda de flujo a veces incluye carriles para celdas de flujo, y el uso de identificadores
10 puede facilitar el análisis de varias muestras en cada carril. Una celda de flujo a menudo es un soporte sólido que puede configurarse para retener y/o permitir el paso ordenado de soluciones de reactivos sobre analitos unidos. Las celdas de flujo con frecuencia son planas en forma, ópticamente transparentes, generalmente en la escala milimétrica o submilimétrica, y con frecuencia tienen canales o carriles en los que se produce la interacción analito/reactivo. En algunos casos, el número de muestras analizadas en un carril de celdas de flujo dado depende del número de
15 identificadores únicos utilizados durante la preparación de la biblioteca y/o del diseño de la sonda en el carril de celda de flujo único. La multiplexación utilizando 12 identificadores, por ejemplo, permite el análisis simultáneo de 96 muestras (por ejemplo, igual al número de pocillos en una placa de micropocillos de 96 pocillos) en una celda de flujo de 8 carriles. De manera similar, la multiplexación utilizando 48 identificadores, por ejemplo, permite el análisis simultáneo de 384 muestras (por ejemplo, igual al número de pocillos en una placa de micropocillos de 384 pocillos)
20 en una celda de flujo de 8 carriles. Los ejemplos no limitantes de kits de secuenciación multiplex disponibles en el mercado incluyen el kit de oligonucleótidos de preparación de muestras multiplexado de Illumina y los cebadores de secuenciación de multiplexación y el kit de control PhiX (por ejemplo, los números de catálogo PE-400-1001 y PE-400-1002 de Illumina, respectivamente).

Se puede usar cualquier método adecuado de secuenciación de ácidos nucleicos, ejemplos no limitantes de estos incluyen Maxim & Gilbert, métodos de terminación de cadena, secuenciación por síntesis, secuenciación por unión, secuenciación por espectrometría de masas, técnicas basadas en microscopía, similares o combinaciones de los mismos. En algunos casos, se puede utilizar una tecnología de primera generación, tal como, por ejemplo, los métodos de secuenciación de Sanger, incluidos los métodos automatizados de secuenciación de Sanger, incluida la
30 secuenciación de Sanger microfluídica, en un método proporcionado en el presente documento. En algunos casos, se pueden usar tecnologías de secuenciación que incluyen el uso de tecnologías de imágenes de ácidos nucleicos (por ejemplo, microscopía electrónica de transmisión (TEM) y microscopía de fuerza atómica (AFM)). En algunos casos se utiliza un método de secuenciación de alto rendimiento. Los métodos de secuenciación de alto rendimiento generalmente involucran moldes de ADN amplificados de forma clonal o moléculas de ADN individuales que se
35 secuencian de forma masiva en paralelo, a veces dentro de una celda de flujo. Las técnicas de secuenciación de nueva generación (por ejemplo, 2da y 3ra generación) capaces de secuenciar el ADN de una forma masiva en paralelo se pueden usar para los métodos descritos en el presente documento y se denominan colectivamente en el presente documento "secuenciación masiva en paralelo" (MPS). En algunos casos, los métodos de secuenciación MPS utilizan una aproximación dirigida, donde los cromosomas, genes o regiones específicos de interés son secuencias. En ciertos
40 casos, se usa una aproximación no dirigida donde la mayoría o la totalidad de los ácidos nucleicos en una muestra se secuencian, amplifican y/o capturan de manera aleatoria.

En algunos casos, se usa una aproximación de enriquecimiento, amplificación y/o secuenciación. Una aproximación dirigida a menudo aísla, selecciona y/o enriquece un subconjunto de ácidos nucleicos en una muestra para su
45 procesamiento posterior usando oligonucleótidos específicos de secuencia. En algunos casos, se utiliza una biblioteca de oligonucleótidos específicos de secuencia para dirigir (por ejemplo, hibridar con) uno o más conjuntos de ácidos nucleicos en una muestra. Los oligonucleótidos y/o cebadores específicos de secuencia a menudo son selectivos para secuencias particulares (por ejemplo, secuencias únicas de ácidos nucleicos) presentes en uno o más cromosomas, genes, exones, intrones y/o regiones reguladoras de interés. Se puede usar cualquier método o combinación de
50 métodos adecuados para el enriquecimiento, amplificación y/o secuenciación de uno o más subconjuntos de ácidos nucleicos dirigidos. En algunos casos, las secuencias dirigidas se aíslan y/o enriquecen por captura en una fase sólida (por ejemplo, una celda de flujo, una perla) utilizando uno o más anclajes específicos de secuencia. En algunos casos, las secuencias dirigidas se enriquecen y/o amplifican mediante un método basado en polimerasa (por ejemplo, un método basado en PCR, mediante cualquier extensión adecuada basada en polimerasa) utilizando cebadores y/o
55 conjuntos de cebadores específicos de secuencia. Los anclajes específicos de secuencia a menudo se pueden usar como cebadores específicos de secuencia.

La secuenciación MPS a veces hace uso de la secuenciación por síntesis y ciertos procesos de imagen. Una tecnología de secuenciación de ácidos nucleicos que se puede usar en un método descrito en el presente documento
60 es la secuenciación por síntesis y la secuenciación basada en terminador reversible (por ejemplo, Genome Analyzer de Illumina; Genome Analyzer II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA)). Con esta tecnología, se pueden secuenciar millones de fragmentos de ácidos nucleicos (por ejemplo, ADN) en paralelo. En un ejemplo de este tipo de tecnología de secuenciación, se usa una celda de flujo que contiene un portaobjetos ópticamente transparente con 8 carriles individuales en cuyas superficies están anclados los oligonucleótidos unidos (por ejemplo, cebadores adaptadores). Una celda de flujo a menudo es un soporte sólido que puede configurarse para retener y/o permitir el
65 paso ordenado de soluciones de reactivos sobre analitos unidos. Las celdas de flujo con frecuencia son planas en

forma, ópticamente transparentes, generalmente en la escala milimétrica o submilimétrica, y con frecuencia tienen canales o carriles en los que se produce la interacción analito/reactivo.

5 En algunos casos, la secuenciación por síntesis comprende la adición iterativa (por ejemplo, por adición covalente) de un nucleótido a un cebador o una cadena de ácido nucleico preexistente de una manera dirigida por molde. Cada adición iterativa de un nucleótido se detecta y el proceso se repite varias veces hasta que se obtiene una secuencia de una hebra de ácido nucleico. La longitud de una secuencia obtenida depende, en parte, del número de etapas de adición y detección que se realizan. En algunos casos de secuenciación por síntesis, se agregan y detectan uno, dos, tres o más nucleótidos del mismo tipo (por ejemplo, A, G, C o T) en una ronda de adición de nucleótidos. Los
10 nucleótidos se pueden agregar por cualquier método adecuado (por ejemplo, enzimáticamente o químicamente). Por ejemplo, en algunos casos, una polimerasa o una ligasa agrega un nucleótido a un cebador o a una hebra de ácido nucleico preexistente de una manera dirigida por molde. En algunos casos de secuenciación por síntesis, se utilizan diferentes tipos de nucleótidos, análogos de nucleótidos y/o identificadores. En algunos casos, se usan terminadores reversibles y/o identificadores removibles (por ejemplo, escindibles). En algunos casos, se usan nucleótidos y/o
15 análogos de nucleótidos etiquetados fluorescentes. En ciertos casos, la secuenciación por síntesis comprende una escisión (por ejemplo, escisión y eliminación de un identificador) y/o una etapa de lavado. En algunos casos, la adición de uno o más nucleótidos se detecta mediante un método adecuado descrito en el presente documento o conocido en la técnica, cuyos ejemplos no limitantes incluyen cualquier máquina de imágenes adecuada, una cámara adecuada, una cámara digital, una máquina de imágenes basada en un CCD (Dispositivo de Carga Acoplada) (por ejemplo, una cámara CCD), una máquina de imágenes basada en CMOS (Silicio Complementario de Óxido Metálico) (por ejemplo, una cámara CMOS), un fotodiodo (por ejemplo, un tubo fotomultiplicador), microscopio electrónico, un transistor de efecto de campo (por ejemplo, un transistor de efecto de campo de ADN), un sensor iónico ISFET (por ejemplo, un sensor CHEMFET), similares o combinaciones de los mismos. Otros métodos de secuenciación que se pueden usar para llevar a cabo métodos en el presente documento incluyen PCR digital y secuenciación por hibridación.

25 Otros métodos de secuenciación que se pueden usar para llevar a cabo métodos del presente documento incluyen PCR digital y secuenciación por hibridación. En algunos casos, la reacción en cadena de la polimerasa digital (PCR digital o dPCR) se puede usar para identificar y cuantificar directamente los ácidos nucleicos en una muestra. La PCR digital se puede realizar en una emulsión. Por ejemplo, los ácidos nucleicos individuales se separan, por ejemplo, en un dispositivo de cámara microfluídica, y cada ácido nucleico se amplifica individualmente mediante PCR. Los ácidos nucleicos se pueden separar de modo que no haya más de un ácido nucleico por pocillo. En algunos casos, se pueden usar diferentes sondas para distinguir varios alelos (por ejemplo, alelos fetales y alelos maternos). Se pueden enumerar los alelos para determinar el número de copias.

35 En ciertos casos, se puede utilizar la secuenciación por hibridación. El método implica poner en contacto una pluralidad de secuencias polinucleotídicas con una pluralidad de sondas polinucleotídicas, donde cada una de la pluralidad de las sondas polinucleotídicas puede anclar opcionalmente a un sustrato. En algunos casos, el sustrato puede ser una superficie plana con una matriz de secuencias de nucleótidos conocidas. El patrón de hibridación con la matriz se puede usar para determinar las secuencias de polinucleótidos presentes en la muestra. En algunos casos, cada sonda se ancla a una perla, por ejemplo, una perla magnética o similar. La hibridación a las perlas puede identificarse y usarse para identificar la pluralidad de secuencias de polinucleótidos dentro de la muestra.

45 En algunos casos, la secuenciación mediante nanoporos se puede usar en un método descrito en el presente documento. La secuenciación mediante nanoporos es una tecnología de secuenciación de molécula única mediante la cual una única molécula de ácido nucleico (por ejemplo, el ADN) se secuencia directamente cuando pasa a través de un nanoporo.

50 Se puede usar un método, sistema o plataforma tecnológica adecuada de MPS para llevar a cabo los métodos descritos en el presente documento para obtener lecturas de secuenciación de ácidos nucleicos. Ejemplos no limitantes de plataformas MPS incluyen Illumina/Solex/HiSeq (por ejemplo, Genome Analyzer de Illumina; Genome Analyzer II; HISEQ 2000; HISEQ), SOLiD, Roche/454, PACBIO y/o SMRT, Secuenciación de Molécula Única Helicos True, Ion Torrent y secuenciación basada en Ion semiconductor (por ejemplo, desarrollada por Life Technologies), WildFire, tecnologías basadas en Genetic Analyzer 5500, 5500 xl W y/o 5500 xl W (por ejemplo, desarrollados y vendidos por Life Technologies, publicación de patente de EE.UU. nº US20130012399); Secuenciación Polony, Pirosecuenciación, Secuenciación de firma masiva en paralelo (MPSS), secuenciación de ARN polimerasa (RNAP), sistemas y métodos LaserGen, plataformas basadas en nanoporos, matriz de transistores de efecto de campo (CHEMFET) sensibles a los productos químicos, secuenciación basada en microscopía electrónica (por ejemplo, desarrollada por ZS Genetics, Halcyon Molecular), secuenciación con nanoesferas.

60 En algunos casos, se realiza secuenciación específica de cromosomas. En algunos casos, la secuenciación específica de cromosomas se realiza utilizando DANSR (análisis digital de regiones seleccionadas). El análisis digital de las regiones seleccionadas permite la cuantificación simultánea de cientos de loci por catenación dependiente de ADNcf de dos oligonucleótidos específicos de locus a través de un oligonucleótido "puente" intermedio para formar un molde de PCR. En algunos casos, la secuenciación específica de cromosomas se realiza generando una biblioteca enriquecida en secuencias específicas de cromosomas. En algunos casos, las lecturas de secuencia se obtienen solo para un conjunto de cromosomas seleccionados. En algunos casos, las lecturas de secuencia se obtienen solo para

los cromosomas 21, 18 y 13.

Lecturas de mapeo

5 Las lecturas de secuencia se pueden mapear y el número de lecturas que mapean a una región de ácido nucleico específica (por ejemplo, un cromosoma, porción o segmento del mismo) se denominan recuentos. Se puede utilizar cualquier método de mapeo adecuado (por ejemplo, proceso, algoritmo, programa, software, módulo, similares o combinación de los mismos). Ciertos aspectos de procesos de mapeo se describen en el presente documento a continuación.

10 El mapeo de las lecturas de secuencias de nucleótidos (es decir, la información de la secuencia de un fragmento cuya posición genómica física es desconocida) se puede realizar de varias maneras, y con frecuencia comprende la alineación de las lecturas de secuencias obtenidas con una secuencia coincidente en un genoma de referencia. En tales alineaciones, las lecturas de secuencias generalmente se alinean con una secuencia de referencia y las que se alinean se designan "mapeadas", "una lectura de secuencia mapeada" o "una lectura mapeada". En ciertos casos, una lectura de secuencia mapeada se denomina "hit" o "recuento". En algunos casos, las lecturas de secuencia mapeadas se agrupan de acuerdo con diversos parámetros y se asignan a porciones particulares, que se explican con más detalle a continuación.

20 Como se usa en el presente documento, los términos "alineada", "alineación" o "que alinea" se refieren a dos o más secuencias de ácido nucleico que pueden identificarse como una coincidencia (por ejemplo, 100% de identidad) o una coincidencia parcial. Las alineaciones se pueden hacer manualmente o por ordenador (por ejemplo, un software, programa, módulo o algoritmo), cuyos ejemplos no limitantes incluyen el programa informático Efficient Local Alignment of Nucleotide Data (ELAND) distribuido como parte de la línea Illumina Genomics Analysis. La alineación de una lectura de secuencia puede ser una coincidencia de secuencia del 100%. En algunos casos, una alineación es menor que una coincidencia de secuencia del 100% (es decir, coincidencia no perfecta, coincidencia parcial, alineación parcial). Una alineación puede ser aproximadamente un 99%, 98 %, 97 %, 96 %, 95 %, 94 %, 93 %, 92 %, 91 %, 90 %, 89 %, 88 %, 87 %, 86 %, 85 %, 84 %, 83 %, 82 %, 81 %, 80 %, 79 %, 78 %, 77 %, 76 % o 75 % de coincidencia. En algunos casos, una alineación comprende un desapareamiento. En algunos casos, una alineación comprende 1, 2, 3, 4 o 5 desapareamientos. Se pueden alinear dos o más secuencias usando cualquiera de las hebras. En ciertos casos, una secuencia de ácido nucleico se alinea con el complemento inverso de otra secuencia de ácido nucleico.

35 Se pueden usar varios métodos informáticos para mapear cada lectura de secuencia a una porción. Ejemplos no limitantes de algoritmos de ordenador que se pueden usar para alinear secuencias incluyen, sin limitación, BLAST, BLITZ, FASTA, BOWTIE 1, BOWTIE 2, ELAND, MAQ, PROBEMATCH, SOAP o SEQMAP, o variaciones de los mismos o combinaciones de los mismos. En algunos casos, las lecturas de secuencia se pueden alinear con secuencias en un genoma de referencia. En algunos casos, las lecturas de secuencia se pueden encontrar y/o alinear con secuencias en bases de datos de ácidos nucleicos conocidas en la técnica, incluyendo, por ejemplo, GenBank, dbEST, dbSTS, EMBL (Laboratorio Europeo de Biología Molecular) y DDBJ (Banco de Datos de ADN de Japón). Se pueden usar herramientas BLAST o similares para buscar las secuencias identificadas en una base de datos de secuencias. Los resultados de búsqueda se pueden usar para ordenar las secuencias identificadas en porciones apropiadas (descritas a continuación), por ejemplo.

45 En algunos casos, una lectura puede mapearse de forma única o no única a porciones de un genoma de referencia. Una lectura se considera como "mapeada de forma única" si se alinea con una única secuencia en el genoma de referencia. Una lectura se considera como "mapeada de forma no única" si se alinea con dos o más secuencias en el genoma de referencia. En algunos casos, las lecturas mapeadas de forma no única se eliminan para un análisis adicional (por ejemplo, la cuantificación). En ciertos casos, se puede permitir que un cierto grado pequeño de desapareamiento (0-1) tenga en cuenta los polimorfismos de un solo nucleótido que pueden existir entre el genoma de referencia y las lecturas de las muestras individuales que se están mapeando. En algunos casos, no se permite un grado de desapareamiento para una lectura mapeada a una secuencia de referencia.

55 Como se usa en el presente documento, la expresión "genoma de referencia" puede referirse a cualquier genoma conocido particular, secuenciado o caracterizado, ya sea parcial o completo, de cualquier organismo o virus que pueda usarse para hacer referencia a secuencias identificadas de un sujeto. Por ejemplo, un genoma de referencia utilizado para sujetos humanos, así como muchos otros organismos, se puede encontrar en el Centro Nacional de Información Biotecnológica en www.ncbi.nlm.nih.gov. Un "genoma" se refiere a la información genética completa de un organismo o virus, expresada en secuencias de ácidos nucleicos. Como se usa en el presente documento, una secuencia de referencia o genoma de referencia a menudo es una secuencia genómica ensamblada o parcialmente ensamblada de un individuo o individuos múltiples. En algunos casos, un genoma de referencia es una secuencia genómica ensamblada o parcialmente ensamblada de uno o más individuos humanos. En algunos casos, un genoma de referencia comprende secuencias asignadas a cromosomas.

65 En ciertos casos, cuando una muestra de ácido nucleico es de una gestante, a veces una secuencia de referencia no es del feto, la madre del feto o el padre del feto, y se denomina en el presente documento como una "referencia externa". En algunos casos, se puede preparar y utilizar una referencia materna. Cuando se prepara una referencia

de la gestante ("secuencia de referencia materna") basada en una referencia externa, las lecturas del ADN de la gestante que contiene sustancialmente ADN no fetal a menudo se mapean a la secuencia de referencia externa y se ensamblan. En ciertos casos, la referencia externa es del ADN de un individuo que tiene sustancialmente la misma etnia que la gestante. Una secuencia de referencia materna puede no cubrir completamente el ADN genómico materno (por ejemplo, puede cubrir aproximadamente el 50%, 60 %, 70 %, 80 %, 90% o más del ADN genómico materno) y la referencia materna puede no coincidir perfectamente con la secuencia del ADN genómico materno (por ejemplo, la secuencia de referencia materna puede incluir múltiples desapareamientos).

En ciertos casos, la capacidad de mapeo se puede evaluar para una región genómica (por ejemplo, porción, porción genómica, porción). La capacidad de mapeo es la capacidad de alinear de forma inequívoca una lectura de una secuencia de nucleótidos con una porción de un genoma de referencia, por lo general hasta un número específico de desapareamientos, incluyendo, por ejemplo, 0, 1, 2 o más desapareamientos. Para una región genómica dada, la capacidad de mapeo esperada se puede estimar utilizando una aproximación de ventana deslizante de una longitud de lectura preestablecida y realizando el promedio de los valores de capacidad de mapeo de nivel de lecturas resultantes. Las regiones genómicas que comprenden tramos de una secuencia de nucleótidos única a veces tienen un alto valor de capacidad de mapeo.

Porciones

En algunos casos, las lecturas de secuencia mapeadas (es decir, marcadores de secuencia) se agrupan de acuerdo con diversos parámetros y se asignan a porciones particulares (por ejemplo, porciones de un genoma de referencia). A menudo, se pueden usar lecturas de secuencias mapeadas individuales para identificar una porción (por ejemplo, la presencia, ausencia o cantidad de una porción) presente en una muestra. En algunos casos, la cantidad de una porción es indicativa de la cantidad de una secuencia más grande (por ejemplo, un cromosoma) en la muestra. El término "porción" también se puede denominar en el presente documento "sección genómica", "fracción", "región", "partición", "porción de un genoma de referencia", "porción de un cromosoma" o "porción genómica". En algunos casos, una porción es un cromosoma completo, un segmento de un cromosoma, un segmento de un genoma de referencia, un segmento que abarca varios cromosomas, segmentos de varios cromosomas y/o combinaciones de los mismos. Una porción se predefine en función de parámetros específicos (por ejemplo, indicadores). En algunos casos, una porción se define de manera arbitraria o no arbitraria basándose en la partición de un genoma (por ejemplo, particionado por tamaño, en el contenido de GC, regiones contiguas, regiones contiguas de un tamaño definido arbitrariamente y similares).

En algunos casos, una porción se delinea en función de uno o más parámetros que incluyen, por ejemplo, la longitud o una característica o características particulares de la secuencia. Las porciones se pueden seleccionar, filtrar y/o eliminar del estudio utilizando cualquier criterio adecuado conocido en la técnica o descrito en el presente documento. En algunos casos, una porción se basa en una longitud particular de secuencia genómica. En algunos casos, un método puede incluir el análisis de múltiples lecturas de secuencias mapeadas en una pluralidad de porciones. Las porciones pueden tener aproximadamente la misma longitud o las porciones pueden tener diferentes longitudes. En algunos casos, las porciones tienen aproximadamente la misma longitud. En algunos casos, porciones de diferentes longitudes se ajustan o ponderan. En algunos casos, una porción es de aproximadamente 10 kilobases (kb) a aproximadamente 100 kb, aproximadamente 20 kb a aproximadamente 80 kb, aproximadamente 30 kb a aproximadamente 70 kb, aproximadamente 40 kb a aproximadamente 60 kb y, algunas veces, de aproximadamente 50 kb. En algunos casos, una porción es de aproximadamente 10 kb a aproximadamente 20 kb. Una porción no se limita a series de secuencia contiguas. Por lo tanto, las porciones pueden estar compuestas por secuencias contiguas y/o no contiguas. Una porción no se limita a un solo cromosoma. En algunos casos, una porción incluye la totalidad o parte de un cromosoma o la totalidad o parte de dos o más cromosomas. En algunos casos, las porciones pueden abarcar uno, dos o más cromosomas completos. Además, las porciones pueden abarcar regiones unidas o no unidas de múltiples cromosomas.

En algunos casos, las porciones pueden ser segmentos cromosómicos particulares en un cromosoma de interés, tal como, por ejemplo, un cromosoma en donde se evalúa una variación genética (por ejemplo, una aneuploidía de los cromosomas 13, 18 y/o 21 o un cromosoma sexual). Una porción también puede ser un genoma patógeno (por ejemplo, bacteriano, fúngico o viral) o un fragmento del mismo. Las porciones pueden ser genes, fragmentos de genes, secuencias reguladoras, intrones, exones y similares.

En algunos casos, un genoma (por ejemplo, un genoma humano) se divide en porciones en función del contenido de información de regiones particulares. En algunos casos, la partición de un genoma puede eliminar regiones similares (por ejemplo, regiones o secuencias idénticas u homólogas) en todo el genoma y solo mantener regiones únicas. Las regiones eliminadas durante la partición pueden estar dentro de un solo cromosoma o pueden abarcar múltiples cromosomas. En algunos casos, un genoma particionado se recorta y optimiza para una alineación más rápida, lo que a menudo permite la aproximación de secuencias identificables de forma única.

En algunos casos, la partición puede reducir la ponderación de regiones similares. Un proceso para la reducción de la ponderación de una porción se describe con más detalle a continuación.

En algunos casos, la partición de un genoma en regiones que trascienden los cromosomas puede basarse en la ganancia de información producida en el contexto de la clasificación. Por ejemplo, el contenido de la información se puede cuantificar utilizando un perfil de valor p que mide la significación de ubicaciones genómicas particulares para distinguir entre grupos de sujetos normales y anormales confirmados (por ejemplo, sujetos euploides y con trisomía, respectivamente). En algunos casos, la partición de un genoma en regiones que trascienden los cromosomas puede basarse en cualquier otro criterio, tal como, por ejemplo, velocidad/conveniencia mientras se alinean marcadores, contenido de GC (por ejemplo, contenido de GC alto o bajo), uniformidad del contenido de GC, otras medidas de contenido de secuencias (por ejemplo, fracción de nucleótidos individuales, fracción de pirimidinas o purinas, fracción de ácidos nucleicos naturales frente a no naturales, fracción de nucleótidos metilados y contenido de CpG), estado de metilación, temperatura de fusión del dúplex, capacidad de secuenciación o PCR, valor de incertidumbre asignado a porciones individuales de un genoma de referencia y/o una búsqueda dirigida para características particulares.

Un "segmento" de un cromosoma generalmente es parte de un cromosoma, y típicamente es una parte diferente de un cromosoma que una porción. Un segmento de un cromosoma a veces se encuentra en una región diferente de un cromosoma respecto a una porción, a veces no comparte un polinucleótido con una porción y, a veces, incluye un polinucleótido que está en una porción. Un segmento de un cromosoma a menudo contiene un número mayor de nucleótidos que una porción (por ejemplo, un segmento a veces incluye una porción), y algunas veces un segmento de un cromosoma contiene un número menor de nucleótidos que una porción (por ejemplo, un segmento a veces está dentro de una porción).

Selección de porciones

Se ha observado que los fragmentos de ácidos nucleicos libres de células circulantes (fragmentos CCF) obtenidos de una gestante comprenden generalmente fragmentos de ácido nucleico que se originan a partir de células fetales (es decir, fragmentos fetales) y fragmentos de ácido nucleico que se originan a partir de células maternas (es decir, fragmentos maternos). Las lecturas de secuencia derivadas de fragmentos CCF procedentes de un feto se denominan en el presente documento "lecturas fetales". Las lecturas de secuencia derivadas de fragmentos de CCF procedentes del genoma de una gestante (por ejemplo, una madre) que tiene un feto se denominan en el presente documento "lecturas maternas". Los fragmentos CCF a partir de los cuales se obtienen lecturas fetales se denominan en el presente documento moldes fetales y los fragmentos CCF a partir de los cuales se obtienen lecturas maternas se denominan en el presente documento moldes maternos.

También se ha observado que en los fragmentos CCF, los fragmentos fetales generalmente son relativamente cortos (por ejemplo, aproximadamente 200 pares de bases de longitud o menos) y que los fragmentos maternos incluyen tales fragmentos relativamente cortos y fragmentos relativamente más largos. Se puede identificar un subconjunto de porciones a las que se mapea una cantidad significativa de lecturas de fragmentos relativamente cortos. Sin quedar ligado a teoría alguna, se espera que las lecturas mapeadas en dichas porciones estén enriquecidas para lecturas fetales, lo que puede mejorar la precisión de un análisis genético fetal (por ejemplo, detectar la presencia o ausencia de una variación genética fetal (por ejemplo, aneuploidía de cromosomas fetales (por ejemplo, T21, T18 y/o T13))).

Sin embargo, a menudo no se considera un número significativo de lecturas, cuando un análisis genético fetal se basa en un subconjunto de lecturas. La selección de un subconjunto de lecturas mapeadas en un subconjunto seleccionado de porciones, y la eliminación de lecturas en porciones no seleccionadas, para un análisis genético fetal puede disminuir la precisión del análisis genético, debido a una mayor varianza, por ejemplo. En algunos casos, aproximadamente del 30% a aproximadamente el 70% (por ejemplo, aproximadamente el 35%, 40%, 45%, 50%, 55%, 60%, o 65%) de las lecturas de secuenciación obtenidas de un mapa del sujeto o muestra, se eliminan del estudio al seleccionar un subconjunto de porciones para un análisis genético fetal. En ciertos casos, aproximadamente del 30% a aproximadamente el 70% (por ejemplo, aproximadamente el 35%, 40%, 45%, 50%, 55%, 60%, o 65%) de las lecturas de secuenciación obtenidas de un sujeto o muestra mapean a un subconjunto de porciones utilizadas para un análisis genético fetal.

Por lo tanto, sin quedar ligado a teoría alguna, en general, existe una relación de compromiso entre mayor precisión asociada con el enriquecimiento de la lectura fetal y menor precisión asociada con una cantidad reducida de datos de lectura (por ejemplo, eliminación de porciones y/o lecturas) para un análisis genético fetal. En algunos casos, un método comprende seleccionar un subconjunto de porciones enriquecidas para lecturas del ácido nucleico fetal (por ejemplo, lecturas fetales) que mejoran, o no disminuyen significativamente, la precisión de un análisis genético fetal. A pesar de este aparente compromiso, se ha determinado, como se describe en el presente documento, que utilizando un subconjunto de porciones, a las que se mapean una cantidad significativa de lecturas de fragmentos relativamente cortos, puede mejorar la precisión de los análisis genéticos fetales.

Un subconjunto de porciones, en las que una cantidad significativa de lecturas de los fragmentos CCF tienen una longitud menor que una longitud de un fragmento seleccionado, puede identificarse seleccionando porciones que cumplan uno o más criterios, y/o identificarse eliminando (por ejemplo, filtrando) porciones que no cumplan uno o más criterios. A veces, un subconjunto de porciones que se identifica después de asignar un conjunto de lecturas derivadas de una muestra, o en algunos casos, un conjunto de lecturas derivadas de múltiples muestras, se mapea en porciones de un genoma de referencia. En ciertos casos, la identificación de un subconjunto de porciones comprende mapear

las lecturas derivadas de fragmentos CCF en porciones en un genoma de referencia, cromosoma o segmento del mismo, y seleccionar un subconjunto de porciones de acuerdo con la cantidad de lecturas derivadas de fragmentos CCF relativamente cortos (por ejemplo, aproximadamente 200 pares de bases o menos) que se han mapeado en porciones individuales.

5 Se puede utilizar cualquier método adecuado para identificar porciones en las que se ha mapeado una cantidad significativa de lecturas de fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado (por ejemplo, una primera longitud de fragmento seleccionado). Los fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado a menudo son fragmentos CCF relativamente cortos, y algunas veces la longitud del fragmento seleccionado es de aproximadamente 200 pares de bases o menos (por ejemplo, fragmentos CCF que tienen aproximadamente 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90 u 80 bases de longitud). La longitud de un fragmento CCF se puede determinar (por ejemplo, deducir o inferir) mediante el mapeo de dos o más lecturas derivadas del fragmento (por ejemplo, una lectura de extremo pareado) a un genoma de referencia. Para lecturas de extremos pareados derivadas de un fragmento CCF, por ejemplo, las lecturas se pueden mapear en un genoma de referencia, se puede determinar la longitud de la secuencia genómica entre las lecturas mapeadas, y la totalidad de las dos longitudes de lectura y la longitud de la secuencia genómica entre las lecturas es igual a la longitud del fragmento CCF. La longitud de un molde de fragmento CCF a veces se determina directamente a partir de la longitud de una lectura derivada del fragmento (por ejemplo, lectura de extremo único).

20 Se pueden utilizar lecturas de secuencia de cualquier número adecuado de muestras para identificar un subconjunto de porciones que cumplan uno o más criterios descritos en este documento. A veces se utilizan lecturas de secuencia de un grupo de muestras de múltiples mujeres gestantes. Se pueden abordar una o más muestras de cada una de las múltiples mujeres gestantes (por ejemplo, de 1 a aproximadamente 20 muestras de cada mujer gestante (por ejemplo, aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 o 19 muestras)), y se puede abordar un número adecuado de mujeres gestantes (por ejemplo, de aproximadamente 2 a aproximadamente 10000 mujeres gestantes (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 mujeres gestantes)). En algunos casos, las lecturas de secuencia de la(s) misma(s) muestra(s) de ensayo de la misma mujer gestante, se mapean en porciones en el genoma de referencia y se utilizan para generar el subconjunto de porciones.

30 Después de que las lecturas se han mapeado en porciones, las porciones a veces se procesan antes de identificar un subconjunto de porciones que cumplen uno o más criterios descritos en este documento. En algunos casos, ciertas porciones se eliminan (por ejemplo, se filtran) de acuerdo con una o más características, cuyos ejemplos no limitantes incluyen mapeo relativamente bajo, variabilidad relativamente alta, fracción relativamente grande de secuencias repetitivas, uno o más procedimientos de normalización (por ejemplo, PERUN y/u otro procedimiento descrito en el presente documento), similares o una combinación de los mismos, y después se selecciona un subconjunto de porciones que cumple uno o más criterios descritos en el presente documento. En algunos casos, las porciones no se procesan antes de identificar un subconjunto de porciones que cumplen uno o más criterios descritos en el presente documento después de que las lecturas se hayan mapeado en las porciones.

40 En algunos casos, un subconjunto de porciones, en donde una cantidad significativa de lecturas de fragmentos CCF tienen una longitud menor que la longitud de un fragmento seleccionado, se identifica en función de si la cantidad de lecturas mapeadas de fragmentos CCF que tienen una longitud menor que la longitud de un primer fragmento seleccionado es mayor que la cantidad de lecturas mapeadas de los fragmentos CCF que tienen una longitud menor que la longitud de un segundo fragmento seleccionado. En ciertos casos, un subconjunto de porciones, en donde una cantidad significativa de lecturas de fragmentos CCF tienen una longitud menor que la longitud de un fragmento seleccionado, se identifica en función de si la cantidad de lecturas mapeadas de fragmentos CCF que tienen una longitud menor que la longitud de un primer fragmento seleccionado es mayor que la cantidad promedio, media o mediana de lecturas mapeadas de fragmentos CCF que tienen una longitud menor que la longitud de un segundo fragmento seleccionado para las porciones analizadas. En algunos casos, un subconjunto de porciones, en donde una cantidad significativa de lecturas de fragmentos CCF tienen una longitud menor que la longitud de un fragmento seleccionado, se identifica, basándose en la relación de longitud de fragmento (FLR) determinada para cada porción. Una "relación de longitud de fragmento" también se denomina en el presente documento una estadística de relación fetal (FRS).

55 En ciertos casos, una FLR se determina, en parte, de acuerdo con la cantidad de lecturas mapeadas en una porción de fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado. En algunos casos, un valor de FLR a menudo puede ser una relación de X a Y, donde X es la cantidad de lecturas derivadas de fragmentos CCF que tienen una longitud menor que la longitud del primer fragmento seleccionado, e Y es la cantidad de lecturas derivadas de fragmentos CCF que tienen una longitud menor que la longitud de un segundo fragmento seleccionado. A menudo, la longitud de un primer fragmento seleccionado se selecciona independientemente de la longitud de un segundo fragmento seleccionado, y viceversa, y la longitud del segundo fragmento seleccionado normalmente es mayor que la longitud del primer fragmento seleccionado. La longitud de un primer fragmento seleccionado puede ser de aproximadamente 200 bases o menos a aproximadamente 30 bases o menos. En algunos casos, una longitud de un primer fragmento seleccionado es de aproximadamente 200, 190, 180, 170, 160, 155, 150, 145, 140, 135, 130, 125, 120, 115, 110, 105, 100, 95, 90, 85, 80, 75, 70, 65, 60, 55 o 50 bases. En algunos casos, una

5 longitud de un primer fragmento seleccionado es de aproximadamente 170 a aproximadamente 130 bases, y algunas veces es de aproximadamente 160 a aproximadamente 140 bases. En algunos casos, la longitud de un segundo fragmento seleccionado es de aproximadamente 2000 bases a aproximadamente 200 bases. En ciertos casos, la longitud de un segundo fragmento seleccionado es de aproximadamente 1000, 950, 800, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250 bases. En algunos casos, la longitud del primer fragmento seleccionado es de aproximadamente 140 a aproximadamente 160 bases (por ejemplo, aproximadamente 150 bases) y la longitud del segundo fragmento seleccionado es de aproximadamente 500 a aproximadamente 700 bases (por ejemplo, aproximadamente 600 bases).

10 En algunos casos, una FLR es un promedio, media o mediana de múltiples valores de FLR. Por ejemplo, a veces, una FLR para una porción dada es un promedio, media o mediana de los valores de FLR para (i) dos o más muestras de ensayo, (ii) dos o más sujetos, o (iii) dos o más muestras de ensayo y dos o más sujetos. En ciertos casos, una FLR promedio, media o mediana deriva de los valores de FLR para dos o más porciones de un genoma, cromosoma o segmento de los mismos. En algunos casos, una FLR promedio, media o mediana está asociada con una incertidumbre (por ejemplo, desviación estándar, desviación absoluta mediana).

15 En algunos casos, se identifica un subconjunto de porciones de acuerdo con uno o más valores de FLR (por ejemplo, una comparación de uno o más valores de FLR). En ciertos casos, se identifica un subconjunto de porciones de acuerdo con una FLR y un umbral (por ejemplo, una comparación de una FLR y un umbral). En ciertos casos, una FLR promedio, media o mediana derivada de una porción dada se compara con una FLR promedio, media o mediana derivada de dos o más porciones de un genoma, cromosoma o segmento de los mismos. Por ejemplo, a veces, una FLR promedio para una porción dada se compara con una FLR mediana para una porción dada. En ciertos casos, una porción se identifica de acuerdo con una FLR promedio, media o mediana determinada para una porción y una FLR promedio, media o mediana determinada para una colección de porciones (por ejemplo, porción de un genoma, cromosoma o segmentos del mismo). En algunos casos, una FLR promedio para una porción puede estar por debajo de un cierto umbral determinado de acuerdo con una FLR mediana y la porción se retira del estudio (por ejemplo, en un análisis genético fetal). En algunos casos, una FLR promedio, media o mediana para una porción está por encima de un cierto umbral determinado de acuerdo con una FLR promedio, media o mediana para un genoma, cromosoma o segmento de los mismos, y la porción se selecciona y/o se agrega a un subconjunto de porciones para su estudio (por ejemplo, al determinar la presencia o ausencia de una variación genética). En algunos casos, una FLR para una porción es igual o mayor que aproximadamente 0,15 a aproximadamente 0,30 (por ejemplo, aproximadamente 0,16, 0,17, 0,18, 0,19, 0,20, 0,21, 0,21, 0,22, 0,23, 0,24, 0,25, 0,26, 0,27, 0,28, 0,29) y la porción se selecciona para su estudio (por ejemplo, se agrega o se incorpora a un subconjunto de porciones para un análisis genético fetal). En algunos casos, una FLR para una porción es igual o menor que aproximadamente 0,20 a aproximadamente 0,10 (por ejemplo, aproximadamente 0,19, 0,18, 0,17, 0,16, 0,15, 0,14, 0,13, 0,12, 0,11) y la porción se elimina del estudio (por ejemplo, se filtra).

20 Las porciones en un subconjunto a veces se identifican en función de, en parte, si una cantidad significativa de lecturas de fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado se mapean en una porción (por ejemplo, de acuerdo con una FLR). En algunos casos, se identifican las porciones de un subconjunto de acuerdo con una o más características o criterios, además de la cantidad de lecturas de secuencia mapeadas desde longitudes de fragmento menores que la longitud de un fragmento seleccionado. En algunos casos, se puede identificar un subconjunto de porciones en función de si una cantidad significativa de lecturas de fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado se mapean en una porción (por ejemplo, de acuerdo con una FLR) y una o más características adicionales. Ejemplos no limitantes de otras características incluyen el número de exones en, y/o el contenido de GC de, un genoma, cromosoma o segmento de los mismos, y/o una o más de las porciones. Por consiguiente, a veces las porciones identificadas en función de si una cantidad significativa de lecturas de fragmentos CCF que tienen una longitud menor que la longitud de un fragmento seleccionado se mapean en una porción (por ejemplo, de acuerdo con una FLR) para un subconjunto, se seleccionan o eliminan adicionalmente de acuerdo con el contenido de GC de la porción y/o el número de exones en la porción. En algunos casos, una porción no se selecciona o elimina del estudio (por ejemplo, filtrada) si el contenido de GC y/o el número de exones en la porción no se correlaciona con una FLR para la porción.

25 En algunos casos, un subconjunto de porciones consiste en, consistir esencialmente en, o comprende porciones que cumplan con uno o más criterios particulares descritos en el presente documento (por ejemplo, las porciones se caracterizan por un FLR igual a o mayor que un cierto valor). En ciertos casos, las porciones que no cumplen un criterio se incluyen en un subconjunto de porciones que cumplen el criterio, por ejemplo, para aumentar la precisión de un análisis genético fetal. En ciertos casos, en un subconjunto de porciones que "consiste esencialmente en" porciones seleccionadas de acuerdo con un criterio (por ejemplo, una FLR igual a o mayor que un cierto valor), aproximadamente el 90% o más (por ejemplo, aproximadamente el 91%, 91%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de las porciones cumple el criterio y aproximadamente el 10% o menos (por ejemplo, aproximadamente el 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, aproximadamente el 1% o menos) de las porciones no cumple el criterio.

30 Un subconjunto de porciones seleccionadas por los métodos descritos en el presente documento se puede utilizar para un análisis genético fetal de diferentes maneras. En ciertos casos, las lecturas derivadas de una muestra se utilizan en un proceso de mapeo utilizando un subconjunto preseleccionado de porciones descritas en el presente

documento, y no utilizando la totalidad o la mayoría de las porciones en un genoma de referencia. Aquellas lecturas que se mapean en el subconjunto de porciones preseleccionadas a menudo se utilizan en etapas adicionales de un análisis genético fetal, y las lecturas que no se mapean en el subconjunto de porciones preseleccionadas a menudo no se utilizan en las etapas adicionales de un análisis genético fetal (por ejemplo, se eliminan o filtran las lecturas que no se mapean).

A veces, las lecturas de secuencia derivadas de una muestra se mapean en todas o en la mayoría de las porciones de un genoma de referencia y se selecciona posteriormente un subconjunto preseleccionado de porciones descrito en el presente documento. Después, las lecturas del subconjunto seleccionado de porciones a menudo se utilizan en etapas adicionales de un análisis genético fetal. En los últimos casos, las lecturas de porciones no seleccionadas a menudo no se utilizan en etapas adicionales de un análisis genético fetal (por ejemplo, se eliminan o filtran las lecturas en las porciones no seleccionadas).

Recuentos

Las lecturas de secuencia que se mapean o particionan basándose en una característica o variable seleccionada se pueden cuantificar para determinar el número de lecturas que se mapean en una o más porciones (por ejemplo, porción de un genoma de referencia). En ciertos casos, la cantidad de lecturas de secuencia que se mapean a una porción se denominan recuentos (por ejemplo, un recuento). A menudo un recuento se asocia con una porción. En ciertos casos, los recuentos de dos o más porciones (por ejemplo, un conjunto de porciones) se manipulan matemáticamente (por ejemplo, promediar, sumar, normalizar, similares o una combinación de los mismos). En algunos casos, se determina un recuento a partir de algunas o todas las lecturas de secuencia mapeadas en (es decir, asociadas con) una porción. En ciertos casos, se determina un recuento a partir de un subconjunto predefinido de lecturas de secuencias mapeadas. Los subconjuntos predefinidos de lecturas de secuencias mapeadas se pueden definir o seleccionar utilizando cualquier característica o variable adecuada. En algunos casos, los subconjuntos predefinidos de lecturas de secuencias mapeadas pueden incluir de 1 a n lecturas de secuencia, donde n representa un número igual a la suma de todas las lecturas de secuencia generadas a partir de un sujeto de ensayo o una muestra de un sujeto de referencia.

En ciertos casos, un recuento deriva de lecturas de secuencia que se procesan o manipulan mediante un método, operación o proceso matemático adecuado conocido en la técnica. Un recuento (por ejemplo, recuentos) puede determinarse mediante un método, operación o proceso matemático adecuado. En ciertos casos, un recuento deriva de lecturas de secuencia asociadas con una porción en donde algunas o todas las lecturas de secuencia se ponderan, eliminan, filtran, normalizan, ajustan, promedian, derivan como una media, suman o restan o procesan mediante una combinación de los mismos. En ciertos casos, un recuento deriva de lecturas de secuencia sin procesar y lecturas de secuencia filtradas. En ciertos casos, un valor de recuento se determina por un proceso matemático. En ciertos casos, un valor de recuento es un promedio, media o suma de lecturas de secuencia mapeadas en una porción. A menudo un recuento es un número medio de recuentos. En algunos casos, un recuento se asocia con un valor de incertidumbre.

En algunos casos, los recuentos se pueden manipular o transformar (por ejemplo, normalizar, combinar, sumar, filtrar, seleccionar, promediar, derivar como una media, similares o una combinación de los mismos). En algunos casos, los recuentos se pueden transformar para producir recuentos normalizados. Los recuentos se pueden procesar (por ejemplo, normalizar) mediante un método conocido en la técnica y/o como se describe en el presente documento (por ejemplo, normalización en porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS de GC, LOWESS, PERUN, RM, GCRM, cQn y/o combinaciones de las mismas).

Los recuentos (por ejemplo, recuentos sin procesar, filtrados y/o normalizados) se pueden procesar y normalizar a uno o más niveles. Los niveles y perfiles se describen con mayor detalle a continuación. En ciertos casos, los recuentos se pueden procesar y/o normalizar a un nivel de referencia. Los niveles de referencia se abordan más adelante en el presente documento. Los recuentos procesados de acuerdo con un nivel (por ejemplo, recuentos procesados) pueden asociarse con un valor de incertidumbre (por ejemplo, una varianza calculada, un error, desviación estándar, puntuación Z, un valor de p, desviación media absoluta, etc.). En ciertos casos, un valor de incertidumbre define un intervalo por encima y por debajo de un nivel. Se puede usar un valor de desviación en lugar de un valor de incertidumbre, y los ejemplos no limitantes de medidas de desviación incluyen desviación estándar, desviación absoluta promedio, desviación absoluta media, puntuación estándar (por ejemplo, puntuación Z, puntuación Z, puntuación normal, variable estandarizada) y similares.

Los recuentos a menudo se obtienen de una muestra de ácido nucleico de una gestante que tiene un feto. Los recuentos de lecturas de secuencias de ácidos nucleicos mapeadas en una o más porciones a menudo son recuentos representativos tanto del feto como de la madre del feto (por ejemplo, una mujer gestante). En ciertos casos, algunos de los recuentos mapeados en una porción son de un genoma fetal y algunos de los recuentos mapeados en la misma porción son de un genoma materno.

Procesamiento y Normalización de Datos

Las lecturas de secuencia mapeadas que se han contabilizado se denominan en el presente documento, datos sin

procesar, ya que los datos representan recuentos no manipulados (por ejemplo, recuentos sin procesar). En algunos casos, los datos de lectura de secuencia en un conjunto de datos pueden procesarse adicionalmente (por ejemplo, manipularse matemáticamente y/o estadísticamente) y/o mostrarse para facilitar proporcionar el resultado. En ciertos casos, los conjuntos de datos, incluidos los conjuntos de datos más grandes, pueden beneficiarse del procesamiento
 5 previo para facilitar un análisis adicional. El procesamiento previo de conjuntos de datos a veces implica la eliminación de porciones redundantes y/o porciones no informativas de un genoma de referencia (p. ej., porciones de un genoma de referencia con datos no informativos, lecturas mapeadas redundantes, porciones con recuentos de mediana cero, secuencias sobrerrepresentadas o representadas de forma insuficiente). Sin quedar ligado a teoría alguna, el procesamiento y/o el procesamiento previo de datos pueden (i) eliminar datos ruidosos, (ii) eliminar datos no
 10 informativos, (iii) eliminar datos redundantes, (iv) reducir la complejidad de conjuntos de datos más grandes y/o (v) facilitar la transformación de los datos de una forma en una o más formas diferentes. Las expresiones "procesamiento previo" y "procesamiento" cuando se utilizan con respecto a datos o conjuntos de datos se denominan colectivamente en el presente documento "procesamiento". En algunos casos, el procesamiento puede hacer que los datos sean más fáciles de analizar y pueden generar un resultado. En algunos casos, uno o más o la totalidad de los métodos de
 15 procesamiento (por ejemplo, métodos de normalización, filtrado de porciones, mapeo, validación, similares o combinaciones de los mismos) se realizan mediante un procesador, un microprocesador, un ordenador, junto con la memoria y/o mediante una máquina controlada por microprocesador.

La expresión "datos ruidosos", como se usa en el presente documento, se refiere a (a) datos que tienen una varianza significativa entre los puntos de datos cuando se analizan o representan gráficamente, (b) datos que tienen una
 20 desviación estándar significativa (por ejemplo, mayor que 3 desviaciones estándar), (c) datos que tienen un error estándar de la media significativo, similares, y combinaciones de los anteriores. Los datos ruidosos a veces se producen debido a la cantidad y/o calidad del material de partida (por ejemplo, muestra de ácidos nucleicos), y a veces se producen como parte de los procesos para preparar o replicar el ADN utilizado para generar lecturas de secuencia.
 25 En ciertos casos, el ruido puede ser el resultado de ciertas secuencias que están sobrerrepresentadas cuando se preparan utilizando métodos basados en PCR. Los métodos descritos en el presente documento pueden reducir o eliminar la contribución de datos ruidosos y, por lo tanto, reducir el efecto de los datos ruidosos en el resultado proporcionado.

Las expresiones "datos no informativos", "porciones no informativas de un genoma de referencia" y "porciones no informativas", como se usan en el presente documento, se refieren a porciones, o datos derivados de ellas, que tienen un valor numérico que es significativamente diferente de un valor umbral predeterminado o se encuentra fuera del
 30 intervalo de valores de corte predeterminado. En el presente documento, las expresiones "umbral" y "valor umbral" se refieren a cualquier número que se calcula utilizando un conjunto de datos calificados y sirve como un límite de diagnóstico de una variación genética (por ejemplo, una variación en el número de copias, una aneuploidía, una aberración cromosómica y similares). En ciertos casos, un umbral se excede por los resultados obtenidos por los métodos descritos en el presente documento y a un sujeto se le diagnostica una variación genética (por ejemplo, trisomía 21). Un valor o intervalo de valores umbral a menudo se calcula manipulando matemática y/o estadísticamente
 35 los datos de lectura de secuencia (por ejemplo, de una referencia y/o sujeto), y los datos de lectura de secuencia se manipulan para generar un valor o intervalo de valores umbral en datos de lectura de secuencia (por ejemplo, de una referencia y/o sujeto). En algunos casos, se determina un valor de incertidumbre. Un valor de incertidumbre generalmente es una medida de varianza o error y puede ser cualquier medida adecuada de varianza o error. En algunos casos, un valor de incertidumbre es una desviación estándar, un error estándar, una varianza calculada, un valor p o una desviación media absoluta (MAD). En algunos casos, un valor de incertidumbre se puede calcular de acuerdo con una fórmula en el Ejemplo 4.
 40
 45

Se puede utilizar cualquier procedimiento adecuado para procesar conjuntos de datos descritos en el presente documento. Ejemplos no limitantes de procedimientos adecuados para su uso para el procesamiento de conjuntos de
 50 datos incluyen filtrado, normalización, ponderación, monitorización de alturas de máximos, monitorización de áreas de máximos, monitorización de bordes de máximos, determinación de relaciones de área, procesamiento matemático de datos, procesamiento estadístico de datos, aplicación de algoritmos estadísticos, análisis con variables fijas, análisis con variables optimizadas, representación gráfica de datos para identificar patrones o tendencias para procesamiento adicional, similares y combinaciones de los anteriores. En algunos casos, los conjuntos de datos se procesan en función de varias características (por ejemplo, en el contenido de GC, lecturas mapeadas redundantes, regiones de
 55 centrómeros, regiones de telómeros, similares y combinaciones de los mismos) y/o variables (por ejemplo, sexo fetal, edad materna, ploidía materna, porcentaje de contribución de ácido nucleico fetal, similares o combinaciones de los mismos). En ciertos casos, el procesamiento de conjuntos de datos, como se describe en el presente documento, puede reducir la complejidad y/o dimensionalidad de conjuntos de datos grandes y/o complejos. Un ejemplo no limitante de un conjunto de datos complejos incluye datos de lectura de secuencia generados a partir de uno o más sujetos de ensayo y una pluralidad de sujetos de referencia de diferentes edades y orígenes étnicos. En algunos casos, los conjuntos de datos pueden incluir desde miles hasta millones de lecturas de secuencia para cada sujeto de ensayo y/o de referencia.
 60

En ciertos casos, el procesamiento de datos se puede realizar en cualquier número de etapas. Por ejemplo, en algunos
 65 casos, los datos pueden procesarse utilizando un solo procedimiento de procesamiento, y en ciertos casos, los datos pueden procesarse utilizando 1 o más, 5 o más, 10 o más o 20 o más etapas de procesamiento (por ejemplo, 1 o más

etapas de procesamiento, 2 o más etapas de procesamiento, 3 o más etapas de procesamiento, 4 o más etapas de procesamiento, 5 o más etapas de procesamiento, 6 o más etapas de procesamiento, 7 o más etapas de procesamiento, 8 o más etapas de procesamiento, 9 o más etapas de procesamiento, 10 o más etapas de procesamiento, 11 o más etapas de procesamiento, 12 o más etapas de procesamiento, 13 o más etapas de procesamiento, 14 o más etapas de procesamiento, 15 o más etapas de procesamiento, 16 o más etapas de procesamiento, 17 o más etapas de procesamiento, 18 o más etapas de procesamiento, 19 o más etapas de procesamiento, 20 o más etapas de procesamiento). En algunos casos, las etapas de procesamiento pueden ser la misma etapa repetida dos o más veces (por ejemplo, filtrado dos o más veces, normalización dos o más veces), y en ciertos casos, las etapas de procesamiento pueden ser dos o más etapas de procesamiento diferentes (por ejemplo, filtrado, normalización; normalización, monitorización de alturas de máximos; filtrado, normalización, normalización de una referencia, manipulación estadística para determinar los valores de p, y similares), realizados de forma simultánea o secuencial. En algunos casos, se puede utilizar cualquier número y/o combinación adecuada de las mismas o diferentes etapas de procesamiento para procesar los datos de lecturas de secuencia para facilitar proporcionar el resultado. En ciertos casos, el procesamiento de conjuntos de datos mediante los criterios descritos en el presente documento puede reducir la complejidad y/o dimensionalidad de conjuntos de datos.

En algunos casos, una o más etapas de procesamiento pueden comprender una o más etapas de filtrado. El término "filtrado", como se usa en el presente documento, se refiere a eliminar porciones o porciones de un genoma de referencia para su estudio. Se pueden seleccionar porciones de un genoma de referencia para su eliminación en función de cualquier criterio adecuado, incluyendo, pero sin limitación, datos redundantes (p. ej., lecturas mapeadas redundantes o solapadas), datos no informativos (p. ej., porciones de un genoma de referencia con recuentos de mediana cero), porciones de un genoma de referencia con secuencias sobrerrepresentadas o representadas de manera insuficiente, datos ruidosos, similares, o combinaciones de los anteriores. Un proceso de filtrado a menudo implica eliminar una o más porciones de un genoma de referencia del estudio y restar los recuentos en una o más porciones de un genoma de referencia seleccionado para eliminar de los recuentos contados o recuentos sumados de las porciones de un genoma de referencia, cromosoma o cromosomas o genoma en estudio. En algunos casos, las porciones de un genoma de referencia se pueden eliminar sucesivamente (por ejemplo, una a la vez para permitir la evaluación del efecto de la eliminación de cada porción individual), y en ciertos casos, se pueden eliminar al mismo tiempo todas las porciones de un genoma de referencia marcadas para su eliminación. En algunos casos, las porciones de un genoma de referencia caracterizadas por una varianza por encima o por debajo de un cierto nivel se eliminan, lo que a veces se denomina en el presente documento filtrado de porciones "ruidosas" de un genoma de referencia. En ciertos casos, un proceso de filtrado comprende obtener puntos de datos de un conjunto de datos que se desvían del nivel de perfil medio de una porción, un cromosoma o segmento de un cromosoma en un múltiplo predeterminado de la varianza del perfil, y en ciertos casos, un proceso de filtrado comprende eliminar puntos de datos, un conjunto de datos, que no se desvían del nivel de perfil medio de una porción, un cromosoma o segmento de un cromosoma por un múltiplo predeterminado de la varianza del perfil. En algunos casos, se utiliza un proceso de filtrado para reducir el número de porciones candidatas de un genoma de referencia analizado por la presencia o ausencia de una variación genética. La reducción del número de porciones candidatas de un genoma de referencia analizado por la presencia o ausencia de una variación genética (p. ej., microdelección, microduplicación) a menudo reduce la complejidad y/o dimensionalidad de un conjunto de datos, y algunas veces aumenta la velocidad de búsqueda y/o identificación de variaciones genéticas y/o aberraciones genéticas en dos o más órdenes de magnitud.

En algunos casos, una o más etapas de procesamiento pueden comprender una o más etapas de normalización. La normalización se puede realizar mediante un método adecuado descrito en el presente documento o conocido en la técnica. En ciertos casos, la normalización comprende ajustar los valores medidos en diferentes escalas a una escala teóricamente común. En ciertos casos, la normalización comprende un ajuste matemático sofisticado para alinear las distribuciones de probabilidad de los valores ajustados. En ciertos casos, la normalización comprende alinear distribuciones a una distribución normal. En ciertos casos, la normalización comprende ajustes matemáticos que permiten la comparación de los valores normalizados correspondientes para diferentes conjuntos de datos de una manera que elimina los efectos de ciertas influencias generales (por ejemplo, errores y anomalías). En ciertos casos, la normalización comprende escalado. La normalización a veces comprende la división de uno o más conjuntos de datos por una variable o fórmula predeterminada. Ejemplos no limitantes de métodos de normalización incluyen normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS (atenuado de regresión local ponderada), PERUN, repetición de enmascaramiento (RM), normalización GC y repetición de enmascaramiento (GCRM), normalización por cuantiles condicionales (cQn) y/o combinaciones de los mismos. La determinación de una presencia o ausencia de una variación genética (por ejemplo, una aneuploidía) puede utilizar un método de normalización (por ejemplo, una normalización en porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS (atenuado de regresión local ponderada), PERUN, repetición de enmascaramiento (RM), normalización GC y repetición de enmascaramiento (GCRM), cQn, un método de normalización conocido en la técnica y/o una combinación de los mismos).

Por ejemplo, LOESS es un método de modelado de regresión conocido en la técnica que combina múltiples modelos de regresión en un metamodelo basado en los k vecinos más cercanos. LOESS se refiere a veces como una regresión polinomial ponderada localmente. En algunos casos, LOESS de GC, aplica un modelo LOESS a la relación entre el recuento de fragmentos (por ejemplo, lecturas de secuencia, recuentos) y la composición en GC para porciones de un

genoma de referencia. La representación gráfica de una curva suave a través de un conjunto de puntos de datos utilizando LOESS a veces se denomina curva LOESS, particularmente cuando cada valor atenuado viene dado por una regresión cuadrática ponderada de mínimos cuadrados sobre el intervalo de valores de la variable de criterio del diagrama de dispersión del eje y. Para cada punto de un conjunto de datos, el método LOESS ajusta un polinomio de bajo grado a un subconjunto de los datos, con valores de variables explicativas cerca del punto cuya respuesta se está estimando. El polinomio se ajusta utilizando mínimos cuadrados ponderados, lo que da más peso a los puntos cercanos al punto cuya respuesta se está estimando y menos a los puntos más alejados. El valor de la función de regresión para un punto se obtiene luego evaluando el polinomio local utilizando los valores de las variables explicativas para ese punto de datos. El ajuste LOESS a veces se considera completo después de que se hayan calculado los valores de la función de regresión para cada uno de los puntos de datos. Muchos de los detalles de este método, como el grado del modelo polinomial y las ponderaciones, son flexibles.

Se puede utilizar cualquier número adecuado de normalizaciones. En algunos casos, los conjuntos de datos pueden normalizarse 1 o más, 5 o más, 10 o más o incluso 20 o más veces. Los conjuntos de datos pueden normalizarse a valores (por ejemplo, valor de normalización) representativos de cualquier característica o variable adecuada (por ejemplo, datos de muestra, datos de referencia, o ambos). Ejemplos no limitantes de los tipos de normalización de datos que se pueden usar incluyen la normalización de los datos de recuento sin procesar para una o más porciones de ensayo o referencia seleccionadas al número total de recuentos mapeados en el cromosoma o en el genoma completo en donde se mapean la porción o las secciones seleccionadas; normalización de los datos de recuento sin procesar para una o más porciones seleccionadas a un recuento de referencia mediana para una o más porciones o el cromosoma en donde se mapea una porción o segmento seleccionados; normalización de los datos de recuento sin procesar con datos previamente normalizados o sus derivados; y normalización previamente a la normalización de datos a una o más variables de normalización predeterminadas. La normalización de un conjunto de datos a veces tiene el efecto de aislar el error estadístico, dependiendo de la característica o propiedad seleccionada como la variable de normalización predeterminada. La normalización de un conjunto de datos a veces también permite la comparación de las características de datos que tienen diferentes escalas, al llevar los datos a una escala común (por ejemplo, una variable de normalización predeterminada). En algunos casos, se pueden utilizar una o más normalizaciones a un valor estadísticamente derivado para minimizar las diferencias de datos y disminuir la importancia de los datos externos. La normalización de porciones, o porciones de un genoma de referencia, con respecto a un valor de normalización a veces se denominan "normalización de porciones".

En ciertos casos, una etapa de procesamiento que comprende la normalización incluye la normalización a una ventana estática, y en algunos casos, una etapa de procesamiento que comprende la normalización incluye la normalización a una ventana móvil o deslizante. El término "ventana", como se usa en el presente documento, se refiere a una o más porciones elegidas para el análisis, y algunas veces se usa como referencia para la comparación (por ejemplo, se usa para la normalización y/u otra manipulación matemática o estadística). El término "normalizar a una ventana estática" como se usa en el presente documento se refiere a un proceso de normalización que usa una o más porciones seleccionadas para comparación entre el conjunto de datos de un sujeto de ensayo y del sujeto de referencia. En algunos casos, las porciones seleccionadas se utilizan para generar un perfil. Una ventana estática generalmente incluye un conjunto predeterminado de porciones que no cambian durante las manipulaciones y/o análisis. Las expresiones "normalizar a una ventana móvil" y "normalizar a una ventana deslizante" como se usan en el presente documento se refieren a normalizaciones realizadas a porciones localizadas en la región genómica (por ejemplo, entorno genético inmediato, porción o secciones adyacentes, y similares) de una porción de ensayo seleccionada, donde una o más porciones de ensayo seleccionadas se normalizan a porciones que rodean inmediatamente la porción de ensayo seleccionada. En ciertos casos, las porciones seleccionadas se utilizan para generar un perfil. Una normalización de ventana deslizante o móvil a menudo incluye moverse o deslizarse repetidamente a una porción de ensayo adyacente, y normalizar la porción de ensayo recién seleccionada a las porciones que rodean inmediatamente o adyacentes a la porción de ensayo recién seleccionada, donde las ventanas adyacentes tienen una o más porciones en común. En ciertos casos, una pluralidad de porciones de ensayo y/o cromosomas seleccionados pueden analizarse mediante un proceso de ventana deslizante.

En algunos casos, la normalización a una ventana deslizante o móvil puede generar uno o más valores, donde cada valor representa la normalización a un conjunto diferente de porciones de referencia seleccionadas de diferentes regiones de un genoma (por ejemplo, un cromosoma). En ciertos casos, el uno o más valores generados son sumas acumulativas (por ejemplo, una estimación numérica del entero del perfil de recuento normalizado sobre la porción, dominio (por ejemplo, parte del cromosoma) o cromosoma seleccionados). Los valores generados por el proceso de ventana deslizante o móvil se pueden utilizar para generar un perfil y facilitar la obtención de un resultado. En algunos casos, las sumas acumulativas de una o más porciones se pueden mostrar en función de la posición genómica. El análisis de ventanas móviles o deslizantes a veces se utiliza para analizar un genoma en busca de la presencia o ausencia de microdeleciones y/o microinserciones. En ciertos casos, la visualización de sumas acumulativas de una o más porciones se usa para identificar la presencia o ausencia de regiones de variación genética (por ejemplo, microdeleciones, microduplicaciones). En algunos casos, el análisis de ventana móvil o deslizante se usa para identificar regiones genómicas que contienen microdeleciones y en ciertos casos el análisis de ventana móvil o deslizante se usa para identificar regiones genómicas que contienen microduplicaciones.

Una metodología de normalización particularmente útil para reducir el error asociado con los indicadores de ácidos

nucleicos se denomina en el presente documento Eliminación de Error Parametrizado y Normalización No-Sesgada (PERUN) descrita en el presente documento y, por ejemplo, en la Solicitud de Patente de Estados Unidos número 13/669.136 y en la solicitud de patente internacional n.º PCT/US12/59123 (documento WO2013/052913). La metodología PERUN se puede aplicar a una variedad de indicadores de ácidos nucleicos (por ejemplo, lecturas de

5 secuencias de ácidos nucleicos) con el propósito de reducir los efectos de error que confunden las predicciones basadas en dichos indicadores.

Por ejemplo, la metodología PERUN se puede aplicar a lecturas de secuencias de ácidos nucleicos de una muestra y reducir los efectos de error que pueden perjudicar las determinaciones de nivel de sección genómica. Dicha aplicación es útil para usar lecturas de secuencias de ácidos nucleicos para determinar la presencia o ausencia de una variación genética en un sujeto que se manifiesta como un nivel variable de una secuencia de nucleótidos (por ejemplo, una porción, un nivel de sección genómica). Ejemplos no limitantes de variaciones en las porciones son las aneuploidías cromosómicas (por ejemplo, trisomía 21, trisomía 18, trisomía 13) y la presencia o ausencia de un cromosoma sexual (por ejemplo, XX en mujeres frente a XY en hombres). Una trisomía de un autosoma (por ejemplo, un cromosoma que no sea un cromosoma sexual) puede denominarse un autosoma afectado. Otros ejemplos no limitantes de variaciones en los niveles de sección genómica incluyen microdeleciones, microinserciones, duplicaciones y mosaicismo.

10

15

En determinadas aplicaciones, la metodología PERUN puede reducir el sesgo experimental al normalizar los indicadores de ácidos nucleicos para grupos genómicos particulares, los últimos de los cuales se denominan porciones. Las porciones incluyen una colección adecuada de indicadores de ácidos nucleicos, un ejemplo no limitante de las cuales, incluye una longitud de nucleótidos contiguos, que se denomina en el presente documento una sección genómica o porción de un genoma de referencia. Las ventanas pueden incluir otros indicadores de ácidos nucleicos como se describe en el presente documento. En tales aplicaciones, la metodología PERUN generalmente normaliza los indicadores de ácidos nucleicos en ventanas particulares en una serie de muestras en tres dimensiones.

20

25

En determinadas aplicaciones, la metodología PERUN puede reducir el sesgo experimental y/o sistemático al normalizar los indicadores de ácidos nucleicos (por ejemplo, recuentos, lecturas) mapeados en segmentos particulares (por ejemplo, porciones) de un genoma de referencia. En tales aplicaciones, la metodología PERUN generalmente normaliza los recuentos de lecturas de ácidos nucleicos en porciones particulares de un genoma de referencia en una serie de muestras en tres dimensiones. Se proporciona una descripción detallada de PERUN y sus aplicaciones en la sección de Ejemplos en el presente documento, en la solicitud de patente internacional n.º PCT/US12/59123 (WO2013/052913) y la publicación de solicitud de patente de EE.UU. n.º US20130085681.

30

En ciertos casos, la metodología PERUN incluye el cálculo de un nivel de sección genómica para porciones de un genoma de referencia a partir de (a) recuentos de lectura de secuencias mapeadas en una porción de un genoma de referencia para una muestra de ensayo, (b) sesgo experimental (por ejemplo, sesgo de GC) para la muestra de ensayo y (c) uno o más parámetros de ajuste (por ejemplo, estimaciones de ajuste) para una relación ajustada entre (i) sesgo experimental para una porción de un genoma de referencia para el cual se mapean las lecturas de secuencias y (ii) los recuentos de lecturas de secuencias se mapean en la porción. El sesgo experimental para cada una de las porciones de un genoma de referencia se puede determinar a través de múltiples muestras de acuerdo con una relación ajustada para cada muestra entre (i) los recuentos de lecturas de secuencias mapeadas en cada una de las porciones de un genoma de referencia, y (ii) una característica de mapeo para cada una de las porciones de un genoma de referencia. Esta relación ajustada para cada muestra se puede ensamblar para múltiples muestras en tres dimensiones. En ciertos casos, el ensamblaje se puede ordenar de acuerdo con el sesgo experimental, aunque la metodología PERUN se puede practicar sin ordenar el ensamblaje de acuerdo con el sesgo experimental. La relación ajustada para cada muestra y la relación ajustada para cada porción del genoma de referencia pueden ajustarse independientemente a una función lineal o a una función no lineal mediante un proceso de ajuste adecuado conocido en la técnica.

35

40

45

En algunos casos, una relación es una relación geométrica y/o gráfica. Los términos "relación" y "correlación", como se usa en el presente documento, son sinónimos. En algunos casos, una relación es una relación matemática. En algunos casos, una relación se representa gráficamente. En algunos casos, una relación es una relación lineal. En ciertos casos, una relación es una relación no lineal. En ciertos casos, una relación es una regresión (por ejemplo, una línea de regresión). Una regresión puede ser una regresión lineal o una regresión no lineal. Una relación puede expresarse por una ecuación matemática. A menudo, una relación se define, en parte, por una o más constantes. Se puede generar una relación por un método conocido en la técnica. En ciertos casos, se puede generar una relación en dos dimensiones para una o más muestras, y se puede seleccionar una probabilidad de error variable, o posiblemente se puede seleccionar una probabilidad de error, para una o más de las dimensiones. Se puede generar una relación, por ejemplo, utilizando un software de gráficos conocido en la técnica que representa un gráfico utilizando valores de dos o más variables proporcionadas por un usuario. Se puede ajustar una relación utilizando un método conocido en la técnica (por ejemplo, software gráfico). Ciertas relaciones pueden ajustarse mediante regresión lineal, y la regresión lineal puede generar un valor de pendiente y un valor de intersección. Ciertas relaciones a veces no son lineales y pueden ajustarse mediante una función no lineal, como una función parabólica, hiperbólica o exponencial (por ejemplo, una función cuadrática), por ejemplo.

50

55

60

En la metodología PERUN, una o más de las relaciones ajustadas pueden ser lineales. Para un análisis de ácido

65

nucleico circulante libre de células de gestantes, donde el sesgo experimental es un sesgo de GC y la característica de mapeo es el contenido de GC, una relación ajustada para una muestra entre (i) los recuentos de lecturas de secuencias mapeadas en cada porción, y (ii) el contenido de GC para cada una de las porciones de un genoma de referencia, puede ser lineal. Para la última relación ajustada, la pendiente corresponde al sesgo de GC, y se puede determinar un coeficiente de sesgo de GC para cada muestra cuando las relaciones ajustadas se ensamblan a través de múltiples muestras. En tales casos, la relación ajustada para múltiples muestras y una porción entre (i) el coeficiente de sesgo de GC para la porción, y (ii) los recuentos de lecturas de secuencias mapeadas en la porción, también puede ser lineal. Se puede obtener una intersección y una pendiente a partir de esta última relación ajustada. En tales aplicaciones, la pendiente aborda el sesgo específico de la muestra en función del contenido de GC y la intersección aborda un patrón de atenuación específico de porción común a todas las muestras. La metodología PERUN puede reducir significativamente dicho sesgo específico de la muestra y la atenuación específica de porción al calcular los niveles de la sección genómica para proporcionar un resultado (por ejemplo, presencia o ausencia de variación genética; determinación del sexo fetal).

En algunos casos, la normalización PERUN hace uso del ajuste a una función lineal y se describe mediante la Ecuación A, la Ecuación B o una derivación de la misma.

Ecuación A:

$$M = LI + GS \quad (A)$$

Ecuación B:

$$L = (M - GS)/I \quad (B)$$

En algunos casos, L es un nivel o perfil normalizado por PERUN. En algunos casos, L es la salida deseada del procedimiento de normalización PERUN. En ciertos casos, L es específico de porción. En algunos casos L se determina según múltiples porciones de un genoma de referencia y representa un nivel normalizado por PERUN de un genoma, cromosoma, porciones o segmento de los mismos. El nivel L se usa a menudo para análisis adicionales (por ejemplo, para determinar los valores Z , deleciones/duplicaciones maternas, microdeleciones/microduplicaciones fetales, sexo fetal, aneuploidías sexuales, etc.). El método de normalización de acuerdo con la Ecuación B se denomina Eliminación de Error Parametrizado y Normalización No-Sesgada (PERUN).

En algunos casos, G es un coeficiente de sesgo de GC medido utilizando un modelo lineal, LOESS, o cualquier aproximación equivalente. En algunos casos, G es una pendiente. En algunos casos, el coeficiente G de sesgo de GC se evalúa como la pendiente de la regresión para los recuentos M (por ejemplo, recuentos sin procesar) para la porción i y el contenido de GC de la porción i determinado a partir de un genoma de referencia. En algunos casos, G representa información secundaria, extraída de M y determinada de acuerdo con una relación. En algunos casos, G representa una relación para un conjunto de recuentos específicos de porción y un conjunto de valores de contenido de GC específicos de porción para una muestra (por ejemplo, una muestra de ensayo). En algunos casos, el contenido de GC específico de porción deriva de un genoma de referencia. En algunos casos, el contenido de GC específico de porción deriva del contenido de GC observado o medido (por ejemplo, medido a partir de la muestra). Un coeficiente de sesgo de GC a menudo se determina para cada muestra en un grupo de muestras y generalmente se determina para una muestra de ensayo. Un coeficiente de sesgo de GC a menudo es específico de la muestra. En algunos casos, un coeficiente de sesgo de GC es una constante. En ciertos casos, un coeficiente de sesgo de GC, una vez derivado de una muestra, no cambia.

En algunos casos, I es una intersección y S es una pendiente derivada de una relación lineal. En algunos casos, la relación de la que derivan I y S es diferente de la relación de la que deriva G . En algunos casos, la relación de la que derivan I y S se fija para una configuración experimental dada. En algunos casos, I y S derivan de una relación lineal de acuerdo con los recuentos (por ejemplo, recuentos sin procesar) y un coeficiente de sesgo de GC de acuerdo con múltiples muestras. En algunos casos, I y S derivan independientemente de la muestra de ensayo. En algunos casos, I y S derivan de múltiples muestras. I y S a menudo son específicas de porción. En algunos casos, I y S se determinan suponiendo que $L = 1$ para todas las porciones de un genoma de referencia en muestras euploides. En algunos casos, se determina una relación lineal para las muestras euploides y se determinan los valores de I y S específicos para una porción seleccionada (suponiendo que $L = 1$). En ciertos casos, se aplica el mismo procedimiento a todas las porciones de un genoma de referencia en un genoma humano y se determina un conjunto de intersecciones I y pendientes S para cada porción.

En algunos casos, se aplica una aproximación de validación cruzada. La validación cruzada, a veces se denomina estimación de rotación. En algunos casos, se aplica una aproximación de validación cruzada para evaluar la precisión con la que se llevará a cabo un modelo predictivo (por ejemplo, como PERUN) en la práctica utilizando una muestra de ensayo. En algunos casos, una ronda de validación cruzada comprende hacer la partición de una muestra de datos en subconjuntos complementarios, realizar un análisis de validación cruzada en un subconjunto (por ejemplo, a veces denominado conjunto de entrenamiento) y validar el análisis utilizando otro subconjunto (por ejemplo, a veces llamado un conjunto de validación o conjunto de ensayo). En ciertos casos, se realizan múltiples rondas de validación cruzada

utilizando diferentes particiones y/o subconjuntos diferentes). Ejemplos no limitantes de las aproximaciones de validación cruzada incluyen dejando-uno-fuera (*leave-one-out*), bordes deslizantes, k iteraciones, de 2 iteraciones, repetición de submuestreo aleatorio, similares o combinaciones de los mismos. En algunos casos, una validación cruzada selecciona aleatoriamente un conjunto de trabajo que contiene el 90% de un conjunto de muestras que comprende fetos euploides conocidos y utiliza ese subconjunto para entrenar un modelo. En ciertos casos, la selección aleatoria se repite 100 veces, produciendo un conjunto de 100 pendientes y 100 intersecciones por cada porción.

En algunos casos, el valor de M es un valor medido derivado de una muestra de ensayo. En algunos casos, M es recuentos sin procesar medidos para una porción. En algunos casos, cuando los valores I y S están disponibles para una porción, la medición M se determina a partir de una muestra de ensayo y se usa para determinar el nivel L normalizado con PERUN para un genoma, cromosoma, segmento o porción del mismo según la Ecuación B.

Por lo tanto, la aplicación de la metodología PERUN a las lecturas de secuencias en múltiples muestras en paralelo puede reducir significativamente el error causado por (i) el sesgo experimental específico de la muestra (por ejemplo, el sesgo de GC) y (ii) la atenuación específica de porción común a las muestras. Otros métodos en los que cada una de estas dos fuentes de error se abordan por separado o en serie a menudo no pueden reducirlos tan eficazmente como la metodología PERUN. Sin quedar ligado a teoría alguna, se espera que la metodología PERUN reduzca el error más eficazmente en parte porque sus procesos generalmente aditivos no magnifican la propagación tanto como los procesos generalmente multiplicativos utilizados en otras aproximaciones de normalización (por ejemplo, GC-LOESS).

Se pueden utilizar técnicas de normalización y estadísticas adicionales en combinación con la metodología PERUN. Se puede aplicar un proceso adicional antes, después y/o durante el empleo de la metodología PERUN. Se describen ejemplos no limitantes de procesos que pueden usarse en combinación con la metodología PERUN en el presente documento a continuación.

En algunos casos, se puede utilizar una normalización o ajuste secundario de un nivel de sección genómica para el contenido de GC junto con la metodología PERUN. Se puede utilizar un procedimiento adecuado de ajuste o normalización del contenido de GC (por ejemplo, GC-LOESS, GCRM). En ciertos casos, se identifica una muestra particular para la aplicación de un proceso de normalización de GC adicional. Por ejemplo, la aplicación de la metodología PERUN puede determinar el sesgo de GC para cada muestra, y puede seleccionarse una muestra asociada con un sesgo de GC por encima de un cierto umbral para un proceso de normalización de GC adicional. En tales casos, se puede usar un nivel de umbral predeterminado para seleccionar tales muestras para la normalización adicional de GC.

En ciertos casos, se puede utilizar un proceso de filtrado o ponderación de porciones junto con la metodología PERUN. Se puede utilizar un proceso adecuado de filtrado o ponderación de porciones, se describen ejemplos no limitantes en el presente documento, en la solicitud de patente internacional no. PCT/US12/59123 (WO2013/052913) y en la publicación de solicitud de patente de EE.UU. n.º US20130085681. En ciertos casos, se utiliza una técnica de normalización que reduce el error asociado con las inserciones, duplicaciones y/o deleciones maternas (por ejemplo, variaciones del número de copias maternas y/o fetales), junto con la metodología PERUN.

Los niveles de sección genómica calculados por la metodología PERUN se pueden utilizar directamente para proporcionar un resultado. En algunos casos, los niveles de sección genómica pueden utilizarse directamente para proporcionar un resultado para muestras en las que la fracción fetal es de aproximadamente un 2% a aproximadamente un 6% o mayor (por ejemplo, una fracción fetal de aproximadamente 4% o mayor). Los niveles de sección genómica calculados por la metodología PERUN a veces se procesan adicionalmente para proporcionar un resultado. En algunos casos, los niveles de sección genómica calculados se estandarizan. En ciertos casos, la suma, la media o la mediana de los niveles calculados de la sección genómica para una porción de ensayo (por ejemplo, el cromosoma 21) se puede dividir por la suma, la media o la mediana de los niveles calculados de la sección genómica para porciones distintas de la porción de ensayo (por ejemplo, autosomas distintos del cromosoma 21), para generar un nivel experimental de sección genómica. Se puede usar un nivel de sección genómica experimental o un nivel de sección genómica sin procesar como parte de un análisis de estandarización, como el cálculo de una puntuación Z . Se puede generar una puntuación Z para una muestra al restar un nivel de sección genómica esperado de un nivel de sección genómica experimental o un nivel de sección genómica sin procesar y el valor resultante puede dividirse por una desviación estándar para las muestras. En ciertos casos, las puntuaciones Z resultantes pueden distribuirse para diferentes muestras y analizarse, o pueden relacionarse con otras variables, como la fracción fetal y otras, y analizarse, para proporcionar un resultado.

Como se observa en el presente documento, la metodología PERUN no se limita a la normalización de acuerdo con el sesgo de GC y el contenido de GC per se, y puede usarse para reducir el error asociado con otras fuentes de error. Un ejemplo no limitante de una fuente de sesgo de contenido no GC es la capacidad de mapeo. Cuando se abordan otros parámetros de normalización distintos del sesgo y el contenido de GC, una o más de las relaciones ajustadas pueden ser no lineales (por ejemplo, hiperbólicas, exponenciales). En algunos casos, cuando el sesgo experimental se determina a partir de una relación no lineal, por ejemplo, se puede analizar una estimación de curvatura del sesgo experimental.

- La metodología PERUN se puede aplicar a una variedad de indicadores de ácidos nucleicos. Ejemplos no limitantes de indicadores de ácidos nucleicos son lecturas de secuencias de ácidos nucleicos y niveles de ácidos nucleicos en una ubicación particular en una micromatriz. Ejemplos no limitantes de lecturas de secuencia incluyen aquellas obtenidas a partir de ADN circulante libre de células, ARN circulante libre de células, ADN celular y ARN celular. La metodología PERUN se puede aplicar a las lecturas de secuencia mapeadas en secuencias de referencia adecuadas, como el ADN de referencia genómico, el ARN de referencia celular (por ejemplo, el transcriptoma) y porciones de los mismos (por ejemplo, parte(s) de un complemento genómico de ADN o ARN del transcriptoma, parte(s) de un cromosoma).
- Por lo tanto, en ciertos casos, el ácido nucleico celular (por ejemplo, ADN o ARN) puede servir como un indicador de ácidos nucleicos. Las lecturas de ácido nucleico celular mapeadas en las porciones del genoma de referencia pueden normalizarse utilizando la metodología PERUN. El ácido nucleico celular unido a una proteína particular a veces se refiere a procesos de inmunoprecipitación de cromatina (ChIP). El ácido nucleico enriquecido con ChIP es un ácido nucleico en asociación con proteínas celulares, como el ADN o el ARN, por ejemplo. Las lecturas de ácidos nucleicos enriquecidos con ChIP se pueden obtener usando tecnología conocida en la técnica. Las lecturas de ácidos nucleicos enriquecidos con ChIP se pueden mapear en una o más porciones de un genoma de referencia, y los resultados se pueden normalizar usando metodología PERUN para proporcionar un resultado.
- En ciertos casos, el ARN celular puede servir como indicadores de ácidos nucleicos. Las lecturas de ARN celular pueden mapearse a las porciones de ARN de referencia y normalizarse utilizando metodología PERUN para proporcionar un resultado. Las secuencias conocidas para el ARN celular, denominadas un transcriptoma, o un segmento del mismo, pueden usarse como una referencia a la que se pueden mapear las lecturas de ARN de una muestra. Las lecturas de ARN de muestra pueden obtenerse usando tecnología conocida en la técnica. Los resultados de las lecturas de ARN mapeadas en una referencia pueden normalizarse utilizando metodología PERUN para proporcionar un resultado.
- En algunos casos, los niveles de ácidos nucleicos en las micromatrices pueden servir como indicadores de ácidos nucleicos. Pueden analizarse usando la metodología PERUN los niveles de ácido nucleico, o ácido nucleico que hibrida, a través de muestras para un abordaje particular, en una matriz, normalizando de este modo, los indicadores de ácidos nucleicos proporcionados por el análisis de micromatrices. De esta manera, un abordaje particular o ácido nucleico que hibrida en una micromatriz es análogo a una porción para lecturas de secuencias de ácidos nucleicos mapeadas, y la metodología PERUN se puede usar para normalizar los datos de la micromatriz para proporcionar un mejor resultado.
- En algunos casos, una etapa de procesamiento comprende una ponderación. Los términos "ponderado", "ponderación" o "función de ponderación" o derivados gramaticales o equivalentes de los mismos, como se usa en el presente documento, se refieren a una manipulación matemática de una porción o la totalidad de un conjunto de datos que a veces se utiliza para alterar la influencia de ciertas características o variables del conjunto de datos con respecto a otras características o variables del conjunto de datos (por ejemplo, aumentar o disminuir la significación y/o la contribución de los datos contenidos en una o más porciones o porciones de un genoma de referencia, en función de la calidad o utilidad de los datos en la porción seleccionada o porciones de un genoma de referencia seleccionado). En algunos casos, se puede usar una función de ponderación para aumentar la influencia de los datos con una varianza de medición relativamente pequeña, y/o para disminuir la influencia de los datos con una varianza de medición relativamente grande. Por ejemplo, las porciones de un genoma de referencia con datos de secuencia de baja calidad o poco representados pueden ser "ponderadas a la baja" para minimizar la influencia en un conjunto de datos, mientras que las porciones seleccionadas de un genoma de referencia pueden ser "ponderadas al alza" para aumentar la influencia en un conjunto de datos. Un ejemplo no limitante de una función de ponderación es $[1/(\text{desviación estándar})^2]$. Una etapa de ponderación a veces se realiza de una manera sustancialmente similar a una etapa de normalización. En algunos casos, un conjunto de datos se divide una variable predeterminada (por ejemplo, variable de ponderación). Una variable predeterminada (por ejemplo, función objetivo minimizada, Phi) a menudo se selecciona para ponderar diferentes partes de un conjunto de datos de manera diferente (por ejemplo, aumentar la influencia de ciertos tipos de datos al tiempo que disminuye la influencia de otros tipos de datos).
- En ciertos casos, una etapa de procesamiento puede comprender una o más manipulaciones matemáticas y/o estadísticas. Cualquier manipulación matemática y/o estadística adecuada, sola o en combinación, puede usarse para analizar y/o manipular un conjunto de datos descrito en el presente documento. Se puede usar cualquier número adecuado de manipulaciones matemáticas y/o estadísticas. En algunos casos, un conjunto de datos puede manipularse, matemática y/o estadísticamente, 1 o más, 5 o más, 10 o más o 20 o más veces. Ejemplos no limitantes de manipulaciones matemáticas y estadísticas que pueden usarse incluyen la suma, resta, multiplicación, división, funciones algebraicas, estimadores de mínimos cuadrados, ajuste de curvas, ecuaciones diferenciales, polinomios racionales, polinomios dobles, polinomios ortogonales, puntuaciones Z, valores de p, valores de chi, valores de phi, análisis de los niveles de máximos, determinación de las ubicaciones de los bordes de los máximos, cálculo de las proporciones de área de los máximos, análisis del nivel cromosómico medio, cálculo de la desviación media absoluta, suma de restos al cuadrado, media, desviación estándar, error estándar, similares o combinaciones de los mismos. Se puede realizar una manipulación matemática y/o estadística en la totalidad o una porción de los datos de lecturas de secuencias, o en los productos procesados de los mismos. Ejemplos no limitantes de variables o características de

conjuntos de datos que pueden manipularse estadísticamente incluyen recuentos sin procesar, recuentos filtrados, recuentos normalizados, alturas de máximos, amplitudes de máximos, áreas de máximos, bordes de máximos, tolerancias laterales, valores de P, niveles mediana, niveles medios, distribución de recuentos dentro de una región genómica, representación relativa de especies de ácidos nucleicos, similares o combinaciones de los mismos.

5 En algunos casos, una etapa de procesamiento puede comprender el uso de uno o más algoritmos estadísticos. Cualquier algoritmo estadístico adecuado, solo o en combinación, puede usarse para analizar y/o manipular un conjunto de datos descrito en el presente documento. Se puede utilizar cualquier número adecuado de algoritmos estadísticos. En algunos casos, un conjunto de datos puede analizarse utilizando 1 o más, 5 o más, 10 o más o 20 o más algoritmos estadísticos. Ejemplos no limitantes de algoritmos estadísticos adecuados para su uso con los métodos descritos en el presente documento incluyen árboles de decisión, recuentos nulos, comparaciones múltiples, prueba ómnibus, problema de Behrens-Fisher, remuestreo, método de Fisher para combinar pruebas independientes de significación, hipótesis nula, error de tipo I, error de tipo II, prueba exacta, prueba Z de una muestra, prueba Z de dos muestras, prueba t de una muestra, prueba t pareada, prueba t combinada de dos muestras con iguales varianzas, prueba t no combinada de dos muestras con varianzas distintas, prueba z para una proporción, prueba z para dos proporciones combinada, prueba z para dos proporciones no combinada, prueba chi cuadrado para una muestra, prueba F para dos muestras para igualdad de varianzas, intervalo de confianza, intervalo creíble, significación, meta análisis, regresión lineal simple, regresión lineal robusta, similares o combinaciones de los anteriores. Ejemplos no limitantes de variables o características de conjuntos de datos que pueden analizarse utilizando algoritmos estadísticos incluyen recuentos sin procesar, recuentos filtrados, recuentos normalizados, alturas de máximos, amplitudes de máximos, bordes de máximos, tolerancias laterales, valores de P, niveles mediana, niveles medios, distribución de recuentos dentro de una región genómica, representación relativa de especies de ácidos nucleicos, similares o combinaciones de los mismos.

25 En ciertos casos, un conjunto de datos puede analizarse utilizando múltiples algoritmos estadísticos (por ejemplo, 2 o más) (por ejemplo, regresión de mínimos cuadrados, análisis de componentes principales, análisis discriminante lineal, análisis discriminante cuadrático, empaquetado, redes neuronales, modelos de máquinas de soporte vectorial, bosques aleatorios, modelos de árboles de clasificación, K vecinos más cercanos, regresión logística y/o atenuado de pérdidas) y/o manipulaciones matemáticas y/o estadísticas (por ejemplo, denominadas en el presente documento, manipulaciones). En algunos casos, el uso de manipulaciones múltiples puede generar un espacio N-dimensional que puede usarse para proporcionar un resultado. En ciertos casos, el análisis de un conjunto de datos utilizando múltiples manipulaciones puede reducir la complejidad y/o dimensionalidad del conjunto de datos. Por ejemplo, el uso de manipulaciones múltiples en un conjunto de datos de referencia puede generar un espacio N-dimensional (por ejemplo, representación gráfica de probabilidad) que puede usarse para representar la presencia o ausencia de una variación genética, dependiendo del estado genético de las muestras de referencia (por ejemplo, positivo o negativo para una variación genética seleccionada). El análisis de muestras de ensayo utilizando un conjunto de manipulaciones sustancialmente similar se puede usar para generar un punto N-dimensional para cada una de las muestras de ensayo. La complejidad y/o dimensionalidad del conjunto de datos de un sujeto de ensayo a veces se reduce a un valor único o punto N-dimensional que se puede comparar fácilmente con el espacio N-dimensional generado a partir de los datos de referencia. Los datos de la muestra de ensayo que se encuentran dentro del espacio N-dimensional poblado por los datos del sujeto de referencia son indicativos de un estado genético sustancialmente similar al de los sujetos de referencia. Los datos de la muestra de ensayo que se encuentran fuera del espacio N-dimensional poblado por los datos del sujeto de referencia son indicativos de un estado genético sustancialmente diferente al de los sujetos de referencia. En algunos casos, las referencias son euploides o por otro lado no tienen una variación genética o afección médica.

Después de que los conjuntos de datos se hayan contado, opcionalmente filtrado y normalizado, en algunos casos, los conjuntos de datos procesados pueden manipularse adicionalmente mediante uno o más procedimientos de filtrado y/o normalización. En ciertos casos, un conjunto de datos que ha sido manipulado adicionalmente por uno o más procedimientos de filtrado y/o normalización se puede usar para generar un perfil. En algunos casos, el uno o más procedimientos de filtrado y/o normalización a veces pueden reducir la complejidad y/o dimensionalidad del conjunto de datos. Se puede proporcionar un resultado basado en un conjunto de datos de complejidad y/o dimensionalidad reducidas.

55 En algunos casos, las porciones se pueden filtrar de acuerdo con una medida de error (por ejemplo, desviación estándar, error estándar, varianza calculada, valor de p, error medio absoluto (MAE), desviación absoluta promedio y/o desviación absoluta media (MAD)). En ciertos casos, una medida de error se refiere a la variabilidad del recuento. En algunos casos, las porciones se filtran de acuerdo a la variabilidad del recuento. En ciertos casos, la variabilidad del recuento es una medida de error determinada para los recuentos mapeados en una porción (es decir, una porción) de un genoma de referencia para muestras múltiples (por ejemplo, muestra múltiple obtenida de múltiples sujetos, por ejemplo, 50 o más, 100 o más, 500 o más, 1000 o más, 5000 o más o 10.000 o más sujetos). En algunos casos, se filtran las porciones con una variabilidad de recuento por encima de un intervalo superior predeterminado (por ejemplo, excluirse del estudio). En algunos casos, un intervalo superior predeterminado es un valor MAD igual o mayor que aproximadamente 50, aproximadamente 52, aproximadamente 54, aproximadamente 56, aproximadamente 58, aproximadamente 60, aproximadamente 62, aproximadamente 64, aproximadamente 66, aproximadamente 68, aproximadamente 70, aproximadamente 72, aproximadamente 74 o igual a o mayor que aproximadamente 76. En

algunos casos, se filtran las porciones con una variabilidad de recuento por debajo de un intervalo inferior predeterminado (por ejemplo, excluirse del estudio). En algunos casos, un intervalo inferior predeterminado es un valor MAD igual a o menor que aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25, aproximadamente 20, aproximadamente 15, aproximadamente 10, aproximadamente 5, aproximadamente 1 o igual a o menor que aproximadamente 0. En algunos casos, se filtran las porciones con una variabilidad de recuento dentro de un intervalo predeterminado (por ejemplo, excluirse del estudio). En algunos casos, un intervalo predeterminado es un valor MAD mayor que cero o menor que aproximadamente 76, menor que aproximadamente 74, menor que aproximadamente 73, menor que aproximadamente 72, menor que aproximadamente 71, menor que aproximadamente 70, menor que aproximadamente 69, menor que aproximadamente 68, menor que aproximadamente 67, menor que aproximadamente 66, menor que aproximadamente 65, menor que aproximadamente 64, menor que aproximadamente 62, menor que aproximadamente 60, menor que aproximadamente 58, menor que aproximadamente 56, menor que aproximadamente 54, menor que aproximadamente 52 o menor que aproximadamente 50. En algunos casos, un intervalo predeterminado es un valor MAD mayor que cero o menor que aproximadamente 67,7. En algunos casos, se seleccionan porciones con una variabilidad de recuento dentro de un intervalo predeterminado (por ejemplo, para determinar la presencia o ausencia de una variación genética).

En algunos casos, la variabilidad de recuento de porciones representa una distribución (por ejemplo, una distribución normal). En algunos casos, las porciones se seleccionan dentro de un cuantil de la distribución. En algunos casos, se seleccionan porciones dentro de un cuantil igual o inferior a aproximadamente el 99,9%, 99,8 %, 99,7 %, 99,6 %, 99,5 %, 99,4 %, 99,3 %, 99,2 %, 99,1 %, 99,0 %, 98,9 %, 98,8 %, 98,7 %, 98,6 %, 98,5 %, 98,4 %, 98,3 %, 98,2 %, 98,1 %, 98,0 %, 97 %, 96 %, 95 %, 94 %, 93 %, 92 %, 91 %, 90 %, 85 %, 80%, o igual o inferior a un cuantil de aproximadamente el 75% para la distribución. En algunos casos, se seleccionan porciones dentro de un cuantil del 99% de la distribución de la variabilidad del recuento. En algunos caos, se seleccionan porciones con una MAD > 0 y una MAD <67,725 dentro del cuantil del 99%, dando como resultado la identificación de un conjunto de porciones estables de un genoma de referencia.

Ejemplos no limitantes de filtrado de porciones con respecto a PERUN se proporcionan en el presente documento y en la solicitud de patente internacional nº PCT/US12/59123 (documento WO2013/052913). Las porciones se pueden filtrar basándose en o basándose en parte en, una medida de error. En ciertos casos, se puede utilizar una medida de error que comprende valores absolutos de desviación, tal como un factor R, para la eliminación o ponderación de porciones. En algunos casos, un factor R, se define como la suma de las desviaciones absolutas de los valores de recuentos predichos de las mediciones reales dividida por los valores de recuentos predichos de las mediciones reales (por ejemplo, la ecuación B en el presente documento). Si bien puede usarse una medida de error que comprende valores absolutos de desviación, se puede emplear alternativamente una medida de error adecuada. En ciertos casos, se puede utilizar una medida de error que no incluya valores absolutos de desviación, como una dispersión basada en cuadrados. En algunos casos, las porciones se filtran o ponderan de acuerdo con una medida de capacidad de mapeo (por ejemplo, puntuación de capacidad de mapeo). A veces, una porción se filtra o se pondera de acuerdo con un número relativamente bajo de lecturas de secuencia mapeadas en la porción (por ejemplo, 0, 1, 2, 3, 4, 5 lecturas mapeadas en la porción). Las porciones se pueden filtrar o ponderar de acuerdo con el tipo de análisis que se realiza. Por ejemplo, para el análisis de aneuploidía del cromosoma 13, 18 y/o 21, pueden filtrarse los cromosomas sexuales y pueden analizarse solo los autosomas, o un subconjunto de autosomas. Para la determinación del sexo fetal, los autosomas pueden filtrarse y se pueden analizar solo los cromosomas sexuales (X e Y) o uno de los cromosomas sexuales (X o Y).

En casos particulares, se puede emplear el siguiente proceso de filtrado. Se selecciona el mismo conjunto de porciones (por ejemplo, porciones de un genoma de referencia) dentro de un cromosoma dado (por ejemplo, el cromosoma 21) y se compara el número de lecturas en muestras afectadas y no afectadas. La diferencia se relaciona con las muestras de trisomía 21 y euploides e implica un conjunto de porciones que cubren la mayor parte del cromosoma 21. El conjunto de porciones es el mismo entre las muestras euploides y T21. La distinción entre un conjunto de porciones y una sola sección no es crucial, ya que se puede definir una porción. La misma región genómica se compara en diferentes pacientes. Este proceso se puede utilizar para un análisis de trisomía, como para T13 o T18 además de, o en lugar de, T21.

Después de que los conjuntos de datos se hayan contado, opcionalmente filtrado y normalizado, en algunos casos, los conjuntos de datos procesados se pueden manipular por ponderación. En ciertos casos, se pueden seleccionar una o más porciones para ponderar para reducir la influencia de los datos (por ejemplo, datos ruidosos, datos no informativos) contenidos en las porciones seleccionadas, y en algunos casos se puede seleccionar una o más porciones para ponderar para mejorar o aumentar la influencia de los datos (por ejemplo, datos con pequeña varianza medida) contenidos en las porciones seleccionadas. En algunos casos, se pondera un conjunto de datos utilizando una función de ponderación única que disminuye la influencia de los datos con grandes varianzas y aumenta la influencia de los datos con pequeñas varianzas. A veces se usa una función de ponderación para reducir la influencia de los datos con grandes varianzas y aumentar la influencia de los datos con pequeñas varianzas (por ejemplo, $[1/(\text{desviación estándar})^2]$). En algunos casos, se genera una representación gráfica de perfil de datos procesados manipulados adicionalmente por ponderación para facilitar la clasificación y/o proporcionar un resultado. Se puede proporcionar un resultado basado en una representación gráfica de perfil de datos ponderados.

El filtrado o la ponderación de las porciones se puede realizar en uno o más puntos adecuados en un análisis. Por ejemplo, las porciones pueden filtrarse o ponderarse antes o después de que las lecturas de secuencias se mapeen a porciones de un genoma de referencia. Las porciones pueden filtrarse o ponderarse antes o después de que se determine un sesgo experimental para las porciones individuales del genoma. En algunos casos, las porciones pueden filtrarse o ponderarse antes o después de que se calculen los niveles de la sección genómica.

Después de que los conjuntos de datos se hayan contado, opcionalmente filtrado, normalizado y opcionalmente ponderado, en algunos casos, los conjuntos de datos procesados pueden manipularse mediante una o más manipulaciones matemáticas y/o estadísticas (por ejemplo, funciones estadísticas o algoritmos estadísticos). En ciertos casos, los conjuntos de datos procesados pueden manipularse adicionalmente calculando las puntuaciones Z para una o más porciones, cromosomas o porciones de cromosomas seleccionados. En ciertos casos, los conjuntos de datos procesados pueden manipularse adicionalmente calculando los valores de P. Un caso de una ecuación para calcular una puntuación Z y un valor p se presenta en la Ecuación 1 (Ejemplo 2). En ciertos casos, las manipulaciones matemáticas y/o estadísticas incluyen uno o más supuestos relacionados con la ploidía y/o la fracción fetal. En algunos casos, se genera una representación gráfica de perfil de datos procesados manipulados adicionalmente por una o más manipulaciones estadísticas y/o matemáticas para facilitar la clasificación y/o proporcionar un resultado. Se puede proporcionar un resultado basado en una representación gráfica de perfil de datos manipulados estadística y/o matemáticamente. Un resultado proporcionado basándose en una representación gráfica de perfil de datos estadísticamente y/o matemáticamente manipulados incluye a menudo uno o más supuestos relacionados con la ploidía y/o la fracción fetal.

En ciertos casos, se realizan múltiples manipulaciones en los conjuntos de datos procesados para generar un espacio N-dimensional y/o un punto N-dimensional, después de que los conjuntos de datos se hayan contado, opcionalmente filtrado y normalizado. Se puede proporcionar un resultado basado en una representación gráfica de perfil de conjuntos de datos analizados en N-dimensiones.

En algunos casos, los conjuntos de datos se procesan utilizando uno o más análisis de nivel de máximo, análisis de amplitud de máximos, análisis de ubicación de bordes de máximo, tolerancias laterales de máximo, similares, derivaciones de los mismos, o combinaciones de los anteriores, como parte de o después de que los conjuntos de datos se hayan procesado y/o manipulado. En algunos casos, se genera una representación gráfica de perfil de los datos procesados utilizando uno o más análisis de nivel de máximo, análisis de amplitudes de máximos, análisis de ubicación de bordes de máximo, tolerancias laterales de máximo, similares, derivaciones de los mismos o combinaciones de los anteriores para facilitar la clasificación y/o proporcionar un resultado. Se puede proporcionar un resultado basado en una representación gráfica de perfil de datos que se han procesado utilizando uno o más análisis de nivel de máximo, análisis de amplitud de máximos, análisis de ubicación de bordes de máximo, tolerancias laterales de máximo, similares, derivaciones de los mismos, o combinaciones de los anteriores.

En algunos casos, el uso de una o más muestras de referencia que están sustancialmente libres de una variación genética en cuestión se puede usar para generar un perfil de recuentos mediana de referencia, lo que puede dar como resultado un valor predeterminado representativo de la ausencia de la variación genética, y a menudo se desvía de un valor predeterminado en áreas correspondientes a la ubicación genómica en donde se encuentra la variación genética en el sujeto de ensayo, si el sujeto de ensayo poseyera la variación genética. En sujetos de ensayo con riesgo de, o que padecen una afección médica asociada con una variación genética, se espera que el valor numérico para la porción o las secciones seleccionadas varíe significativamente del valor predeterminado para las ubicaciones genómicas no afectadas. En ciertos casos, el uso de una o más muestras de referencia que se sabe que portan la variación genética en cuestión se puede usar para generar un perfil de recuentos mediana de referencia, lo que puede dar como resultado un valor predeterminado representativo de la presencia de la variación genética, y a menudo se desvía de un valor predeterminado en áreas correspondientes a la ubicación genómica en donde un sujeto de ensayo no porta la variación genética. En sujetos de ensayo con riesgo de, o que padecen una afección médica asociada con una variación genética, se espera que el valor numérico para la porción o las secciones seleccionadas varíe significativamente del valor predeterminado para las ubicaciones genómicas afectadas.

En algunos casos, el análisis y procesamiento de datos puede incluir el uso de uno o más supuestos. Se puede utilizar un número o tipo de supuestos adecuados para analizar o procesar un conjunto de datos. Ejemplos no limitantes de supuestos que se pueden usar para el procesamiento y/o análisis de datos incluyen ploidía materna, contribución fetal, prevalencia de ciertas secuencias en una población de referencia, antecedentes étnicos, prevalencia de una afección médica seleccionada en miembros de la familia relacionados, paralelismo entre los perfiles de recuentos sin procesar de diferentes pacientes y/o series después de la normalización de GC y la repetición del enmascaramiento (por ejemplo, GCRM), las coincidencias idénticas representan artefactos de PCR (por ejemplo, posición de base idéntica), supuestos inherentes a un ensayo de cuantificación fetal (por ejemplo, FQA), supuestos con respecto a gemelos (p. ej., 2 gemelos y solo 1 está afectado, la fracción fetal eficaz es solo el 50% de la fracción fetal medida total (de manera similar para los trillizos, los cuatrillizos y similares)), el ADN libre de células fetal (p. ej., ADNcf) cubre uniformemente el genoma completo, similares y combinaciones de los mismos.

En aquellos casos en los que la calidad y/o la profundidad de las lecturas de secuencias mapeadas no permiten una predicción de resultados de la presencia o ausencia de una variación genética a un nivel de confianza deseado (por

ejemplo, un nivel de confianza del 95% o superior), basándose en los perfiles de recuentos normalizados, pueden utilizarse uno o más algoritmos de manipulación matemática y/o algoritmos de predicción estadística adicionales, para generar valores numéricos adicionales útiles para el análisis de datos y/o para proporcionar un resultado. La expresión "perfil de recuento normalizado", como se usa en el presente documento, se refiere a un perfil generado usando recuentos normalizados. En el presente documento se describen ejemplos de métodos que se pueden usar para generar recuentos normalizados y perfiles de recuentos normalizados. Como se ha indicado, se pueden normalizar las lecturas de secuencia mapeadas que se han contado con respecto a los recuentos de muestras de ensayo o recuentos de muestras de referencia. En algunos casos, un perfil de recuento normalizado se puede presentar como una representación gráfica.

Perfiles

En algunos casos, una etapa de procesamiento puede comprender generar uno o más perfiles (por ejemplo, representación gráfica de perfil) a partir de diversos aspectos de un conjunto de datos o su derivación (por ejemplo, producto de una o más etapas de procesamiento matemático y/o estadístico de datos conocidas en la técnica y/o descritas en el presente documento). El término "perfil" como se usa en el presente documento se refiere a un producto de una manipulación matemática y/o estadística de datos que puede facilitar la identificación de patrones y/o correlaciones en grandes cantidades de datos. Un "perfil" a menudo incluye valores resultantes de una o más manipulaciones de datos o conjuntos de datos, basadas en uno o más criterios. Un perfil a menudo incluye múltiples puntos de datos. Cualquier número adecuado de puntos de datos puede incluirse en un perfil dependiendo de la naturaleza y/o la complejidad de un conjunto de datos. En ciertos casos, los perfiles pueden incluir 2 o más puntos de datos, 3 o más puntos de datos, 5 o más puntos de datos, 10 o más puntos de datos, 24 o más puntos de datos, 25 o más puntos de datos, 50 o más puntos de datos, 100 o más puntos de datos, 500 o más puntos de datos, 1000 o más puntos de datos, 5000 o más puntos de datos, 10.000 o más puntos de datos, o 100.000 o más puntos de datos.

En algunos casos, un perfil es representativo de la totalidad de un conjunto de datos, y en ciertos casos, un perfil es representativo de una parte o subconjunto de un conjunto de datos. Es decir, un perfil a veces incluye o se genera a partir de puntos de datos representativos de datos que no se han filtrado para eliminar cualquier dato, y en ocasiones un perfil incluye o se genera a partir de puntos de datos representativos de datos que se han filtrado para eliminar datos no deseados. En algunos casos, un punto de datos en un perfil representa los resultados de la manipulación de datos para una porción. En ciertos casos, un punto de datos en un perfil incluye resultados de manipulación de datos para grupos de porciones. En algunos casos, los grupos de porciones pueden ser adyacentes entre sí, y en ciertos casos, los grupos de porciones pueden ser de diferentes partes de un cromosoma o genoma.

Los puntos de datos en un perfil derivado de un conjunto de datos pueden ser representativos de cualquier clasificación de datos adecuada. Ejemplos no limitantes de categorías en las que se pueden agrupar los datos para generar puntos de datos de perfil incluyen: porciones basadas en el tamaño, porciones basadas en características de secuencia (por ejemplo, en el contenido de GC, en el contenido de AT, posición en un cromosoma (por ejemplo, brazo corto, brazo largo, centrómero, telómero y similares), niveles de expresión, cromosoma, similares o combinaciones de los mismos. En algunos casos, se puede generar un perfil a partir de puntos de datos obtenidos de otro perfil (por ejemplo, perfil de datos normalizado renormalizado a un valor de normalización diferente para generar un perfil de datos renormalizado). En ciertos casos, un perfil generado a partir de puntos de datos obtenidos de otro perfil reduce el número de puntos de datos y/o la complejidad del conjunto de datos. La reducción del número de puntos de datos y/o la complejidad de un conjunto de datos a menudo facilita la interpretación de los datos y/o facilita proporcionar un resultado.

Un perfil (por ejemplo, un perfil genómico, un perfil cromosómico, un perfil de un segmento de un cromosoma) a menudo es una colección de recuentos normalizados o no normalizados para dos o más porciones. Un perfil a menudo incluye al menos un nivel (por ejemplo, un nivel de sección genómica), y a menudo comprende dos o más niveles (por ejemplo, un perfil a menudo tiene múltiples niveles). Un nivel generalmente es para un conjunto de porciones que tienen aproximadamente los mismos recuentos o recuentos normalizados. Los niveles se describen en mayor detalle en el presente documento. En ciertos casos, un perfil comprende una o más porciones, cuyas porciones se pueden ponderar, eliminar, filtrar, normalizar, ajustar, promediar, derivar como una media, sumar, restar, procesar o transformar por cualquier combinación de los mismos. Un perfil a menudo comprende recuentos normalizados mapeados en porciones que definen dos o más niveles, donde los recuentos se normalizan adicionalmente de acuerdo con uno de los niveles mediante un método adecuado. A menudo, los recuentos de un perfil (por ejemplo, un nivel de perfil) están asociados con un valor de incertidumbre.

Un perfil que comprende uno o más niveles a veces se rellena (por ejemplo, ajuste de huecos (*hole padding*)). *Padding* (por ejemplo, *hole padding*) se refiere a un proceso de identificación y ajuste de niveles en un perfil que se debe a microdelecciones o duplicaciones maternas (por ejemplo, variaciones en el número de copias). En algunos casos, los niveles pueden encontrarse ajustados debido a microduplicaciones fetales o microdelecciones fetales. Las microduplicaciones o microdelecciones en un perfil pueden aumentar o disminuir artificialmente el nivel general de un perfil (por ejemplo, un perfil de un cromosoma) conduciendo a determinaciones positivas falsas o negativas falsas de una aneuploidía cromosómica (por ejemplo, una trisomía). En algunos casos, los niveles en un perfil que se deben a microduplicaciones y/o deleciones se identifican y ajustan (por ejemplo, se rellenan y/o eliminan) mediante un proceso

a veces denominado *padding* o *hole padding*. En ciertos casos, un perfil comprende uno o más primeros niveles que son significativamente diferentes a un segundo nivel dentro del perfil, cada uno de los primeros niveles comprende una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales y se ajustan uno o más de los primeros niveles.

Un perfil que comprende uno o más niveles puede incluir un primer nivel y un segundo nivel. En algunos casos, un primer nivel es diferente (por ejemplo, significativamente diferente) a un segundo nivel. En algunos casos, un primer nivel comprende un primer conjunto de porciones, un segundo nivel comprende un segundo conjunto de porciones y el primer conjunto de porciones no es un subconjunto del segundo conjunto de porciones. En ciertos casos, un primer conjunto de porciones es diferente de un segundo conjunto de porciones a partir del cual se determinan un primer y segundo nivel. En algunos casos, un perfil puede tener varios primeros niveles diferentes (por ejemplo, significativamente diferentes, por ejemplo, tienen un valor significativamente diferente) que un segundo nivel dentro del perfil. En algunos casos, un perfil comprende uno o más primeros niveles que son significativamente diferentes a un segundo nivel dentro del perfil y se ajustan uno o más de los primeros niveles. En algunos casos, un perfil comprende uno o más primeros niveles que son significativamente diferentes a un segundo nivel dentro del perfil, cada uno de los primeros niveles comprende una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales y se ajustan uno o más de los primeros niveles. En algunos casos, un primer nivel dentro de un perfil se elimina del perfil o se ajusta (por ejemplo, se rellena). Un perfil puede comprender varios niveles que incluyen uno o más primeros niveles significativamente diferentes de uno o más segundos niveles y, a menudo, la mayoría de los niveles en un perfil son segundos niveles, cuyos segundos niveles son aproximadamente iguales entre sí. En algunos casos, más del 50%, más del 60 %, más del 70 %, más del 80 %, más del 90% o más del 95% de los niveles en un perfil son segundos niveles.

Un perfil a veces se muestra como una representación gráfica. Por ejemplo, se pueden representar gráficamente y visualizar uno o más niveles que representan recuentos (por ejemplo, recuentos normalizados) de porciones. Ejemplos no limitantes de representaciones gráficas de perfil que se pueden generar incluyen recuento sin procesar (p. ej., perfil de recuentos sin procesar o perfil sin procesar), recuento normalizado, ponderado de porción, puntuación z, valor de p, relación de área frente a ploidía ajustada, nivel mediana frente a relación entre la fracción fetal ajustada y medida, componentes principales, similares, o combinaciones de los mismos. En algunos casos, las representaciones gráficas de perfil permiten la visualización de los datos manipulados. En ciertos casos, se puede utilizar una representación gráfica de perfil para proporcionar un resultado (por ejemplo, relación de área frente a ploidía ajustada, nivel mediana frente a relación entre la fracción fetal ajustada y medida, componentes principales).

Las expresiones "representación gráfica de perfil de recuentos sin procesar" o "representación gráfica de perfil sin procesar" como se usan en el presente documento se refieren a una representación gráfica de recuentos en cada porción en una región normalizada a recuentos totales en una región (por ejemplo, genoma, porción, cromosoma, porciones de un genoma de referencia o un segmento de un cromosoma). En algunos casos, un perfil puede generarse utilizando un proceso de ventana estática, y un perfil puede generarse utilizando un proceso de ventana deslizante.

Un perfil generado para un sujeto de ensayo a veces se compara con un perfil generado para uno o más sujetos de referencia, para facilitar la interpretación de manipulaciones matemáticas y/o estadísticas de un conjunto de datos y/o para proporcionar un resultado. En algunos casos, se genera un perfil basado en uno o más supuestos iniciales (por ejemplo, contribución materna de ácido nucleico (por ejemplo, fracción materna), contribución fetal de ácido nucleico (por ejemplo, fracción fetal), ploidía de la muestra de referencia, similares o combinaciones de los mismos). En ciertos casos, un perfil de ensayo a menudo se centra alrededor de un valor predeterminado representativo de la ausencia de una variación genética, y a menudo se desvía de un valor predeterminado en áreas correspondientes a la ubicación genómica en donde se encuentra la variación genética en el sujeto de ensayo, si el sujeto de ensayo poseyera la variación genética. En sujetos de ensayo con riesgo de, o que padecen una afección médica asociada con una variación genética, se espera que el valor numérico para la porción seleccionada varíe significativamente del valor predeterminado para las ubicaciones genómicas no afectadas. Dependiendo de las suposiciones iniciales (p. ej., ploidía fija o ploidía optimizada, fracción fetal fija o fracción fetal optimizada o combinaciones de las mismas), el umbral o valor de corte predeterminado o el intervalo de valores del umbral indicativo de la presencia o ausencia de una variación genética puede variar mientras se proporciona un resultado útil para determinar la presencia o ausencia de una variación genética. En algunos casos, un perfil es indicativo y/o representativo de un fenotipo.

A modo de ejemplo no limitativo, pueden obtenerse muestras normalizadas y/o perfiles de recuento de referencia a partir de datos de lecturas de secuencias sin procesar (a) calculando recuentos de la mediana de referencia para cromosomas seleccionados, porciones o segmentos de los mismos de un conjunto de referencias que se sabe que no portan una variación genética, (b) eliminando porciones no informativas de los recuentos sin procesar de la muestra de referencia (por ejemplo, filtrado); (c) normalizando los recuentos de referencia para todas las porciones restantes de un genoma de referencia al número total residual de recuentos (por ejemplo, suma de los recuentos restantes después de la eliminación de las porciones no informativas de un genoma de referencia) para el cromosoma seleccionado de la muestra de referencia o la ubicación genómica seleccionada, generando, de este modo, un perfil de sujeto de referencia normalizado; (d) eliminando las porciones correspondientes de la muestra del sujeto de ensayo;

y (e) normalizando los recuentos de sujetos de ensayo restantes para una o más ubicaciones genómicas seleccionadas a la suma de los recuentos mediana de referencia residuales para el cromosoma o cromosomas que contienen las ubicaciones genómicas seleccionadas, generando, de este modo, un perfil de sujeto de ensayo normalizado. En ciertos casos, entre (c) y (d) puede incluirse una etapa de normalización adicional con respecto al
5 genoma completo, reducido por las porciones filtradas en (b).

Se puede generar un perfil de conjunto de datos mediante una o más manipulaciones de datos de lecturas de secuencias mapeadas contadas. Algunos casos incluyen los siguientes. Se mapean las lecturas de secuencia y se determina el número de recuentos (es decir, marcadores de secuencia) para cada porción genómica (por ejemplo,
10 contada). Se genera un perfil de recuentos sin procesar a partir de las lecturas de secuencia mapeadas que se cuentan. En ciertos casos, se proporciona un resultado comparando un perfil de recuentos sin procesar de un sujeto de ensayo con un perfil de recuento mediana de referencia para los cromosomas, porciones o segmentos de los mismos de un conjunto de sujetos de referencia que no poseen una variación genética.

15 Opcionalmente, en algunos casos, se filtran los datos de lecturas de secuencias para eliminar datos ruidosos o porciones no informativas. Después del filtrado, los recuentos restantes normalmente se suman para generar un conjunto de datos filtrados. En ciertos casos, se genera un perfil de recuentos filtrados a partir de un conjunto de datos filtrados.

20 Después de que los datos de lecturas de secuencias se hayan contado y filtrado opcionalmente, los conjuntos de datos se pueden normalizar para generar niveles o perfiles. Un conjunto de datos puede normalizarse normalizando una o más porciones seleccionadas a un valor de referencia de normalización adecuado. En algunos casos, un valor de referencia de normalización es representativo de los recuentos totales para el cromosoma o cromosomas a partir de los cuales se seleccionan las porciones. En ciertos casos, un valor de referencia de normalización es representativo
25 de una o más porciones correspondientes, porciones de cromosomas o cromosomas de un conjunto de datos de referencia preparado a partir de un conjunto de sujetos de referencia que se sabe que no poseen una variación genética. En algunos casos, un valor de referencia de normalización es representativo de una o más porciones correspondientes, porciones de cromosomas o cromosomas de un conjunto de datos de sujetos de ensayo preparado a partir de un sujeto de ensayo que se está analizando para detectar la presencia o ausencia de una variación genética.
30 En ciertos casos, el proceso de normalización se realiza utilizando una aproximación de ventana estática, y el proceso de normalización se realiza utilizando una aproximación de ventana móvil o deslizante. En ciertos casos, se genera un perfil que comprende recuentos normalizados para facilitar la clasificación y/o proporcionar un resultado. Se puede proporcionar un resultado basado en una representación gráfica de un perfil que comprende los recuentos normalizados (por ejemplo, utilizando una representación gráfica de dicho perfil).

35 *Niveles*

En algunos casos, a un nivel se le asigna un valor (por ejemplo, un número, un valor cuantitativo). Un nivel puede determinarse mediante un método, operación o proceso matemático adecuado (por ejemplo, un nivel procesado). Un
40 nivel a menudo es, o deriva de, recuentos (por ejemplo, recuentos normalizados) para un conjunto de porciones. En algunos casos, un nivel de una porción es sustancialmente igual al número total de recuentos mapeados en una porción (por ejemplo, recuentos, recuentos normalizados). A menudo, un nivel se determina a partir de los recuentos que se procesan, transforman o manipulan mediante un método, operación o proceso matemático adecuado conocido en la técnica. En algunos casos, un nivel deriva de los recuentos que se procesan y los ejemplos no limitantes de
45 recuentos procesados incluyen los recuentos ponderados, eliminados, filtrados, normalizados, ajustados, promediados, derivados como una media (por ejemplo, nivel medio), sumados, restados, transformados o una combinación de los mismos. En algunos casos, un nivel comprende recuentos que están normalizados (por ejemplo, recuentos normalizados de porciones). Un nivel puede ser para los recuentos normalizados por un proceso adecuado, cuyos ejemplos no limitantes incluyen normalización por porciones, normalización por contenido de GC, regresión de
50 mínimos cuadrados lineal y no lineal, LOESS de GC, LOWESS, PERUN, RM, GCRM, cQn, similares y/o combinaciones de los mismos. Un nivel puede comprender recuentos normalizados o cantidades relativas de recuentos. En algunos casos, un nivel es para recuentos o recuentos normalizados de dos o más porciones que se promedian y el nivel se conoce como un nivel promedio. En algunos casos, un nivel es para un conjunto de porciones que tienen un recuento medio o una media de recuentos normalizados que se conoce como un nivel medio. En algunos
55 casos, un nivel deriva para porciones que comprenden recuentos sin procesar y/o filtrados. En algunos casos, un nivel se basa en recuentos que están sin procesar. En algunos casos, un nivel está asociado con un valor de incertidumbre (por ejemplo, una desviación estándar, una MAD). En algunos casos, un nivel está representado por una puntuación Z o valor de p.

60 Un nivel para una o más porciones es sinónimo de "nivel de sección genómica" en el presente documento. El término "nivel" como se usa en el presente documento a veces es sinónimo del término "elevación". La determinación del significado del término "nivel" se puede determinar a partir del contexto en donde se utiliza. Por ejemplo, el término "nivel", cuando se usa en el contexto de secciones genómicas, perfiles, lecturas y/o recuentos a menudo significa elevación. El término "nivel", cuando se usa en el contexto de una sustancia o composición (p.ej., nivel de ARN o nivel
65 de plexión) a menudo se refiere a una cantidad. El término "nivel", cuando se usa en el contexto de incertidumbre (p.ej., nivel de error, nivel de confianza, nivel de desviación, nivel de incertidumbre) a menudo se refiere a una cantidad.

Los recuentos normalizados o no normalizados para dos o más niveles (por ejemplo, dos o más niveles en un perfil) a veces pueden manipularse matemáticamente (por ejemplo, sumarse, multiplicarse, promediarse, normalizarse, similares, o combinaciones de los mismos) de acuerdo con los niveles. Por ejemplo, los recuentos normalizados o no normalizados para dos o más niveles pueden normalizarse de acuerdo con uno, algunos o todos los niveles en un perfil. En algunos casos, los recuentos normalizados o no normalizados de todos los niveles en un perfil se normalizan de acuerdo con un nivel en el perfil. En algunos casos, los recuentos normalizados o no normalizados de un primer nivel en un perfil se normalizan de acuerdo con los recuentos normalizados o no normalizados de un segundo nivel en el perfil.

Ejemplos no limitantes de un nivel (por ejemplo, un primer nivel, un segundo nivel) son un nivel para un conjunto de porciones que comprenden recuentos procesados, un nivel para un conjunto de porciones que comprende una media, mediana o promedio de recuentos, un nivel para un conjunto de porciones que comprenden recuentos normalizados, similares o cualquier combinación de los mismos. En algunos casos, un primer nivel y un segundo nivel en un perfil derivan de recuentos de porciones mapeadas en el mismo cromosoma. En algunos casos, un primer nivel y un segundo nivel en un perfil derivan de recuentos de porciones mapeadas en diferentes cromosomas.

En algunos casos, un nivel se determina a partir de recuentos normalizados o no normalizados mapeados en una o más porciones. En algunos casos, un nivel se determina a partir de recuentos normalizados o no normalizados mapeados en dos o más porciones, donde los recuentos normalizados para cada porción a menudo son aproximadamente iguales. Puede haber variación en los recuentos (por ejemplo, recuentos normalizados) en un conjunto de porciones para un nivel. En un conjunto de porciones para un nivel, puede haber una o más porciones que tengan recuentos que sean significativamente diferentes a los de otras porciones del conjunto (por ejemplo, máximos y/o mínimos). Cualquier número adecuado de recuentos normalizados o no normalizados asociados con cualquier número adecuado de porciones puede definir un nivel.

En algunos casos, se pueden determinar uno o más niveles a partir de recuentos normalizados o no normalizados de todas o algunas de las porciones de un genoma. A menudo, un nivel puede determinarse a partir de todos o algunos de los recuentos normalizados o no normalizados de un cromosoma, o segmento del mismo. En algunos casos, dos o más recuentos derivados de dos o más porciones (por ejemplo, un conjunto de porciones) determinan un nivel. En algunos casos, dos o más recuentos (por ejemplo, recuentos de dos o más porciones) determinan un nivel. En algunos casos, los recuentos de 2 a aproximadamente 100.000 porciones determinan un nivel. En algunos casos, los recuentos de 2 a aproximadamente 50.000, 2 a aproximadamente 40.000, 2 a aproximadamente 30.000, 2 a aproximadamente 20.000, 2 a aproximadamente 10.000, 2 a aproximadamente 5.000, 2 a aproximadamente 2.500, 2 a aproximadamente 1.250, 2 a aproximadamente 1.000, 2 a aproximadamente 500, 2 a aproximadamente 250, 2 a aproximadamente 100 o 2 a aproximadamente 60 porciones determinan un nivel. En algunos casos, los recuentos de aproximadamente 10 a aproximadamente 50 porciones determinan un nivel. En algunos casos, los recuentos desde aproximadamente 20 hasta aproximadamente 40 o más porciones determinan un nivel. En algunos casos, un nivel comprende recuentos de aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60 o más porciones. En algunos casos, un nivel corresponde a un conjunto de porciones (por ejemplo, un conjunto de porciones de un genoma de referencia, un conjunto de porciones de un cromosoma o un conjunto de porciones de un segmento de un cromosoma).

En algunos casos, se determina un nivel para los recuentos normalizados o no normalizados de porciones contiguas. En algunos casos, las porciones (por ejemplo, un conjunto de porciones) que son contiguas representan segmentos vecinos de un genoma o segmentos vecinos de un cromosoma o gen. Por ejemplo, dos o más porciones contiguas, cuando se alinean mediante la unión de las porciones de extremo a extremo, pueden representar un ensamblaje de secuencias de una secuencia de ADN más larga que cada porción. Por ejemplo, dos o más porciones contiguas pueden representar un genoma intacto, cromosoma, gen, intrón, exón o segmento de los mismos. En algunos casos, se determina un nivel a partir de una colección (por ejemplo, un conjunto) de porciones contiguas y/o porciones no contiguas.

Niveles Diferentes

En algunos casos, un perfil de recuentos normalizados comprende un nivel (por ejemplo, un primer nivel) significativamente diferente de otro nivel (por ejemplo, un segundo nivel) dentro del perfil. Un primer nivel puede ser más alto o más bajo que un segundo nivel. En algunos casos, un primer nivel es para un conjunto de porciones que comprenden una o más lecturas que comprenden una variación en el número de copias (por ejemplo, una variación en el número de copias maternas, una variación en el número de copias fetales o una variación en el número de copias maternas y una variación en el número de copias fetales) y el segundo el nivel es para un conjunto de porciones que comprende lecturas que no tienen sustancialmente ninguna variación en el número de copias. En algunos casos, significativamente diferente se refiere a una diferencia observable. En algunos casos, significativamente diferente se refiere a una diferencia estadísticamente diferente o estadísticamente significativa. Una diferencia estadísticamente significativa es a veces una evaluación estadística de una diferencia observada. Una diferencia estadísticamente significativa puede evaluarse mediante un método adecuado en la técnica. Se puede usar cualquier umbral o intervalo adecuado para determinar que dos niveles son significativamente diferentes. En ciertos casos, dos niveles (por ejemplo, niveles medios) que difieren en aproximadamente un 0,01 por ciento o más (por ejemplo, 0,01 por ciento de

uno o cualquiera de los valores de nivel) son significativamente diferentes. En algunos casos, dos niveles (por ejemplo, niveles medios) que difieren en aproximadamente un 0,1 por ciento o más son significativamente diferentes. En ciertos casos, dos niveles (por ejemplo, niveles medios) que difieren en aproximadamente un 0,5 por ciento o más son significativamente diferentes. En algunos casos, dos niveles (por ejemplo, niveles medios) que difieren en aproximadamente un 0,5, 0,75, 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 o más de aproximadamente un 10% son significativamente diferentes. En algunos casos, dos niveles (por ejemplo, niveles medios) son significativamente diferentes y no hay superposición en ninguno de los niveles y/o no se superponen en un intervalo definido por un valor de incertidumbre calculado para uno o ambos niveles. En ciertos casos, el valor de incertidumbre es una desviación estándar expresada como sigma. En algunos casos, dos niveles (por ejemplo, niveles medios) son significativamente diferentes y difieren en aproximadamente 1 o más veces el valor de incertidumbre (por ejemplo, 1 sigma). En algunos casos, dos niveles (por ejemplo, niveles medios) son significativamente diferentes y difieren en aproximadamente 2 o más veces el valor de incertidumbre (por ejemplo, 2 sigma), aproximadamente 3 o más, aproximadamente 4 o más, aproximadamente 5 o más, aproximadamente 6 o más, aproximadamente 7 o más, aproximadamente 8 o más, aproximadamente 9 o más, o aproximadamente 10 o más veces el valor de incertidumbre. En algunos casos, dos niveles (por ejemplo, niveles medios) son significativamente diferentes cuando difieren en aproximadamente 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 2,1, 2,2, 2,3, 2,4, 2,5, 2,6, 2,7, 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,7, 3,8, 3,9 o 4,0 veces el valor de incertidumbre o más. En algunos casos, el nivel de confianza aumenta a medida que aumenta la diferencia entre dos niveles. En ciertos casos, el nivel de confianza disminuye a medida que la diferencia entre dos niveles disminuye y/o el valor de incertidumbre aumenta. Por ejemplo, a veces el nivel de confianza aumenta con la proporción de la diferencia entre niveles y la desviación estándar (por ejemplo, MAD).

Se pueden usar uno o más algoritmos de predicción para determinar la significación o dar un significado a los datos de detección recopilados en condiciones variables que pueden ponderarse independientemente o dependientemente entre sí. El término "variable" como se usa en el presente documento se refiere a un factor, cantidad o función de un algoritmo que tiene un valor o conjunto de valores.

En algunos casos, un primer conjunto de porciones a menudo incluye porciones que son diferentes a (por ejemplo, no superpuestas con) un segundo conjunto de porciones. Por ejemplo, a veces, un primer nivel de recuentos normalizados es significativamente diferente a un segundo nivel de recuentos normalizados en un perfil, y el primer nivel es para un primer conjunto de porciones, el segundo nivel es para un segundo conjunto de porciones y las porciones no se superponen en el primer conjunto y el segundo conjunto de porciones. En ciertos casos, un primer conjunto de porciones no es un segundo conjunto de porciones a partir del cual se determinan un primer nivel y segundo nivel, respectivamente. En algunos casos, un primer conjunto de porciones es diferente y/o distinto de un segundo conjunto de porciones a partir del cual se determinan un primer nivel y un segundo nivel, respectivamente.

En algunos casos, un primer conjunto de porciones es un subconjunto de un segundo conjunto de porciones en un perfil. Por ejemplo, a veces, un segundo nivel de recuentos normalizados para un segundo conjunto de porciones en un perfil comprende recuentos normalizados de un primer conjunto de porciones para un primer nivel en el perfil y el primer conjunto de porciones es un subconjunto del segundo conjunto de porciones en el perfil. En algunos casos, un nivel promedio, medio o mediana deriva de un segundo nivel donde el segundo nivel comprende un primer nivel. En algunos casos, un segundo nivel comprende un segundo conjunto de porciones que representan un cromosoma completo y un primer nivel comprende un primer conjunto de porciones donde el primer conjunto es un subconjunto del segundo conjunto de porciones y el primer nivel representa una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales que está presente en el cromosoma.

En algunos casos, un valor de un segundo nivel está más cerca del valor medio, promedio o mediana de un perfil de recuentos para un cromosoma, o segmento del mismo, que el primer nivel. En algunos casos, un segundo nivel es un nivel medio de un cromosoma, una porción de un cromosoma o un segmento del mismo. En algunos casos, un primer nivel es significativamente diferente de un nivel predominante (por ejemplo, un segundo nivel) que representa un cromosoma, o un segmento del mismo. Un perfil puede incluir múltiples primeros niveles que difieren significativamente de un segundo nivel, y cada primer nivel independientemente puede ser más alto o más bajo que el segundo nivel. En algunos casos, un primer nivel y un segundo nivel derivan del mismo cromosoma y el primer nivel es más alto o más bajo que el segundo nivel, y el segundo nivel es el nivel predominante del cromosoma. En algunos casos, un primer nivel y un segundo nivel derivan del mismo cromosoma, un primer nivel es indicativo de una variación en el número de copias (por ejemplo, una variación, delección, inserción, duplicación del número de copias maternas y/o fetales) y un segundo nivel es un nivel medio o nivel predominante de porciones para un cromosoma, o un segmento del mismo.

En ciertos casos, una lectura en un segundo conjunto de porciones para un segundo nivel no incluye sustancialmente una variación genética (por ejemplo, una variación en el número de copias, una variación en el número de copias maternas y/o fetales). A menudo, un segundo conjunto de porciones para un segundo nivel incluye cierta variabilidad (por ejemplo, variabilidad en el nivel, variabilidad en los recuentos para las porciones). En algunos casos, una o más porciones en un conjunto de porciones para un nivel asociado con sustancialmente ninguna variación en el número de copias incluyen una o más lecturas con una variación en el número de copias presente en un genoma materno y/o fetal. Por ejemplo, a veces, un conjunto de porciones incluye una variación en el número de copias que está presente

en un pequeño segmento de un cromosoma (por ejemplo, menos de 10 porciones) y el conjunto de porciones es para un nivel asociado con ninguna variación sustancial en el número de copias. Por lo tanto, un conjunto de porciones que no incluyen sustancialmente ninguna variación en el número de copias puede incluir una variación en el número de copias que está presente en menos de aproximadamente 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 porciones de un nivel.

En algunos casos un primer nivel es para un primer conjunto de porciones y un segundo nivel es para un segundo conjunto de porciones y el primer conjunto de porciones y el segundo conjunto de porciones son contiguos (por ejemplo, adyacentes con respecto a la secuencia de ácido nucleico de un cromosoma o segmento del mismo). En algunos casos, el primer conjunto de porciones y el segundo conjunto de porciones no son contiguos.

Se pueden utilizar lecturas de secuencias relativamente cortas de una mezcla de ácidos nucleicos materno y fetal para proporcionar recuentos que pueden transformarse en un nivel y/o un perfil. Los recuentos, niveles y perfiles se pueden representar en forma electrónica o tangible y se pueden visualizar. Los recuentos mapeados en porciones (p. ej., representados como niveles y/o perfiles) pueden proporcionar una representación visual de un genoma fetal y/o materno, un cromosoma o una porción o un segmento de un cromosoma que está presente en un feto y/o gestante.

Nivel de Referencia y Valor de Referencia Normalizado

En algunos casos, un perfil comprende un nivel de referencia (por ejemplo, un nivel utilizado como referencia). A menudo, un perfil de recuentos normalizados proporciona un nivel de referencia a partir del cual se determinan los niveles esperados y los intervalos esperados (véase el análisis a continuación, sobre los niveles e intervalos esperados). Un nivel de referencia a menudo es para recuentos normalizados de porciones que comprenden lecturas mapeadas tanto de una madre como de un feto. Un nivel de referencia es a menudo la suma de los recuentos normalizados de lecturas mapeadas de un feto y una madre (por ejemplo, una gestante). En algunos casos, un nivel de referencia es para porciones que comprenden lecturas mapeadas de una madre euploide y/o un feto euploide. En algunos casos, un nivel de referencia es para porciones que comprenden lecturas mapeadas que tienen una variación genética fetal y/o materna (por ejemplo, una aneuploidía (por ejemplo, una trisomía), una variación en el número de copias, una microduplicación, una microdelección, una inserción). En algunos casos, un nivel de referencia es para porciones que no incluyen sustancialmente variaciones genéticas maternas y/o fetales (por ejemplo, una aneuploidía (por ejemplo, una trisomía), una variación en el número de copias, una microduplicación, una microdelección, una inserción). En algunos casos, un segundo nivel se utiliza como un nivel de referencia. En ciertos casos, un perfil comprende un primer nivel de recuentos normalizados y un segundo nivel de recuentos normalizados, el primer nivel es significativamente diferente del segundo nivel y el segundo nivel es el nivel de referencia. En ciertos casos, un perfil comprende un primer nivel de recuentos normalizados para un primer conjunto de porciones, un segundo nivel de recuentos normalizados para un segundo conjunto de porciones, el primer conjunto de porciones incluye lecturas mapeadas que tienen una variación en el número de copias maternas y/o fetales, el segundo conjunto de porciones comprende lecturas mapeadas que no tienen sustancialmente ninguna variación en el número de copias maternas y/o una variación en el número de copias fetales, y el segundo nivel es un nivel de referencia.

En algunos casos, los recuentos mapeados en porciones para uno o más niveles de un perfil se normalizan según los recuentos de un nivel de referencia. En algunos casos, la normalización de los recuentos de un nivel según los recuentos de un nivel de referencia comprende dividir los recuentos de un nivel por los recuentos de un nivel de referencia o un múltiplo o fracción del mismo. Los recuentos normalizados según los recuentos de un nivel de referencia a menudo se han normalizado de acuerdo con otro proceso (por ejemplo, PERUN) y los recuentos de un nivel de referencia a menudo también se han normalizado (por ejemplo, por PERUN). En algunos casos, los recuentos de un nivel se normalizan según los recuentos de un nivel de referencia y los recuentos del nivel de referencia son escalables a un valor adecuado antes de o después de la normalización. El proceso de escalar los recuentos de un nivel de referencia puede comprender cualquier constante adecuada (es decir, número) y puede aplicarse cualquier manipulación matemática adecuada a los recuentos de un nivel de referencia.

Un valor de referencia normalizado (VRN) a menudo se determina de acuerdo con los recuentos normalizados de un nivel de referencia. La determinación de un VRN puede comprender cualquier proceso de normalización adecuado (por ejemplo, manipulación matemática) aplicado a los recuentos de un nivel de referencia donde se usa el mismo proceso de normalización para normalizar los recuentos de otros niveles dentro del mismo perfil. Determinar un VRN a menudo comprende dividir un nivel de referencia por sí mismo. Determinar un VRN a menudo comprende dividir un nivel de referencia por un múltiplo de sí mismo. Determinar un VRN a menudo comprende dividir un nivel de referencia por la suma o diferencia del nivel de referencia y una constante (por ejemplo, cualquier número).

Un VRN a veces se denomina valor nulo. Un VRN puede ser cualquier valor adecuado. En algunos casos, un VRN es cualquier valor distinto de cero. En algunos casos, un VRN es un número entero. En algunos casos, un VRN es un número entero positivo. En algunos casos, un VRN es 1, 10, 100 o 1000. A menudo, un VRN es igual a 1. En algunos casos, un VRN es igual a cero. Los recuentos de un nivel de referencia se pueden normalizar a cualquier VRN adecuado. En algunos casos, los recuentos de un nivel de referencia se normalizan a un VRN de cero. A menudo, los recuentos de un nivel de referencia se normalizan a un VRN de 1.

Niveles Esperados

Un nivel esperado es a veces un nivel predefinido (por ejemplo, un nivel teórico, nivel previsto). Un "nivel esperado" se denomina a veces en el presente documento, un "valor de nivel predeterminado". En algunos casos, un nivel esperado es un valor previsto para un nivel de recuentos normalizados para un conjunto de porciones que incluyen una variación en el número de copias. En ciertos casos, se determina un nivel esperado para un conjunto de porciones que no incluyen sustancialmente ninguna variación en el número de copias. Se puede determinar un nivel esperado para una ploidía cromosómica (por ejemplo, 0, 1, 2 (es decir, diploides), 3 o 4 cromosomas) o una microploidía (deleción, duplicación, inserción o ausencia de homocigotos o heterocigotos). A menudo, se determina un nivel esperado para una microploidía materna (por ejemplo, una variación en el número de copias maternas y/o fetales).

Un nivel esperado para una variación genética o una variación en el número de copias se puede determinar de cualquier manera adecuada. A menudo, un nivel esperado se determina mediante una manipulación matemática adecuada de un nivel (por ejemplo, recuentos mapeados en un conjunto de porciones para un nivel). En algunos casos, un nivel esperado se determina utilizando una constante que a veces se denomina constante de nivel esperado. Un nivel esperado para una variación en el número de copias a veces se calcula multiplicando un nivel de referencia, recuentos normalizados de un nivel de referencia o un VRN por una constante de nivel esperado, sumando una constante de nivel esperado, restando una constante de nivel esperado, dividiendo por una constante de nivel esperado, o por una combinación de los mismos. A menudo, un nivel esperado (por ejemplo, un nivel esperado de una variación en el número de copias maternas y/o fetales) determinado para el mismo sujeto, grupo de muestra o de ensayo se determina de acuerdo con el mismo nivel de referencia o VRN.

A menudo, un nivel esperado se determina multiplicando un nivel de referencia, recuentos normalizados de un nivel de referencia o un VRN por una constante de nivel esperado donde el nivel de referencia, los recuentos normalizados de un nivel de referencia o VNR no es igual a cero. En algunos casos, un nivel esperado se determina sumando una constante de nivel esperado al nivel de referencia, recuentos normalizados de un nivel de referencia o a un VRN que es igual a cero. En algunos casos, un nivel esperado, recuentos normalizados de un nivel de referencia, VRN y la constante de nivel esperado son escalables. El proceso de escalado puede comprender cualquier constante adecuada (es decir, número) y cualquier manipulación matemática adecuada donde se aplique el mismo proceso de escalado a todos los valores en estudio.

Constante de Nivel Esperado

Una constante de nivel esperado puede determinarse por un método adecuado. En algunos casos, una constante de nivel esperado se determina arbitrariamente. A menudo, una constante de nivel esperado se determina empíricamente. En algunos casos, una constante de nivel esperado se determina de acuerdo con una manipulación matemática. En algunos casos, una constante de nivel esperado se determina de acuerdo con una referencia (p. ej., un genoma de referencia, una muestra de referencia, datos de ensayo de referencia). En algunos casos, una constante de nivel esperado se predetermina para un nivel representativo de la presencia o ausencia de una variación genética o una variación en el número de copias (por ejemplo, una duplicación, inserción o deleción). En algunos casos, una constante de nivel esperado se predetermina para un nivel representativo de la presencia o ausencia de una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales. Una constante de nivel esperado para una variación en el número de copias puede ser cualquier constante o conjunto de constantes adecuadas.

En algunos casos, la constante de nivel esperado para una duplicación homocigótica (por ejemplo, una duplicación homocigótica) puede ser de aproximadamente 1,6 a aproximadamente 2,4, de aproximadamente 1,7 a aproximadamente 2,3, de aproximadamente 1,8 a aproximadamente 2,2, o de aproximadamente 1,9 a aproximadamente 2,1. En algunos casos, la constante de nivel esperado para una duplicación homocigótica es de aproximadamente 1,6, 1,7, 1,8, 1,9, 2,0, 2,1, 2,2, 2,3 o aproximadamente 2,4. A menudo, la constante de nivel esperado para una duplicación homocigótica es aproximadamente 1,90, 1,92, 1,94, 1,96, 1,98, 2,0, 2,02, 2,04, 2,06, 2,08 o aproximadamente 2,10. A menudo, la constante de nivel esperado para una duplicación homocigótica es aproximadamente 2.

En algunos casos, la constante de nivel esperado para una duplicación heterocigótica (por ejemplo, una duplicación homocigótica) es de aproximadamente 1,2 a aproximadamente 1,8, de aproximadamente 1,3 a aproximadamente 1,7, o de aproximadamente 1,4 a aproximadamente 1,6. En algunos casos, la constante de nivel esperado para una duplicación heterocigótica es de aproximadamente 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, o aproximadamente 1,8. A menudo, la constante de nivel esperado para una duplicación heterocigótica es aproximadamente 1,40, 1,42, 1,44, 1,46, 1,48, 1,5, 1,52, 1,54, 1,56, 1,58 o aproximadamente 1,60. En algunos casos, la constante de nivel esperado para una duplicación heterocigótica es de aproximadamente 1,5.

En algunos casos, la constante de nivel esperado para la ausencia de una variación en el número de copias (por ejemplo, la ausencia de una variación en el número de copias maternas y/o variación en el número de copias fetales) es de aproximadamente 1,3 a aproximadamente 0,7, de aproximadamente 1,2 a aproximadamente 0,8, o de aproximadamente 1,1 a aproximadamente 0,9. En algunos casos, la constante de nivel esperado para la ausencia de

una variación en el número de copias es de aproximadamente 1,3, 1,2, 1,1, 1,0, 0,9, 0,8, o aproximadamente 0,7. A menudo, la constante de nivel esperado para la ausencia de una variación en el número de copias es aproximadamente 1,09, 1,08, 1,06, 1,04, 1,02, 1,0, 0,98, 0,96, 0,94, o aproximadamente 0,92. En algunos casos, la constante de nivel esperado para la ausencia de una variación en el número de copias es de aproximadamente 1.

5 La constante de nivel esperado para una delección heterocigótica (por ejemplo, una delección heterocigótica materna, fetal o materna y fetal) puede ser de aproximadamente 0,2 a aproximadamente 0,8, de aproximadamente 0,3 a aproximadamente 0,7, o de aproximadamente 0,4 a aproximadamente 0,6. La constante de nivel esperado para una delección heterocigótica puede ser aproximadamente 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, o aproximadamente 0,8. A menudo, la constante de nivel esperado para una delección heterocigótica es aproximadamente 0,40, 0,42, 0,44, 0,46, 0,48, 0,5, 10 0,52, 0,54, 0,56, 0,58 o aproximadamente 0,60. La constante de nivel esperado para una delección heterocigótica puede ser aproximadamente 0,5.

15 En algunos casos, la constante de nivel esperado para una delección homocigótica (por ejemplo, una delección homocigótica) es de aproximadamente -0,4 a aproximadamente 0,4, de aproximadamente -0,3 a aproximadamente 0,3, de aproximadamente -0,2 a aproximadamente 0,2, o de aproximadamente -0,1 a aproximadamente 0,1. EN algunos casos, la constante de nivel esperado para una delección homocigótica es de aproximadamente -0,4, -0,3, -0,2, -0,1, 0,0, 0,1, 0,2, 0,3, o aproximadamente 0,4. A menudo, la constante de nivel esperado para una delección homocigótica es de aproximadamente -0,1, -0,08, -0,06, -0,04, -0,02, 0,0, 0,02, 0,04, 0,06, 0,08, o de aproximadamente 0,10. A menudo, la constante de nivel esperado para una delección homocigótica es de aproximadamente 0.

Intervalo de Nivel Esperado

25 En algunos casos, la presencia o ausencia de una variación genética o una variación en el número de copias (por ejemplo, una variación en el número de copias maternas, una variación en el número de copias fetales o una variación en el número de copias maternas y una variación en el número de copias fetales) se determina por un nivel que se encuentra dentro o fuera de un intervalo de nivel esperado. Un intervalo de nivel esperado a menudo se determina de acuerdo con un nivel esperado. En algunos casos, se determina un intervalo de nivel esperado para un nivel que no comprende sustancialmente ninguna variación genética o sustancialmente ninguna variación en el número de copias. 30 Se puede usar un método adecuado para determinar un intervalo de nivel esperado.

En algunos casos, un intervalo de nivel esperado se define de acuerdo con un valor de incertidumbre adecuado calculado para un nivel. Ejemplos no limitantes de un valor de incertidumbre son una desviación estándar, error estándar, varianza calculada, valor de p o desviación media absoluta (MAD). En algunos casos, un intervalo de nivel 35 esperado para una variación genética o una variación en el número de copias se determina, en parte, calculando el nivel de incertidumbre para un nivel (p. ej., un primer nivel, un segundo nivel, un primer nivel y un segundo nivel). En algunos casos, un intervalo de nivel esperado se define de acuerdo con un valor de incertidumbre calculado para un perfil (por ejemplo, un perfil de recuentos normalizados para un cromosoma o segmento del mismo). En algunos casos, se calcula un valor de incertidumbre para un nivel que no comprende sustancialmente ninguna variación genética o 40 sustancialmente ninguna variación en el número de copias. En algunos casos, se calcula un valor de incertidumbre para un primer nivel, un segundo nivel o un primer nivel y un segundo nivel. En algunos casos, se determina un valor de incertidumbre para un primer nivel, un segundo nivel o un segundo nivel que comprende un primer nivel.

45 Un intervalo de nivel esperado a veces se calcula, en parte, multiplicando, sumando, restando o dividiendo un valor de incertidumbre por una constante n (por ejemplo, una constante predeterminada). Se puede utilizar un procedimiento o una combinación de procedimientos matemáticos adecuados. La constante n (por ejemplo, la constante predeterminada n) a veces se denomina intervalo de confianza. Un intervalo de confianza seleccionado se determina de acuerdo con la constante n que se selecciona. La constante n (por ejemplo, la constante predeterminada n , el intervalo de confianza) se puede determinar de una manera adecuada. La constante n puede ser un número o fracción 50 de un número mayor que cero. La constante n puede ser un número entero. A menudo, la constante n es un número menor que 10. En algunos casos, la constante n es un número menor que aproximadamente 10, menor que aproximadamente 9, menor que aproximadamente 8, menor que aproximadamente 7, menor que aproximadamente 6, menor que aproximadamente 5, menor que aproximadamente 4, menor que aproximadamente 3 o menor que aproximadamente 2. En algunos casos, la constante n es de aproximadamente 10, 9,5, 9, 8,5, 8, 7,5, 7, 6,5, 6, 5,5, 5, 55 4,5, 4, 3,5, 3, 2,5, 2 o 1. La constante n puede determinarse empíricamente a partir de datos derivados de sujetos (una gestante y/o un feto) con una disposición genética conocida.

A menudo, un valor de incertidumbre y una constante n definen un intervalo (por ejemplo, un límite de incertidumbre). Por ejemplo, a veces un valor de incertidumbre es una desviación estándar (por ejemplo, ± 5) y se multiplica por una 60 constante n (por ejemplo, un intervalo de confianza) definiendo, de este modo, un límite de intervalo o incertidumbre (por ejemplo, $5n$ a $-5n$).

En algunos casos, un intervalo de nivel esperado para una variación genética (por ejemplo, una variación en el número de copias maternas, variación en el número de copias fetales o una variación en el número de copias maternas y una 65 variación en el número de copias fetales) es la suma de un nivel esperado más una constante n veces la incertidumbre (por ejemplo, $n \times \text{sigma}$ (por ejemplo, 6 sigma)). En algunos casos, el intervalo de nivel esperado para una variación

genética o variación en el número de copias designado por k se puede definir mediante la fórmula:

Fórmula R: $(\text{Intervalo de Nivel Esperado})_k = (\text{Nivel Esperado})_k + n\sigma$

5 donde σ es un valor de incertidumbre, n es una constante (por ejemplo, una constante predeterminada) y el intervalo de nivel esperado y el nivel esperado son para la variación genética k (por ejemplo, $k =$ una delección heterocigótica, por ejemplo, $k =$ la ausencia de una variación genética). Por ejemplo, para un nivel esperado igual a 1 (por ejemplo, la ausencia de una variación en el número de copias), un valor de incertidumbre (es decir, σ) igual a $\pm 0,05$, y $n = 3$, el intervalo de nivel esperado se define como 1,15 a 0,85. En algunos casos, el intervalo de nivel esperado para una duplicación heterocigótica se determina entre 1,65 y 1,35 cuando el nivel esperado para una duplicación heterocigótica es 1,5, $n = 3$ y el valor de incertidumbre σ es $\pm 0,05$. En algunos casos, el intervalo de nivel esperado para una delección heterocigótica se determinarse como de 0,65 a 0,35 cuando el nivel esperado para una duplicación heterocigótica es 0,5, $n = 3$ y el valor de incertidumbre σ es $\pm 0,05$. En algunos casos, el intervalo de nivel esperado para una duplicación homocigótica se determina entre 2,15 y 1,85 cuando el nivel esperado para una duplicación heterocigótica es 2,0, $n = 3$ y el valor de incertidumbre σ es $\pm 0,05$.

En algunos casos, el intervalo de nivel esperado para una delección homocigótica se determina como de 0,15 a -0,15 cuando el nivel esperado para una duplicación heterocigótica es 0,0, $n = 3$ y el valor de incertidumbre σ es $\pm 0,05$.

20 En algunos casos, un intervalo de nivel esperado para una variación en el número de copias homocigóticas (por ejemplo, una variación en el número de copias homocigótica maternas, fetales o maternas y fetales) se determina, en parte, de acuerdo con un intervalo de nivel esperado para una variación en el número de copias heterocigótica correspondiente. Por ejemplo, a veces, un intervalo de nivel esperado para una duplicación homocigótica comprende todos los valores mayores que un límite superior de un intervalo de nivel esperado para una duplicación heterocigótica.

25 En algunos casos, un intervalo de nivel esperado para una duplicación homocigótica comprende todos los valores mayores o iguales a un límite superior de un intervalo de nivel esperado para una duplicación heterocigótica. En algunos casos, un intervalo de nivel esperado para una duplicación homocigótica comprende todos los valores mayores que un límite superior de un intervalo de nivel esperado para una duplicación heterocigótica y menores que el límite superior definido por la fórmula R donde σ es un valor de incertidumbre y es un valor positivo, n es una constante y k es una duplicación homocigótica. En algunos casos, un intervalo de nivel esperado para una duplicación homocigótica comprende todos los valores mayores o iguales a un límite superior de un intervalo de nivel esperado para una duplicación heterocigótica y menores o iguales al límite superior definido por la fórmula R, donde σ es un valor de incertidumbre, σ es un valor positivo, n es una constante y k es una duplicación homocigótica.

35 Un intervalo de nivel esperado para una delección homocigótica puede comprender todos los valores inferiores a un límite inferior de un intervalo de nivel esperado para una delección heterocigótica. Un intervalo de nivel esperado para una delección homocigótica puede comprender todos los valores menores que o iguales a un límite inferior de un intervalo de nivel esperado para una delección heterocigótica. Un intervalo de nivel esperado para una delección homocigótica puede comprender todos los valores menores que un límite inferior de un intervalo de nivel esperado para una delección heterocigótica y mayores que el límite inferior definido por la fórmula R, donde σ es un valor de incertidumbre, σ es un valor negativo, n es una constante y k es una delección homocigótica. Un intervalo de nivel esperado para una delección homocigótica puede comprender todos los valores menores que o iguales a un límite inferior de un intervalo de nivel esperado para una delección heterocigótica y mayores que o iguales al límite inferior definido por la fórmula R, donde σ es un valor de incertidumbre, σ es un valor negativo, n es una constante y k es una delección homocigótica.

Se puede utilizar un valor de incertidumbre para determinar un valor umbral. En algunos casos, se obtiene un intervalo (por ejemplo, un intervalo umbral) calculando el valor de incertidumbre determinado a partir de recuentos sin procesar, filtrados y/o normalizados. Se puede determinar un intervalo multiplicando el valor de incertidumbre para un nivel (por ejemplo, recuentos normalizados de un nivel) por una constante predeterminada (por ejemplo, 1, 2, 3, 4, 5, 6, etc.) que representa el múltiplo de incertidumbre (por ejemplo, número de desviaciones estándar) elegidas como un umbral límite (por ejemplo, multiplicar por 3 para 3 desviaciones estándar), por lo que, en algunos casos, se genera un intervalo. Se puede determinar un intervalo sumando y/o restando un valor (por ejemplo, un valor predeterminado, un valor de incertidumbre, un valor de incertidumbre multiplicado por una constante predeterminada) a y/o desde un nivel por el cual, en algunos casos, se genera un intervalo. Por ejemplo, para un nivel igual a 1, una desviación estándar de $\pm 0,2$, donde una constante predeterminada es 3, el intervalo se puede calcular como $(1 + 3(0,2))$ a $(1 + 3(-0,2))$, o 1,6 a 0,4. Un intervalo a veces puede definir un intervalo esperado o un intervalo de nivel esperado para una variación en el número de copias. En ciertos casos, algunas o todas las porciones que exceden un valor umbral, que se encuentran fuera de un intervalo o que se encuentran dentro de un intervalo de valores, se eliminan como parte de, antes o después de un proceso de normalización. En algunos casos, algunas o todas las porciones que exceden un valor umbral, que se encuentran fuera de un intervalo o que se encuentran dentro de un intervalo se ponderan o ajustan como parte de, o antes del proceso de normalización o clasificación. En el presente documento, se describen ejemplos de ponderación. Las expresiones "datos redundantes" y "lecturas mapeadas redundantes", como se usan en el presente documento, se refieren a lecturas de secuencias derivadas de muestras que se han identificado como asignadas a una ubicación genómica (por ejemplo, posición base) y/o contabilizadas para una porción.

En algunos casos, se determina un valor de incertidumbre de acuerdo con la siguiente fórmula:

$$Z = \frac{L_A - L_O}{\sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_O^2}{N_O}}}$$

- 5 Donde Z representa la desviación estandarizada entre dos niveles, L es el nivel medio (o mediana) y sigma es la desviación estándar (o MAD). El subíndice O denota un segmento de un perfil (por ejemplo, un segundo nivel, un cromosoma, un VRN, un "nivel euploide", un nivel sin una variación en el número de copias), y A denota otro segmento de un perfil (por ejemplo, un primer nivel, un nivel que representa una variación en el número de copias, un nivel que representa una aneuploidía (por ejemplo, una trisomía). La variable N_o representa el número total de porciones en el segmento del perfil indicado por el subíndice O. N_A representa el número total de porciones en el segmento del perfil indicado por el subíndice A.

Clasificación de una Variación en el Número de Copias

- 15 Un nivel (por ejemplo, un primer nivel) que difiere significativamente de otro nivel (por ejemplo, un segundo nivel) a menudo se puede clasificar como una variación en el número de copias (por ejemplo, una variación en el número de copias maternas y/o fetales, una variación en el número de copias fetales, una delección, duplicación, inserción) de acuerdo con un intervalo de nivel esperado. En algunos casos, la presencia de una variación en el número de copias se clasifica cuando un primer nivel es significativamente diferente de un segundo nivel y el primer nivel se encuentra dentro del intervalo del nivel esperado para una variación en el número de copias. Por ejemplo, una variación en el número de copias (p. ej., una variación en el número de copias maternas y/o fetales, una variación en el número de copias fetales) se puede clasificar cuando un primer nivel es significativamente diferente de un segundo nivel y el primer nivel se encuentra dentro del intervalo de nivel esperado para una variación en el número de copias. En algunos casos, una duplicación heterocigótica (por ejemplo, una duplicación heterocigótica materna o fetal, o materna y fetal) o una delección heterocigótica (por ejemplo, una delección heterocigótica materna o fetal, o materna y fetal) se clasifica cuando un primer nivel es significativamente diferente de un segundo nivel y el primer nivel se encuentra dentro del intervalo de nivel esperado para una duplicación heterocigótica o una delección heterocigótica, respectivamente. En algunos casos, una duplicación homocigótica o una delección homocigótica puede clasificarse cuando un primer nivel es significativamente diferente de un segundo nivel y el primer nivel se encuentra dentro del intervalo de nivel esperado para una duplicación homocigótica o una delección homocigótica, respectivamente.

Ajustes de nivel

- En algunos casos, se ajustan uno o más niveles. Un proceso para ajustar un nivel a menudo se conoce como *padding*. En algunos casos, se ajustan múltiples niveles en un perfil (por ejemplo, un perfil de un genoma, un perfil cromosómico, un perfil de una porción o segmento de un cromosoma). En algunos casos, se ajustan de aproximadamente 1 a aproximadamente 10.000 o más niveles en un perfil. En algunos casos, se ajustan aproximadamente 1 a aproximadamente 1000, 1 a aproximadamente 900, 1 a aproximadamente 800, 1 a aproximadamente 700, 1 a aproximadamente 600, 1 a aproximadamente 500, 1 a aproximadamente 400, 1 a aproximadamente 300, 1 a aproximadamente 200, 1 a aproximadamente 100, 1 a aproximadamente 50, 1 a aproximadamente 25, 1 a aproximadamente 20, 1 a aproximadamente 15, 1 a aproximadamente 10, o 1 a aproximadamente 5 niveles en un perfil. En algunos casos se ajusta un nivel. En algunos casos, se ajusta un nivel (por ejemplo, un primer nivel de un perfil de recuento normalizado) que difiera significativamente de un segundo nivel. En algunos casos, se ajusta un nivel clasificado como una variación en el número de copias. En algunos casos se clasifica un nivel (por ejemplo, un primer nivel de un perfil de recuento normalizado) que difiere significativamente de un segundo nivel como una variación en el número de copias (por ejemplo, una variación en el número de copias, por ejemplo, una variación en el número de copias maternas) y se ajusta. En algunos casos, un nivel (por ejemplo, un primer nivel) está dentro de un intervalo de nivel esperado para una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales y el nivel se ajusta. En algunos casos, no se ajustan uno o más niveles (por ejemplo, niveles en un perfil). En algunos casos, un nivel (por ejemplo, un primer nivel) está fuera de un intervalo de nivel esperado para una variación en el número de copias y el nivel no se ajusta. A menudo, no se ajusta un nivel dentro de un intervalo de nivel esperado para la ausencia de una variación en el número de copias. Se puede realizar cualquier número adecuado de ajustes en uno o más niveles en un perfil. En algunos casos, se ajustan uno o más niveles. En algunos casos, se ajustan 2 o más, 3 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más y, a veces, 10 o más niveles.

- En algunos casos, se ajusta un valor de un primer nivel de acuerdo con un valor de un segundo nivel. En algunos casos, un primer nivel, identificado como representativo de una variación en el número de copias, se ajusta al valor de un segundo nivel, donde el segundo nivel a menudo se asocia con ninguna variación en el número de copias. En ciertos casos, un valor de un primer nivel, identificado como representativo de una variación en el número de copias, se ajusta para que el valor del primer nivel sea aproximadamente igual al valor de un segundo nivel.

Un ajuste puede comprender una operación matemática adecuada. En algunos casos, un ajuste comprende una o

más operaciones matemáticas. En algunos casos, un nivel se ajusta por normalización, filtrado, promediación, multiplicación, división, suma o resta o combinación de los mismos. En algunos casos, se ajusta un nivel por un valor predeterminado o una constante. En algunos casos, un nivel se ajusta modificando el valor del nivel al valor de otro nivel.). Por ejemplo, un primer nivel puede ajustarse modificando su valor al valor de un segundo nivel. Un valor en tales casos puede ser un valor procesado (por ejemplo, media, valor normalizado y similares).

En algunos casos, un nivel se clasifica como una variación en el número de copias (por ejemplo, una variación en el número de copias maternas) y se ajusta de acuerdo con un valor predeterminado al que se denomina en el presente documento valor de ajuste predeterminado (VAP). A menudo, un VAP se determina para una variación específica en el número de copias. A menudo, se utiliza un VAP determinado para una variación específica en el número de copias (por ejemplo, duplicación homocigótica, deleción homocigótica, duplicación heterocigótica, eliminación heterocigótica) para ajustar un nivel clasificado como una variación específica en el número de copias (por ejemplo, duplicación homocigótica, deleción homocigótica, duplicación heterocigótica, deleción heterocigótica). En ciertos casos, un nivel se clasifica como una variación en el número de copias y a continuación se ajusta de acuerdo con un VAP específico para el tipo de variación en el número de copias clasificado. En algunos casos, un nivel (por ejemplo, un primer nivel) se clasifica como una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales y se ajusta sumando o restando un VAP del nivel. A menudo, un nivel (por ejemplo, un primer nivel) se clasifica como una variación en el número de copias maternas y se ajusta agregando un VAP al nivel. Por ejemplo, un nivel clasificado como una duplicación (por ejemplo, una duplicación homocigótica materna, fetal o materna y fetal) se puede ajustar agregando un VAP determinado para una duplicación específica (por ejemplo, una duplicación homocigótica) proporcionando, de este modo, un nivel ajustado. A menudo, un VAP determinado para una duplicación de número de copias es un valor negativo. En algunos casos, el proporcionar un ajuste a un nivel representativo de una duplicación utilizando un VAP determinado para una duplicación da como resultado una reducción en el valor del nivel. En algunos casos, un nivel (por ejemplo, un primer nivel) que difiere significativamente de un segundo nivel se clasifica como una deleción de número de copias (por ejemplo, una deleción homocigótica, deleción heterocigótica, duplicación homocigótica, duplicación homocigótica) y el primer nivel se ajusta agregando un VAP determinado para una deleción de número de copias. A menudo, un VAP determinado para una deleción de un número de copias es un valor positivo. En algunos casos, el proporcionar un ajuste a un nivel representativo de una deleción utilizando un VAP determinado para una deleción da como resultado un aumento en el valor del nivel.

Un VAP puede ser cualquier valor adecuado. A menudo, un VAP se determina de acuerdo con y es específico para una variación en el número de copias (por ejemplo, una variación en el número de copias clasificada). En ciertos casos, un VAP se determina de acuerdo con un nivel esperado para una variación en el número de copias (por ejemplo, una variación en el número de copias clasificada) y/o un factor VAP. Un VAP a veces se determina multiplicando un nivel esperado por un factor VAP. Por ejemplo, un VAP para una variación en el número de copias puede determinarse multiplicando un nivel esperado determinado para una variación en el número de copias (por ejemplo, una deleción heterocigótica) por un factor VAP determinado para la misma variación en el número de copias (por ejemplo, una deleción heterocigótica). Por ejemplo, El VAP se puede determinar mediante la siguiente fórmula:

$$\text{VAP}_k = (\text{Nivel Esperado})_k \times (\text{factor VAP})_k$$

para la variación en el número de copias k (por ejemplo, k = una deleción heterocigótica)

Un factor VAP puede ser cualquier valor adecuado. En algunos casos, un factor VAP para una duplicación homocigótica está entre aproximadamente -0,6 y aproximadamente -0,4. En algunos casos, un factor VAP para una duplicación homocigótica es de aproximadamente -0,60, -0,59, -0,58, -0,57, -0,56, -0,55, -0,54, -0,53, -0,52, -0,51, -0,50, -0,49, -0,48, -0,47, -0,46, -0,45, -0,44, -0,43, -0,42, -0,41 y -0,40. A menudo, un factor VAP para una duplicación homocigótica es aproximadamente -0,5.

Por ejemplo, para un VRN de aproximadamente 1 y un nivel esperado de una duplicación homocigótica igual a aproximadamente 2, el VAP para la duplicación homocigótica se determina como aproximadamente -1 de acuerdo con la fórmula anterior. En este caso, un primer nivel clasificado como una duplicación homocigótica se ajusta agregando aproximadamente -1 al valor del primer nivel, por ejemplo.

En algunos casos, un factor VAP para una duplicación heterocigótica está entre aproximadamente -0,4 y aproximadamente -0,2. En algunos casos, un factor VAP para una duplicación heterocigótica es de aproximadamente -0,40, -0,39, -0,38, -0,37, -0,36, -0,35, -0,34, -0,33, -0,32, -0,31, -0,30, -0,29, -0,28, -0,27, -0,26, -0,25, -0,24, -0,23, -0,22, -0,21 y -0,20. A menudo, un factor VAP para una duplicación heterocigótica es aproximadamente -0,33.

Por ejemplo, para un VRN de aproximadamente 1 y un nivel esperado de una duplicación heterocigótica igual a aproximadamente 1,5, el VAP para la duplicación homocigótica se determina como aproximadamente -0,495 de acuerdo con la fórmula anterior. En este caso, un primer nivel clasificado como una duplicación heterocigótica se ajusta agregando aproximadamente -0,495 al valor del primer nivel, por ejemplo.

En algunos casos, un factor VAP para una deleción heterocigótica está entre aproximadamente 0,4 y

aproximadamente 0,2. En algunos casos, un factor VAP para una deleción heterocigótica es de aproximadamente 0,40, 0,39, 0,38, 0,37, 0,36, 0,35, 0,34, 0,33, 0,32, 0,31, 0,30, 0,29, 0,28, 0,27, 0,26, 0,25, 0,24, 0,23, 0,22, 0,21 y 0,20. A menudo, un factor VAP para una deleción heterocigótica es aproximadamente 0,33.

- 5 Por ejemplo, para un VRN de aproximadamente 1 y un nivel esperado de una deleción heterocigótica igual a aproximadamente 0,5, el VAP para la deleción heterocigótica se determina como aproximadamente 0,495 de acuerdo con la fórmula anterior. En este caso, un primer nivel clasificado como una deleción heterocigótica se ajusta agregando aproximadamente 0,495 al valor del primer nivel, por ejemplo.
- 10 En algunos casos, un factor VAP para una deleción homocigótica está entre aproximadamente 0,6 y aproximadamente 0,4. En algunos casos, un factor VAP para una deleción homocigótica es de aproximadamente 0,60, 0,59, 0,58, 0,57, 0,56, 0,55, 0,54, 0,53, 0,52, 0,51, 0,50, 0,49, 0,48, 0,47, 0,46, 0,45, 0,44, 0,43, 0,42, 0,41 y 0,40. A menudo, un factor VAP para una deleción homocigótica es aproximadamente 0,5.
- 15 Por ejemplo, para un VRN de aproximadamente 1 y un nivel esperado de una deleción homocigótica igual a aproximadamente 0, el VAP para la deleción homocigótica se determina como aproximadamente 1 de acuerdo con la fórmula anterior. En este caso, un primer nivel clasificado como una deleción homocigótica se ajusta agregando aproximadamente 1 al valor del primer nivel, por ejemplo.
- 20 En ciertos casos, un VAP es aproximadamente igual a o igual a un nivel esperado para una variación en el número de copias (por ejemplo, el nivel esperado de una variación en el número de copias).

En algunos casos, los recuentos de un nivel se normalizan antes de hacer un ajuste. En ciertos casos, los recuentos de algunos o todos los niveles en un perfil se normalizan antes de realizar un ajuste. Por ejemplo, los recuentos de un nivel pueden normalizarse de acuerdo con los recuentos de un nivel de referencia o un VRN. En ciertos casos, los recuentos de un nivel (por ejemplo, un segundo nivel) se normalizan según los recuentos de un nivel de referencia o un VRN y los recuentos de los otros niveles (por ejemplo, un primer nivel) en un perfil se normalizan en relación con los recuentos del mismo nivel de referencia o VRN antes de realizar un ajuste.

30 En algunos casos, un nivel de un perfil es el resultado de uno o más ajustes. En ciertos casos, el nivel de un perfil se determina después de que se ajusten uno o más niveles en el perfil. En algunos casos, un nivel de un perfil se recalcula después de que se realicen uno o más ajustes.

En algunos casos, una variación en el número de copias (por ejemplo, una variación en el número de copias maternas, variación en el número de copias fetales o una variación en el número de copias maternas y una variación en el número de copias fetales) se determina (por ejemplo, directa o indirectamente) a partir de un ajuste. Por ejemplo, un nivel en un perfil que se ajustó (por ejemplo, un primer nivel ajustado) se puede identificar como una variación en el número de copias maternas. En algunos casos, la magnitud del ajuste indica el tipo de variación en el número de copias (por ejemplo, deleción heterocigótica, duplicación homocigótica y similares). En ciertos casos, un nivel ajustado en un perfil puede identificarse como representativo de una variación en el número de copias de acuerdo con el valor de un VAP para la variación en el número de copias. Por ejemplo, para un perfil dado, VAP es aproximadamente -1 para una duplicación homocigótica, aproximadamente -0,5 para una duplicación heterocigótica, aproximadamente 0,5 para una deleción heterocigótica y aproximadamente 1 para una deleción homocigótica. En el ejemplo anterior, un nivel ajustado por aproximadamente -1 puede identificarse como una duplicación homocigótica, por ejemplo. En algunos casos, se pueden determinar una o más variaciones en el número de copias a partir de un perfil o un nivel que comprende uno o más ajustes.

En ciertos casos, se comparan los niveles ajustados dentro de un perfil. En algunos casos, se identifican anomalías y errores comparando niveles ajustados. Por ejemplo, a menudo se comparan uno o más niveles ajustados en un perfil y un nivel particular puede identificarse como una anomalía o error. En algunos casos, se identifica una anomalía o error dentro de una o más porciones que forman un nivel. Se puede identificar una anomalía o error dentro del mismo nivel (por ejemplo, en un perfil) o en uno o más niveles que representan porciones adyacentes, contiguas, adjuntas o colindantes. En algunos casos, uno o más niveles ajustados son niveles de porciones que son adyacentes, contiguas, adjuntas o colindantes donde se comparan el uno o más niveles ajustados y se identifica una anomalía o error. Una anomalía o error puede ser un máximo o mínimo en un perfil o nivel donde se conoce o se desconoce la causa del máximo o mínimo. En ciertos casos, los niveles ajustados se comparan y se identifica una anomalía o error cuando la anomalía o error se debe a un error estocástico, sistemático, aleatorio o de usuario. En algunos casos, los niveles ajustados se comparan y se elimina una anomalía o error de un perfil. En ciertos casos, los niveles ajustados se comparan y se ajusta una anomalía o error.

60 *Determinación de la Fracción Fetal Basada en el Nivel*

En algunos casos, se determina una fracción fetal de acuerdo con un nivel clasificado como representativo de una variación en el número de copias maternas y/o fetales. Por ejemplo, la determinación de la fracción fetal a menudo comprende evaluar un nivel esperado para una variación en el número de copias maternas y/o fetales utilizada para la determinación de la fracción fetal. En algunos casos, se determina una fracción fetal para un nivel (por ejemplo, un

5 primer nivel) clasificado como representativo de una variación en el número de copias de acuerdo con un intervalo de nivel esperado determinado para el mismo tipo de variación en el número de copias. A menudo, una fracción fetal se determina de acuerdo con un nivel observado que se encuentra dentro de un intervalo de nivel esperado y, por lo tanto, se clasifica como una variación en el número de copias maternas y/o fetales. En algunos casos, una fracción fetal se determina cuando un nivel observado (por ejemplo, un primer nivel) clasificado como una variación en el número de copias maternas y/o fetales es diferente del nivel esperado determinado para la misma variación en el número de copias maternas y/o fetales.

10 En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) es significativamente diferente de un segundo nivel, el primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, un primer nivel es un nivel observado y/u obtenido experimentalmente que es significativamente diferente de un segundo nivel en un perfil y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, el primer nivel es un nivel promedio, medio o sumado y se determina una fracción fetal de acuerdo con el primer nivel. En ciertos casos, se observa un primer nivel y un segundo nivel y/o niveles obtenidos experimentalmente y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, un primer nivel comprende recuentos normalizados para un primer conjunto de porciones y un segundo nivel comprende recuentos normalizados para un segundo conjunto de porciones y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, un primer conjunto de porciones de un primer nivel incluye una variación en el número de copias (por ejemplo, el primer nivel es representativo de una variación en el número de copias) y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, el primer conjunto de porciones de un primer nivel incluye una variación en el número de copias maternas homocigótica o heterocigótica y se determina una fracción fetal de acuerdo con el primer nivel. En algunos casos, un perfil comprende un primer nivel para un primer conjunto de porciones y un segundo nivel para un segundo conjunto de porciones, el segundo conjunto de porciones no incluye prácticamente ninguna variación en el número de copias (por ejemplo, una variación en el número de copias maternas, una variación en el número de copias fetales, o una variación en el número de copias maternas y una variación en el número de copias fetales) y se determina una fracción fetal de acuerdo con el primer nivel.

30 En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) es significativamente diferente de un segundo nivel, el primer nivel se clasificar como para una variación en el número de copias maternas y/o fetales, y se determina una fracción fetal de acuerdo con el primer nivel y/o un nivel esperado de la variación en el número de copias. En algunos casos, un primer nivel se clasifica como para una variación en el número de copias de acuerdo con un nivel esperado para una variación en el número de copias y se determina una fracción fetal de acuerdo con una diferencia entre el primer nivel y el nivel esperado. En ciertos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) se clasifica como una variación en el número de copias maternas y/o fetales, y se determina una fracción fetal como el doble de la diferencia entre el primer nivel y el nivel esperado de la variación en el número de copias. En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) se clasifica como una variación en el número de copias maternas y/o fetales, el primer nivel se puede restar del nivel esperado, proporcionando, de este modo, una diferencia, y se determina una fracción fetal como el doble de la diferencia. En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) se clasifica como una variación en el número de copias maternas y/o fetales, un nivel esperado se resta de un primer nivel, proporcionando, de este modo, una diferencia, y se determina la fracción fetal como el doble de la diferencia.

45 A menudo, una fracción fetal se proporciona como un porcentaje. Por ejemplo, una fracción fetal se puede dividir por 100, proporcionando, de este modo, un valor porcentual. Por ejemplo, para un primer nivel representativo de una duplicación homocigótica materna y que tiene un nivel de 155 y un nivel esperado para una duplicación homocigótica materna que tiene un nivel de 150, se puede determinar una fracción fetal como el 10% (por ejemplo, (fracción fetal = $2 \times (155 - 150)$)).

50 En algunos casos, una fracción fetal se determina a partir de dos o más niveles dentro de un perfil que se clasifican como variaciones en el número de copias. Por ejemplo, a veces dos o más niveles (por ejemplo, dos o más primeros niveles) en un perfil se identifican como significativamente diferentes a un nivel de referencia (por ejemplo, un segundo nivel, un nivel que no incluye prácticamente ninguna variación en el número de copias), los dos o más niveles se clasifican como representativos de una variación en el número de copias maternas y/o fetales y se determina una fracción fetal a partir de cada uno de los dos o más niveles. En algunos casos, una fracción fetal se determina a partir de aproximadamente 3 o más, aproximadamente 4 o más, aproximadamente 5 o más, aproximadamente 6 o más, aproximadamente 7 o más, aproximadamente 8 o más, o aproximadamente 9 o más determinaciones de fracción fetal dentro de un perfil. En algunos casos, una fracción fetal se determina a partir de aproximadamente 10 o más, aproximadamente 20 o más, aproximadamente 30 o más, aproximadamente 40 o más, aproximadamente 50 o más, aproximadamente 60 o más, aproximadamente 70 o más, aproximadamente 80 o más, o aproximadamente 90 o más determinaciones de fracción fetal dentro de un perfil. En algunos casos, una fracción fetal se determina a partir de aproximadamente 100 o más, aproximadamente 200 o más, aproximadamente 300 o más, aproximadamente 400 o más, aproximadamente 500 o más, aproximadamente 600 o más, aproximadamente 700 o más, aproximadamente 800 o más, aproximadamente 900 o más, o aproximadamente 1000 o más determinaciones de fracción fetal dentro de un perfil. En algunos casos, una fracción fetal se determina a partir de aproximadamente 10 a aproximadamente 1000, aproximadamente 20 a aproximadamente 900, aproximadamente 30 a aproximadamente 700,

aproximadamente 40 a aproximadamente 600, aproximadamente 50 a aproximadamente 500, aproximadamente 50 a aproximadamente 400, aproximadamente 50 a aproximadamente 300, aproximadamente 50 a aproximadamente 200, o aproximadamente 50 a aproximadamente 100 determinaciones de fracción fetal dentro de un perfil.

- 5 En algunos casos, se determina una fracción fetal como el promedio o la media de múltiples determinaciones de fracciones fetales dentro de un perfil. En ciertos casos, una fracción fetal determinada a partir de determinaciones de fracciones fetales múltiples es una media (por ejemplo, un promedio, una media, un promedio estándar, una mediana o similares) de determinaciones de fracciones fetales múltiples. A menudo, una fracción fetal determinada a partir de determinaciones fracciones fetales múltiples es un valor medio determinado por un método adecuado conocido en la técnica o descrito en el presente documento. En algunos casos, un valor medio de una determinación de fracción fetal es una media ponderada. En algunos casos, un valor medio de una determinación de fracción fetal es una media no ponderada. Una determinación de la fracción fetal media, mediana o promedio (es decir, un valor de determinación de la fracción fetal media, mediana o promedio) generada a partir de determinaciones de fracciones fetales múltiples a veces se asocia con un valor de incertidumbre (por ejemplo, una varianza, desviación estándar, MAD, o similares). En algunos casos, antes de determinar un valor de fracción fetal medio, mediana o promedio a partir de múltiples determinaciones, se eliminan una o más determinaciones desviadas (descritas con mayor detalle en el presente documento).

- 20 Algunas determinaciones de la fracción fetal dentro de un perfil a veces no se incluyen en la determinación general de una fracción fetal (por ejemplo, la determinación de la fracción fetal media o promedio). En algunos casos, una determinación de la fracción fetal deriva de un primer nivel (por ejemplo, un primer nivel que es significativamente diferente de un segundo nivel) en un perfil y el primer nivel no es indicativo de una variación genética. Por ejemplo, algunos primeros niveles (por ejemplo, máximos o mínimos) en un perfil se generan a partir de anomalías o causas desconocidas. Tales valores a menudo generan determinaciones de fracciones fetales que difieren significativamente de otras determinaciones de fracciones fetales obtenidas a partir de variaciones en el número de copias verdaderas. En algunos casos, las determinaciones de fracciones fetales que difieren significativamente de otras determinaciones de fracciones fetales en un perfil se identifican y eliminan de la determinación de la fracción fetal. Por ejemplo, algunas determinaciones de fracciones fetales obtenidas a partir de máximos y mínimos anómalos se identifican comparándolas con otras determinaciones de fracciones fetales dentro de un perfil y se excluyen de la determinación general de la fracción fetal.

- En algunos casos, una determinación de fracción fetal independiente que difiere significativamente de una determinación de fracción fetal media, mediana o promedio es una diferencia identificada, reconocida y/u observable. En ciertos casos, la expresión "difiere significativamente" puede significar una diferencia estadísticamente diferente y/o una diferencia estadísticamente significativa. Una determinación de fracción fetal "independiente" puede ser una fracción fetal determinada (por ejemplo, en algunos casos una determinación única) a partir de un nivel específico clasificado como una variación en el número de copias. Se puede usar cualquier umbral o intervalo adecuado para determinar que la determinación de fracción fetal difiere significativamente de la determinación de fracción fetal media, mediana o promedio. En ciertos casos, una determinación de fracción fetal difiere significativamente de una determinación de fracción fetal media, mediana o promedio y la determinación puede expresarse como una desviación porcentual del valor promedio o medio. En ciertos casos, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, mediana o promedio difiere en aproximadamente el 15 por ciento o más. En algunos casos, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, mediana o promedio difiere en aproximadamente el 15% a aproximadamente el 100% o más.

- En ciertos casos, una determinación de fracción fetal difiere significativamente de una determinación de fracción fetal media, mediana o promedio de acuerdo con un múltiplo de un valor de incertidumbre asociado con la determinación de fracción fetal media o promedio. A menudo, un valor de incertidumbre y una constante n (por ejemplo, un intervalo de confianza) definen un intervalo (por ejemplo, un límite de incertidumbre). Por ejemplo, a veces un valor de incertidumbre es una desviación estándar para determinaciones de fracción fetal (por ejemplo, ± 5) y se multiplica por una constante n (por ejemplo, un intervalo de confianza) definiendo, de este modo, un límite de intervalo o incertidumbre (por ejemplo, $5n$ a $-5n$, a veces denominado 5 sigma). En algunos casos, una determinación de fracción fetal independiente está fuera del intervalo definido por el límite de incertidumbre y se considera significativamente diferente de una determinación de fracción fetal media, mediana o promedio. Por ejemplo, para un valor medio de 10 y un límite de incertidumbre de 3, una fracción fetal independiente mayor que 13 o menor que 7 es significativamente diferente. En algunos casos, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, mediana o promedio difiere en más de n veces el valor de incertidumbre (por ejemplo, $n \times$ sigma) donde n es aproximadamente igual a o mayor que 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10. En ciertos casos, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, mediana o promedio difiere en más de n veces el valor de incertidumbre (por ejemplo, $n \times$ sigma) donde n es aproximadamente igual a o mayor que 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 2,1, 2,2, 2,3, 2,4, 2,5, 2,6, 2,7, 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,7, 3,8, 3,9, o 4,0.

- 65 En algunos casos, un nivel es representativo de una microploidía fetal y/o materna. En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado), es significativamente diferente a un segundo nivel, el primer nivel se

clasifica como una variación en el número de copias maternas y/o fetales, y el primer nivel y/o el segundo nivel son representativos de una microploidía fetal y/o una microploidía materna. En ciertos casos, un primer nivel es representativo de una microploidía fetal. En algunos casos, un primer nivel es representativo de una microploidía materna. A menudo, un primer nivel es representativo de una microploidía fetal y una microploidía materna. En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado) es significativamente diferente de un segundo nivel, el primer nivel puede clasificarse como una variación en el número de copias maternas y/o fetales, el primer nivel es representativo de una microploidía fetal y/o materna y se una fracción fetal de acuerdo con la microploidía fetal y/o materna. En algunos casos, un primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, el primer nivel es representativo de una microploidía fetal y se determina una fracción fetal de acuerdo con la microploidía fetal. En algunos casos, un primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, el primer nivel es representativo de una microploidía materna y se determina una fracción fetal de acuerdo con la microploidía materna. En algunos casos, un primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, el primer nivel es representativo de una microploidía materna y fetal y se determina una fracción fetal de acuerdo con la microploidía materna y fetal.

En algunos casos, una determinación de una fracción fetal comprende determinar una microploidía fetal y/o materna. En algunos casos, un nivel (por ejemplo, un primer nivel, un nivel observado), es significativamente diferente a un segundo nivel, el primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, una microploidía fetal y/o materna se determina de acuerdo con el primer nivel y/o segundo nivel y se determina una fracción fetal. En algunos casos, en primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, una microploidía fetal se determina de acuerdo con el primer nivel y/o segundo nivel y se determina una fracción fetal de acuerdo con la microploidía fetal. En ciertos casos, un primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, una microploidía materna se determina de acuerdo con el primer nivel y/o segundo nivel y se determina una fracción fetal de acuerdo con la microploidía materna. En algunos casos, un primer nivel se clasifica como una variación en el número de copias maternas y/o fetales, una microploidía materna y fetal se determina de acuerdo con el primer nivel y/o segundo nivel y se determina una fracción fetal de acuerdo con la microploidía materna y fetal.

Una fracción fetal a menudo se determina cuando la microploidía de la madre es diferente de (por ejemplo, no es lo mismo que) la microploidía del feto para un nivel determinado o para un nivel clasificado como una variación en el número de copias. En algunos casos, se determina una fracción fetal cuando la madre es homocigótica para una duplicación (por ejemplo, una microploidía de 2) y el feto es heterocigótico para la misma duplicación (por ejemplo, una microploidía de 1,5). En algunos casos, se determina una fracción fetal cuando la madre es heterocigótica para una duplicación (por ejemplo, una microploidía de 1,5) y el feto es homocigótico para la misma duplicación (por ejemplo, una microploidía de 2) o la duplicación está ausente en el feto (por ejemplo, una microploidía de 1). En algunos casos, se determina una fracción fetal cuando la madre es homocigótica para una delección (por ejemplo, una microploidía de 0) y el feto es heterocigótico para la misma delección (por ejemplo, una microploidía de 0,5). En algunos casos, se determina una fracción fetal cuando la madre es heterocigótica para una delección (por ejemplo, una microploidía de 0,5) y el feto es homocigótico para la misma delección (por ejemplo, una microploidía de 0) o la delección está ausente en el feto (por ejemplo, una microploidía de 1).

En ciertos casos, una fracción fetal a menudo no puede determinarse cuando la microploidía de la madre es igual a (por ejemplo, identificado como la misma) la microploidía del feto para un nivel determinado identificado como una variación en el número de copias. En algunos casos, por ejemplo, para un nivel dado donde tanto la madre como el feto portan el mismo número de copias de una variación en el número de copias, no se determina una fracción fetal. Por ejemplo, una fracción fetal no se puede determinar para un nivel clasificado como una variación en el número de copias cuando tanto la madre como el feto son homocigóticos para la misma delección o homocigóticos para la misma duplicación. En ciertos casos, una fracción fetal no puede determinarse para un nivel clasificado como una variación en el número de copias cuando tanto la madre como el feto son heterocigóticos para la misma delección o heterocigóticos para la misma duplicación. En los casos en que se realizan determinaciones de fracciones fetales múltiples para una muestra, las determinaciones que se desvían significativamente de un valor medio, mediana o promedio pueden resultar de una variación en el número de copias para la cual la ploidía materna es igual a la ploidía fetal, y dichas determinaciones se pueden eliminar del estudio.

En algunos casos, la microploidía de una variación en el número de copias maternas y la variación en el número de copias fetales se desconoce. En algunos casos, en los casos en que no hay una determinación de microploidía fetal y/o materna para una variación en el número de copias, se genera una fracción fetal y se compara con una determinación de fracción fetal media, mediana o promedio. La determinación de fracción fetal para una variación en el número de copias que difiere significativamente de la determinación de fracción fetal media, mediana o promedio se debe, a veces, a que la microploidía de la madre y el feto es la misma para la variación en el número de copias. Una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, mediana o promedio a menudo se excluye de una determinación de fracción fetal general independientemente de la fuente o la causa de la diferencia. En algunos casos, la microploidía de la madre y/o el feto se determina y/o verifica mediante un método conocido en la técnica (por ejemplo, mediante métodos de secuenciación dirigida).

Ploidía fetal

En algunos casos, se usa una determinación de ploidía fetal, en parte, para determinar la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica, una trisomía). Se puede determinar una ploidía fetal, en parte, a partir de una medida de la fracción fetal determinada por un método adecuado de determinación de la fracción fetal, incluidos los métodos descritos en el presente documento. En algunos casos, la ploidía fetal se determina de acuerdo con una determinación de fracción fetal y la ecuación (8), (20), (21) o una variación o derivación de la misma (Ejemplo 2). En algunos casos, la ploidía fetal se determina mediante un método descrito a continuación. En algunos casos, cada método descrito a continuación requiere un recuento de referencia calculado F_i (a veces representado como f_i) determinado para una porción (es decir, una porción, i) de un genoma para muestras múltiples donde la ploidía del feto para la porción i del genoma es euploide. En algunos casos, se determina un valor de incertidumbre (por ejemplo, una desviación estándar, σ) para el recuento de referencia f_i . En algunos casos, se usa un recuento de referencia f_i para un valor de incertidumbre, un recuento de muestra de ensayo y/o una fracción fetal medida (F) para determinar la ploidía fetal de acuerdo con un método que se describe a continuación. En algunos casos, se normaliza un recuento de referencia (por ejemplo, un recuento de referencia promedio, medio o mediana) mediante un método descrito en el presente documento (por ejemplo, normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS, PERUN, RM, GCRM y/o combinaciones de las mismas). En algunos casos, un recuento de referencia de un segmento de un genoma que es euploide es igual a 1 cuando el recuento de referencia se normaliza por PERUN. En algunos casos, tanto el recuento de referencia (por ejemplo, para un feto que se sabe que es euploide) como los recuentos de una muestra de ensayo para una porción o segmento de un genoma se normalizan por PERUN y el recuento de referencia es igual a 1. Análogamente, en algunos casos, un recuento de referencia de una porción o segmento de un genoma que es euploide es igual a 1 cuando los recuentos se normalizan (es decir, dividen por) una mediana del recuento de referencia. Por ejemplo, en algunos casos, tanto el recuento de referencia (por ejemplo, para un feto que es euploide) como los recuentos de una muestra de ensayo para una porción o segmento de un genoma se normalizan mediante un recuento de referencia promedio, el recuento de referencia normalizado es igual a 1 y el recuento de la muestra de ensayo se normaliza (p. ej., dividido por) el recuento de referencia medio. En algunos casos, tanto el recuento de referencia (por ejemplo, para un feto que es euploide) como los recuentos de una muestra de ensayo para una porción o segmento de un genoma se normalizan por GCRM, GC, RM o un método adecuado. En algunos casos, un recuento de referencia es un recuento de referencia promedio, medio o mediana. Un recuento de referencia suele ser un recuento normalizado de una porción (por ejemplo, un nivel de sección genómica normalizado). En algunos casos, un recuento de referencia y los recuentos para una muestra de ensayo son recuentos sin procesar. En algunos casos, un recuento de referencia se determina a partir de un perfil de recuento promedio, medio o mediana. En algunos casos, un recuento de referencia es un nivel de sección genómica calculado. En algunos casos, un recuento de referencia de una muestra de referencia y un recuento de una muestra de ensayo (por ejemplo, una muestra del paciente, por ejemplo, y_i) se normalizan mediante el mismo método o proceso.

En algunos casos, se determina una medida de la fracción fetal (F). Este valor de la fracción fetal se usa después para determinar la ploidía fetal de acuerdo con la ecuación (8), una derivación o una variación de la misma. En algunos casos, se devuelve un valor negativo si el feto es euploide y se devuelve un valor positivo si el feto no es euploide. En algunos casos, un valor negativo indica que el feto es euploide para el segmento del genoma considerado. En ciertos casos, un valor que no es negativo indica que el feto comprende una trisomía. En ciertos casos, cualquier valor positivo indica que el feto comprende una aneuploidía (por ejemplo, una trisomía, una duplicación).

En algunos casos, se determina una suma de restos al cuadrado. Por ejemplo, una ecuación que representa la suma de restos al cuadrado derivados de la ecuación (8) se ilustra en la ecuación (18). En algunos casos, se determina una suma de restos al cuadrado a partir de la ecuación (8) para un valor de ploidía X establecido en un valor de 1 (véase la ecuación (9)) y para un valor de ploidía establecido en un valor de $3/2$ (véase la ecuación (13)). En algunos casos, la suma de restos al cuadrado (ecuaciones (9) y (13)) se determina para un segmento de un genoma o cromosoma (por ejemplo, para todas las porciones de un genoma de referencia i en un segmento del genoma). Por ejemplo, la suma de restos al cuadrado (por ejemplo, las ecuaciones (9) y (13)) se puede determinar para el cromosoma 21, 13, 18 o una porción de los mismos. En algunos casos, para determinar el estado de ploidía de un feto, el resultado de la ecuación (13) se resta de la ecuación (9) para llegar a un valor, ϕ (por ejemplo, véase la ecuación (14)). En ciertos casos, el signo (es decir, positivo o negativo) del valor ϕ determina la presencia o ausencia de una aneuploidía fetal. En ciertos casos, un valor de ϕ (por ejemplo, de la ecuación (14)) que es negativo indica la ausencia de una aneuploidía (por ejemplo, el feto es euploide para porciones de un genoma de referencia i) y un valor de ϕ que no es negativo indica la presencia de una aneuploidía (por ejemplo, una trisomía).

En algunos casos, el recuento de referencia f_i , el valor de incertidumbre para el recuento de referencia σ y/o la fracción fetal medida (F) se usan en las ecuaciones (9) y (13) para determinar la suma de restos al cuadrado de la suma de todas las porciones de genoma de referencia i . En algunos casos, el recuento de referencia f_i , el valor de incertidumbre para el recuento de referencia σ y/o la fracción fetal medida (F) se usan en las ecuaciones (9) y (13) para determinar la ploidía fetal. En algunos casos, los recuentos (por ejemplo, recuentos normalizados, por ejemplo, nivel de sección genómica calculada), representados por y_i para la porción i , para una muestra de ensayo se usan para determinar el estado de ploidía del feto para la porción i . Por ejemplo, en ciertos casos, el estado de ploidía para un segmento de un genoma se determina de acuerdo con un recuento de referencia f_i , un valor de incertidumbre (por ejemplo, del

recuento de referencia), una fracción fetal (F) determinada para una muestra de ensayo y los recuentos y_i determinados para la muestra de ensayo donde el estado de ploidía se determina de acuerdo con la ecuación (14) o una derivación o variación de la misma. En algunos casos, los recuentos y_i y/o los recuentos de referencia se normalizan mediante un método descrito en el presente documento (por ejemplo, normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS, PERUN, RM, GCRM y combinaciones de los mismos). En algunos casos, un estado de ploidía fetal (por ejemplo, euploide, aneuploide, trisomía) para una porción o segmento de un genoma o cromosoma se determina mediante el ejemplo no limitante descrito anteriormente y en la sección de Ejemplos.

En algunos casos, se determina una fracción fetal a partir de una muestra de ensayo, se determinan los recuentos y_i para una muestra de ensayo y ambos se usan para determinar una ploidía para un feto de una muestra de ensayo. En ciertos casos del método descrito en el presente documento, el valor de la ploidía fetal representado por X, no es fijo o asumido. En ciertos casos del método descrito en el presente documento, la fracción fetal F de fija. En algunos casos, se determina una ploidía (por ejemplo, un valor de ploidía) para una porción o segmento de un genoma de acuerdo con la ecuación (20) o (21) (Ejemplo 2). En algunos casos de este método, se determina un valor de ploidía, donde el valor es cercano a 1, 3/2 o 5/4. En algunos casos, un valor de ploidía de aproximadamente 1 indica un feto euploide, un valor de aproximadamente 3/2 indica una trisomía fetal y, en el caso de los gemelos, un valor de aproximadamente 5/4 indica que un feto comprende una trisomía y el otro es euploide para la porción o segmento del genoma considerado. La información adicional con respecto a la determinación de la presencia o ausencia de una aneuploidía fetal a partir de una determinación de ploidía fetal se describe en otra sección a continuación.

En algunos casos, la fracción fetal se determina, se fija, a su valor determinado y la ploidía fetal se determina a partir de una regresión. Se puede utilizar cualquier regresión adecuada, cuyos ejemplos no limitantes incluyen una regresión lineal, una regresión no lineal (p.ej., una regresión polinomial), y similares. En algunos casos, se usa una regresión lineal de acuerdo con la ecuación (8), (20), (21) y/o una derivación o variación de las mismas. En algunos casos, la regresión lineal utilizada es de acuerdo con una suma de restos al cuadrado derivados de la ecuación (8), (20), (21) y/o una derivación o variación de las mismas. En algunos casos, la ploidía fetal se determina de acuerdo con la ecuación (8), (20), (21) y/o una derivación o variación de las mismas y no se usa una regresión. En algunos casos, la ploidía fetal se determina de acuerdo con una suma de restos al cuadrado derivados de la ecuación (8), (20), (21) y/o una derivación o variación de las mismas para múltiples porciones de un genoma de referencia i y no se usa una regresión. Una derivación de una ecuación es cualquier variación de la ecuación obtenida de una prueba matemática de una ecuación.

En algunos casos, se usa un recuento de referencia f_i (descrito anteriormente en el presente documento), un valor de incertidumbre σ y/o una fracción fetal medida (F) en las ecuaciones (20) y (21) para determinar una ploidía fetal. En algunos casos, se usa un recuento de referencia f_i , un valor de incertidumbre σ y/o una fracción fetal medida (F) en las ecuaciones (20) o (21) para determinar una ploidía X fetal para la porción i o para una suma de múltiples porciones de un genoma de referencia i (por ejemplo, para la suma de todas las porciones de un genoma de referencia i para un cromosoma o segmento del mismo). En algunos casos, los recuentos (por ejemplo, recuentos normalizados, nivel de sección genómica calculado), representados por y_i para la porción i , para una muestra de ensayo se usan para determinar la ploidía de un feto para un segmento de un genoma representado por múltiples porciones de un genoma de referencia i . Por ejemplo, en ciertos casos, la ploidía X para un segmento de un genoma se determina de acuerdo con un recuento de referencia f_i , un valor de incertidumbre, una fracción fetal (F) determinada para una muestra de ensayo y los recuentos y_i determinados para la muestra de ensayo donde se determina la ploidía de acuerdo con la ecuación (20), (21) o una derivación o variación de las mismas. En algunos casos, los recuentos y_i y/o los recuentos de referencia se normalizan mediante un método descrito en el presente documento (por ejemplo, normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS, PERUN, RM, GCRM y combinaciones de los mismos). En algunos casos, los recuentos y_i y/o los recuentos de referencia se normalizan y/o procesan mediante el mismo método (por ejemplo, normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS, PERUN, RM, GCRM, un método descrito en el presente documento o combinaciones de los mismos). En algunos casos, los recuentos y_i y f_i son recuentos mapeados en la misma porción o segmento de un genoma o cromosoma.

El valor de incertidumbre σ puede ser una medida adecuada de error, cuyos ejemplos no limitantes incluyen desviación estándar, error estándar, varianza calculada, valor de p o desviación media absoluta (MAD). Puede determinarse el valor de incertidumbre σ para cualquier medición adecuada, cuyos ejemplos no limitantes incluyen puntuaciones Z, valores de Z, valores de t , valores de p , error de validación cruzada, nivel de sección genómica, niveles de sección genómica calculados, niveles, recuentos, similares, o combinaciones de los mismos.

En algunos casos, σ se establece a un valor de 1. En algunos casos, σ no se establece a un valor de 1. En algunos casos, el valor de σ se estima y algunas veces se mide y/o calcula.

En algunos casos, M_i es la ploidía de la madre (es decir, ploidía materna) para una porción del genoma i . En algunos casos, se determina M_i para el mismo paciente (por ejemplo, la misma muestra de ensayo) a partir de la que se determina y_i . En algunos casos, la ploidía materna M_i se conoce o determina de acuerdo con un método descrito en

el presente documento. En algunos casos, la ploidía materna se determina antes o después del *padding* (por ejemplo, después de realizar ajustes de nivel). En algunos casos, M_i se estima o determina por la visualización de un perfil. En algunos casos, la ploidía materna M_i no se conoce. En algunos casos, la ploidía materna M_i se asume. Por ejemplo, en algunos casos, se asume o se sabe que la madre no tiene deleciones y/o duplicaciones en el segmento del genoma que se está evaluando. En algunos casos, se asume o se sabe que la ploidía materna es 1. En algunos casos, la ploidía materna se establece en un valor de 1 después del *padding* (por ejemplo, después de realizar ajustes de nivel). En algunos casos, la ploidía materna se ignora y se establece en un valor de 1. En algunos casos, la ecuación (21) deriva de la ecuación (20) con la suposición de que la madre no tiene deleciones ni duplicaciones en el segmento del genoma que se está evaluando.

En algunos casos, un método para determinar la ploidía fetal es de acuerdo con las lecturas de secuencias de ácidos nucleicos para una muestra de ensayo obtenida de una gestante. En algunos casos, las lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra (por ejemplo, una muestra de ensayo). En algunos casos, un método para determinar la ploidía fetal comprende obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia. En algunos casos, las lecturas de secuencia se mapean a un subconjunto de porciones del genoma de referencia. En algunos casos, determinar la ploidía fetal comprende determinar una fracción fetal. En algunos casos, determinar la ploidía fetal comprende calcular o determinar los niveles de sección genómica. En algunos casos, determinar la ploidía fetal comprende determinar una fracción fetal y calcular o determinar los niveles de sección genómica. En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir de la misma muestra de ensayo (por ejemplo, la misma parte de la muestra de ensayo). En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir de las mismas lecturas obtenidas de la misma muestra de ensayo (por ejemplo, la misma parte de la muestra de ensayo). En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir de las mismas lecturas obtenidas de la misma ejecución de secuenciación y/o de la misma celda de flujo. En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir del mismo equipo y/o máquina (por ejemplo, máquina de secuenciación, celda de flujo o similares).

En algunos casos, se determina un método para determinar la ploidía fetal de acuerdo con una determinación de la fracción fetal y recuentos normalizados (por ejemplo, niveles de sección genómica calculados) donde la determinación de la fracción fetal y los recuentos normalizados (por ejemplo, niveles de sección genómica calculados) se determinan a partir de diferentes partes de una muestra de ensayo (por ejemplo, diferentes alícuotas o, por ejemplo, diferentes muestras de ensayo tomadas aproximadamente al mismo tiempo del mismo sujeto o paciente). Por ejemplo, a veces se determina una fracción fetal a partir de una primera parte de una muestra de ensayo y los recuentos normalizados y/o los niveles de sección genómica se determinan a partir de una segunda parte de la muestra de ensayo. En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir de diferentes muestras de ensayo (por ejemplo, diferentes partes de una muestra de ensayo) tomadas del mismo sujeto (por ejemplo, paciente). En algunos casos, la fracción fetal y los niveles de sección genómica calculados se determinan a partir de lecturas obtenidas en diferentes momentos. En algunos casos, la determinación de la fracción fetal y los recuentos normalizados (por ejemplo, los niveles de sección genómica calculados) se determinan a partir de diferentes equipos y/o de diferentes máquinas (por ejemplo, máquina de secuenciación, celda de flujo o similares).

Resultado

Los métodos descritos en el presente documento pueden proporcionar una determinación de la presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal) para una muestra, proporcionando, de este modo, un resultado (por ejemplo, proporcionando, de este modo, un resultado determinante de la presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal)). Una variación genética, con frecuencia, incluye una ganancia, una pérdida y/o alteración (por ejemplo, duplicación, deleción, fusión, inserción, mutación, reorganización, sustitución o metilación aberrante) de información genética (por ejemplo, cromosomas, segmentos de cromosomas, regiones polimórficas, regiones traslocadas, secuencia de nucleótidos alterada, similares o combinaciones de los anteriores) que dan como resultado un cambio detectable en el genoma o información genética de un sujeto de ensayo con respecto a una referencia. La presencia o ausencia de una variación genética se puede determinar mediante transformación, análisis y/o manipulación de lecturas de secuencias que se han mapeado en porciones (por ejemplo, recuentos, recuentos de porciones genómicas de un genoma de referencia). En algunos casos, determinar un resultado comprende analizar el ácido nucleico de una gestante. En ciertos casos, un resultado se determina de acuerdo con los recuentos (por ejemplo, recuentos normalizados) obtenidos de una gestante, donde los recuentos son del ácido nucleico obtenido de la gestante.

Los métodos descritos en el presente documento a veces determinan la presencia o ausencia de una aneuploidía fetal (por ejemplo, aneuploidía cromosómica completa, aneuploidía cromosómica parcial o aberración cromosómica segmentaria (por ejemplo, mosaicismo, deleción y/o inserción)) para una muestra de ensayo de una gestante con un feto. En ciertos casos, los métodos descritos en el presente documento detectan euploidía o falta de euploidía (no euploidía) para una muestra de una gestante con un feto. En ciertos casos, los métodos descritos en el presente documento detectan trisomía para uno o más cromosomas (por ejemplo, cromosoma 13, cromosoma 18, cromosoma 21 o una combinación de los mismos) o un segmento de los mismos.

En algunos casos, la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía fetal) se determina mediante un método descrito en el presente documento, mediante un método conocido en la técnica o una combinación de los mismos. La presencia o ausencia de una variación genética generalmente se determina a partir de los recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia. Los recuentos de lecturas de secuencia utilizados para determinar la presencia o ausencia de una variación genética a veces son recuentos sin procesar y/o recuentos filtrados, y con frecuencia son recuentos normalizados. Se puede usar un proceso o procesos de normalización adecuados para generar recuentos normalizados, cuyos ejemplos no limitantes incluyen normalización por porciones, normalización por contenido de GC, regresión de mínimos cuadrados lineal y no lineal, LOESS, LOESS de GC, LOWESS, PERUN, RM, GCRM y combinaciones de los mismos. Los recuentos normalizados a veces se expresan como uno o más niveles o niveles en un perfil para un conjunto o conjuntos particulares de porciones. En ocasiones, los recuentos normalizados se ajustan o rellenan antes de determinar la presencia o ausencia de una variación genética.

En algunos casos, un resultado se determina de acuerdo con uno o más niveles. En algunos casos, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con uno o más niveles ajustados. En algunos casos, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con un perfil que comprende de 1 a aproximadamente 10.000 niveles ajustados. Con frecuencia, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con un perfil que comprende de aproximadamente 1 a aproximadamente 1000, de 1 a aproximadamente 900, de 1 a aproximadamente 800, de 1 a aproximadamente de 700, de 1 a aproximadamente 600, 1 a aproximadamente 500, 1 a aproximadamente 400, 1 a aproximadamente 300, 1 a aproximadamente 200, 1 a aproximadamente 100, 1 a aproximadamente 50, 1 a aproximadamente 25, 1 a aproximadamente 20, 1 a aproximadamente 15, 1 a aproximadamente alrededor de 10, o 1 a alrededor de 5 ajustes. En algunos casos, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con un perfil que comprende aproximadamente 1 ajuste (por ejemplo un nivel ajustado). En algunos casos, un resultado se determina de acuerdo con uno o más perfiles (por ejemplo, un perfil de un cromosoma o segmento del mismo) que comprende uno o más, 2 o más, 3 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más o a veces 10 o más ajustes. En algunos casos, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con un perfil donde algunos niveles en un perfil no están ajustados. En algunos casos, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina de acuerdo con un perfil donde no se han realizado ajustes.

En algunos casos, un ajuste de un nivel (por ejemplo, un primer nivel) en un perfil reduce una determinación falsa o un resultado falso. En algunos casos, un ajuste de un nivel (por ejemplo, un primer nivel) en un perfil reduce la frecuencia y/o probabilidad (por ejemplo, probabilidad estadística, posibilidad) de una determinación falsa o resultado falso. Una determinación falsa o un resultado falso puede ser una determinación o un resultado que no son exactos. Una determinación falsa o resultado falso puede ser una determinación o resultado que no refleje la composición genética real o verdadera o la disposición genética real o verdadera (por ejemplo, la presencia o ausencia de una variación genética) de un sujeto (por ejemplo, una gestante, un feto y/o una combinación de ambos). En algunos casos, una determinación falsa o resultado falso es una determinación falsa negativa. En algunos casos, una determinación negativa o un resultado negativo es la ausencia de una variación genética (por ejemplo, aneuploidía, variación en el número de copias). En algunos casos, una determinación falsa o un resultado falso es una determinación de positivos falsos o un resultado positivo falso. En algunos casos, una determinación positiva o un resultado positivo es la presencia de una variación genética (por ejemplo, aneuploidía, variación en el número de copias). En algunos casos, se utiliza una determinación o resultado en un diagnóstico. En algunos casos, una determinación o resultado es para un feto.

La presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal) a veces se determina sin comparar los recuentos de un conjunto de porciones con una referencia. Los recuentos medidos para una muestra de ensayo y que están en una región de ensayo (por ejemplo, un conjunto de porciones de interés) se denominan "recuentos de ensayo" en el presente documento. Los recuentos de ensayo a veces son recuentos procesados, recuentos promediados o sumados, una representación, recuentos normalizados o uno o más niveles o niveles como se describe en el presente documento. En ciertos casos, los recuentos de ensayo se promedian o suman (por ejemplo, se calcula un promedio, media, mediana, modo o suma) para un conjunto de porciones, y los recuentos promediados o sumados se comparan con un umbral o intervalo. Los recuentos de ensayo a veces se expresan como una representación, que se puede expresar como una proporción o porcentaje de recuentos para un primer conjunto de porciones a recuentos para un segundo conjunto de porciones. En ciertos casos, un primer conjunto de porciones es para uno o más cromosomas de ensayo (por ejemplo, cromosoma 13, cromosoma 18, cromosoma 21 o una combinación de los mismos) y, a veces, un segundo conjunto de porciones es para un genoma o una parte de un genoma (por ejemplo, autosomas o autosomas y cromosomas sexuales). En algunos casos, un primer conjunto de porciones es para uno o más cromosomas sexuales (por ejemplo, cromosoma X, cromosoma Y o una combinación de los mismos) y, a veces, un segundo conjunto de porciones es para uno o más autosomas. En algunos casos, un primer conjunto de porciones es para una o más primeras regiones de un cromosoma de ensayo (por ejemplo, cromosoma X, cromosoma Y, o una combinación de los mismos) y, a veces, un segundo conjunto de porciones es para una o más segundas regiones de un cromosoma de ensayo (por ejemplo, cromosoma X, cromosoma Y, o una combinación de los mismos) o el

cromosoma de ensayo completo. En ciertos casos, una representación se compara con un umbral o intervalo. En ciertos casos, los recuentos de ensayo se expresan como uno o más niveles o niveles para recuentos normalizados sobre un conjunto de porciones, y el uno o más niveles o niveles se comparan con un umbral o intervalo. Los recuentos de ensayo (por ejemplo, recuentos promediados o sumados, representación, recuentos normalizados, uno o más niveles o niveles) por encima o por debajo de un umbral particular, en un intervalo particular o fuera de un intervalo particular a veces son determinantes de la presencia de una variación genética o falta de euploidía (por ejemplo, no euploidía). Los recuentos de ensayo (por ejemplo, recuentos promediados o sumados, representación, recuentos normalizados, uno o más niveles o niveles) por debajo o por encima de un umbral particular, en un intervalo particular o fuera de un intervalo particular a veces son determinantes de la ausencia de una variación genética o euploidía.

La presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal) a veces se determina mediante comparación de recuentos, cuyos ejemplos no limitantes incluyen recuentos de ensayo, recuentos de referencia, recuentos sin procesar, recuentos filtrados, recuentos promediados o sumados, representaciones (por ejemplo, representaciones cromosómicas), recuentos normalizados, uno o más niveles (por ejemplo, para un conjunto de porciones, por ejemplo, nivel de sección genómica, perfiles), puntuaciones Z, similares o combinaciones de los mismos. En algunos casos, los recuentos de ensayo se comparan con una referencia (por ejemplo, recuentos de referencia). Una referencia (por ejemplo, un recuento de referencia) puede ser una determinación adecuada de recuentos, cuyos ejemplos no limitantes incluyen recuentos sin procesar, recuentos filtrados, recuentos promediados o sumados, representaciones (por ejemplo, representaciones cromosómicas), recuentos normalizados, uno o más niveles (por ejemplo, para un conjunto de porciones, por ejemplo, nivel de sección genómica, perfiles), puntuaciones Z, similares o combinaciones de los mismos. Con frecuencia, los recuentos de referencia son recuentos para una región de ensayo euploide o de un segmento de un genoma o cromosoma que es euploide. En algunos casos, los recuentos de referencia y los recuentos de ensayo se obtienen de la misma muestra y/o del mismo sujeto. En algunos casos, los recuentos de referencia son de diferentes muestras y/o de diferentes sujetos. En algunos casos, los recuentos de referencia se determinan y/o comparan con un segmento correspondiente del genoma del que derivan y/o determinan los recuentos de ensayo. Un segmento correspondiente se refiere a un segmento, porción o conjunto de porciones que mapean a la misma ubicación de un genoma de referencia. En algunos casos, los recuentos de referencia se determinan y/o comparan con un segmento diferente del genoma del que derivan y/o determinan los recuentos de ensayo.

En ciertos casos, los recuentos de ensayo a veces son para un primer conjunto de porciones y una referencia incluye recuentos para un segundo conjunto de porciones diferente al primer conjunto de porciones. Los recuentos de referencia a veces son para una muestra de ácido nucleico de la misma gestante de la que se obtiene la muestra de ensayo. En ciertos casos, los recuentos de referencia son para una muestra de ácido nucleico de una o más gestantes diferentes de la hembra de la que se obtuvo la muestra de ensayo. En algunos casos, un primer conjunto de porciones está en el cromosoma 13, cromosoma 18, cromosoma 21, un segmento de los mismos o una combinación de los anteriores, y el segundo conjunto de porciones está en otro cromosoma o cromosomas o un segmento de los mismos. En un ejemplo no limitante, donde un primer conjunto de porciones está en el cromosoma 21 o un segmento del mismo, con frecuencia, un segundo conjunto de porciones está en otro cromosoma (por ejemplo, cromosoma 1, cromosoma 13, cromosoma 14, cromosoma 18, cromosoma 19, segmento de los mismos o combinaciones de los anteriores). Con frecuencia, una referencia se encuentra en un cromosoma o segmento del mismo que es típicamente euploide. Por ejemplo, el cromosoma 1 y el cromosoma 19, con frecuencia, son euploides en los fetos debido a una alta tasa de mortalidad fetal temprana asociada con las aneuploidías del cromosoma 1 y del cromosoma 19. Se puede generar una medida de desviación entre los recuentos de ensayo y los recuentos de referencia.

En ciertos casos, una referencia comprende recuentos para el mismo conjunto de porciones que para los recuentos de ensayo, donde los recuentos para la referencia son de una o más muestras de referencia (por ejemplo, con frecuencia, múltiples muestras de referencia de múltiples sujetos de referencia). Con frecuencia, una muestra de referencia procede de una o más gestantes diferentes de una hembra de la que se obtiene una muestra de ensayo. Se puede generar una medida de desviación (por ejemplo, una medida de incertidumbre, un valor de incertidumbre) entre los recuentos de ensayo y los recuentos de referencia. En algunos casos, se determina una medida de desviación a partir de los recuentos de ensayo. En algunos casos, se determina una medida de desviación a partir de los recuentos de referencia. En algunos casos, se determina una medida de desviación a partir de un perfil completo o un subconjunto de porciones dentro de un perfil.

Se puede seleccionar una medida de desviación adecuada, cuyos ejemplos no limitantes incluyen desviación estándar, desviación absoluta promedio, desviación absoluta media, desviación absoluta máxima, puntuación estándar (por ejemplo, valor de Z, puntuación z, puntuación normal, variable estandarizada) y similares. En algunos casos, las muestras de referencia son euploides para una región de ensayo y se evalúa la desviación entre los recuentos de ensayo y los recuentos de referencia. En algunos casos, una determinación de la presencia o ausencia de una variación genética es de acuerdo con el número de desviaciones (por ejemplo, medidas de desviaciones, MAD) entre los recuentos de ensayo y los recuentos de referencia para un segmento o porción de un genoma o cromosoma. En algunos casos, la presencia de una variación genética se determina cuando el número de desviaciones entre los recuentos de ensayo y los recuentos de referencia es mayor que aproximadamente 1, mayor que aproximadamente 1,5, mayor que aproximadamente 2, mayor que aproximadamente 2,5, mayor que aproximadamente 2,6, mayor que aproximadamente 2,7, mayor que aproximadamente 2,8, mayor que aproximadamente 2,9, mayor que

aproximadamente 3, mayor que aproximadamente 3,1, mayor que aproximadamente 3,2, mayor que aproximadamente 3,3, mayor que aproximadamente 3,4, mayor que aproximadamente 3,5, mayor que aproximadamente 4, mayor que aproximadamente 5, o mayor que aproximadamente 6. Por ejemplo, a veces, un recuento de ensayo difiere de un recuento de referencia en más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 MAD) y se determina la presencia de una variación genética. En algunos casos, un recuento de ensayo obtenido de una gestante es mayor que un recuento de referencia por más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 MAD) y se determina la presencia de una aneuploidía cromosómica fetal (por ejemplo, una trisomía fetal). Una desviación de más de tres entre los recuentos de ensayo y los recuentos de referencia, con frecuencia, es indicativa de una región de ensayo no euploide (por ejemplo, la presencia de una variación genética). Los recuentos de ensayo significativamente por encima de los recuentos de referencia, cuyos recuentos de referencia son indicativos de euploidía, a veces son determinantes de una trisomía. En algunos casos, un recuento de ensayo obtenido de una gestante es menor que un recuento de referencia por más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 MAD) y se determina la presencia de una aneuploidía cromosómica fetal (por ejemplo, una monosomía fetal). Los recuentos de ensayo significativamente por debajo de los recuentos de referencia, cuyos recuentos de referencia son indicativos de euploidía, a veces son determinantes de una monosomía.

En algunos casos, la ausencia de una variación genética se determina cuando el número de desviaciones entre los recuentos de ensayo y los recuentos de referencia es menor que aproximadamente 3,5, menor que aproximadamente 3,4, menor que aproximadamente 3,3, menor que aproximadamente 3,2, menor que aproximadamente 3,1, menor que aproximadamente 3,0, menor que aproximadamente 2,9, menor que aproximadamente 2,8, menor que aproximadamente 2,7, menor que aproximadamente 2,6, menor que aproximadamente 2,5, menor que aproximadamente 2,0, menor que aproximadamente 1,5 o menor que aproximadamente 1,0. Por ejemplo, a veces, un recuento de ensayo difiere de un recuento de referencia en menos de 3 medidas de desviación (por ejemplo, 3 sigma, 3 MAD) y se determina la ausencia de una variación genética. En algunos casos, un recuento de ensayo obtenido de una gestante difiere de recuento de referencia por menos de 3 medidas de desviación (por ejemplo, 3 sigma, 3 MAD) y se determina la ausencia de una aneuploidía cromosómica fetal (por ejemplo, una euploidía fetal). En algunos casos, por ejemplo, la desviación de menos de tres entre los recuentos de ensayo y los recuentos de referencia (por ejemplo, 3-sigma para la desviación estándar) con frecuencia, puede ser indicativa de una región de ensayo euploide (por ejemplo, ausencia de una variación genética). Se puede representar gráficamente y visualizar una medida de la desviación entre los recuentos de ensayo para una muestra de ensayo y los recuentos de referencia para uno o más sujetos de referencia (por ejemplo, representación gráfica de puntuación z).

Cualquier otra referencia adecuada puede factorizarse con los recuentos de ensayo para determinar la presencia o ausencia de una variación genética (o determinación de euploide o no euploide) para una región de ensayo de una muestra de ensayo. Por ejemplo, se puede factorizar una determinación de la fracción fetal con los recuentos de ensayo para determinar la presencia o ausencia de una variación genética. Se puede utilizar un proceso adecuado para cuantificar la fracción fetal, cuyos ejemplos no limitantes incluyen un proceso de espectrometría de masas, proceso de secuenciación o combinación de los mismos.

En algunos casos, la presencia o ausencia de una aneuploidía cromosómica fetal (por ejemplo, una trisomía), se determina, en parte, a partir de una determinación de ploidía fetal. En algunos casos, se determina una ploidía fetal mediante un método descrito en el presente documento. En algunos casos determinados, una determinación de ploidía fetal de aproximadamente 1,20 o mayor, 1,25 o mayor, 1,30 o mayor, aproximadamente 1,35 o mayor, aproximadamente 1,4 o mayor, o aproximadamente 1,45 o mayor indica la presencia de una aneuploidía cromosómica fetal (por ejemplo, la presencia de una trisomía fetal). En algunos casos, una determinación de ploidía fetal de aproximadamente 1,20 a aproximadamente 2,0, aproximadamente 1,20 a aproximadamente 1,9, aproximadamente 1,20 a aproximadamente 1,85, aproximadamente 1,20 a aproximadamente 1,8, aproximadamente 1,25 a aproximadamente 2,0, aproximadamente 1,25 a aproximadamente 1,9, aproximadamente 1,25 a aproximadamente 1,85, aproximadamente 1,25 a aproximadamente 1,8, aproximadamente 1,3 a aproximadamente 2,0, aproximadamente 1,3 a aproximadamente 1,9, aproximadamente 1,3 a aproximadamente 1,85, aproximadamente 1,3 a aproximadamente 1,8, aproximadamente 1,35 a aproximadamente 2,0, aproximadamente 1,35 a aproximadamente 1,9, aproximadamente 1,35 a aproximadamente 1,8, aproximadamente 1,4 a aproximadamente 2,0, aproximadamente 1,4 a aproximadamente 1,85 o aproximadamente 1,4 a aproximadamente 1,8 indica la presencia de una aneuploidía cromosómica fetal (por ejemplo, la presencia de una trisomía fetal). En algunos casos, la aneuploidía fetal es una trisomía. En algunos casos, la aneuploidía fetal es una trisomía del cromosoma 13, 18 y/o 21.

En algunos casos, una ploidía fetal menor que aproximadamente 1,35, menor que aproximadamente 1,30, menor que aproximadamente 1,25, menor que aproximadamente 1,20 o menor que aproximadamente 1,15 indica la ausencia de una aneuploidía fetal (por ejemplo, la ausencia de una trisomía fetal, por ejemplo, euploide). En algunos casos, una determinación de ploidía fetal de aproximadamente 0,7 a aproximadamente 1,35, aproximadamente 0,7 a aproximadamente 1,30, aproximadamente 0,7 a aproximadamente 1,25, aproximadamente 0,7 a aproximadamente 1,20, aproximadamente 0,7 a aproximadamente 1,15, aproximadamente 0,75 a aproximadamente 1,35, aproximadamente 0,75 a aproximadamente 1,30, aproximadamente 0,75 a aproximadamente 1,25, aproximadamente 0,75 a aproximadamente 1,20, aproximadamente 0,75 a aproximadamente 1,15, aproximadamente 0,8 a aproximadamente 1,35, aproximadamente 0,8 a aproximadamente 1,30, aproximadamente 0,8 a aproximadamente 1,25, aproximadamente 0,8 a aproximadamente 1,20 o aproximadamente 0,8 a aproximadamente 1,15 indica la

ausencia de una aneuploidía cromosómica fetal (por ejemplo, la ausencia de una trisomía fetal, por ejemplo, euploide).

En algunos casos, una ploidía fetal menor que aproximadamente 0,8, menor que aproximadamente 0,75, menor que aproximadamente 0,70 o menor que aproximadamente 0,6 indica la presencia de una aneuploidía fetal (por ejemplo, la presencia de una delección cromosómica). En algunos casos, una determinación de ploidía fetal de aproximadamente 0 a aproximadamente 0,8, aproximadamente 0 a aproximadamente 0,75, aproximadamente 0 a aproximadamente 0,70, aproximadamente 0 a aproximadamente 0,65, aproximadamente 0 a aproximadamente 0,60, aproximadamente 0,1 a aproximadamente 0,8, aproximadamente 0,1 a aproximadamente 0,75, aproximadamente 0,1 a aproximadamente 0,70, aproximadamente 0,1 a aproximadamente 0,65, aproximadamente 0,1 a aproximadamente 0,60, aproximadamente 0,2 a aproximadamente 0,8, aproximadamente 0,2 a aproximadamente 0,75, aproximadamente 0,2 a aproximadamente 0,70, aproximadamente 0,2 a aproximadamente 0,65, aproximadamente 0,2 a aproximadamente 0,60, aproximadamente 0,25 a aproximadamente 0,8, aproximadamente 0,25 a aproximadamente 0,75, aproximadamente 0,25 a aproximadamente 0,70, aproximadamente 0,25 a aproximadamente 0,65, aproximadamente 0,25 a aproximadamente 0,60, aproximadamente 0,3 a aproximadamente 0,8, aproximadamente 0,3 a aproximadamente 0,75, aproximadamente 0,3 a aproximadamente 0,70, aproximadamente 0,3 a aproximadamente 0,65, aproximadamente 0,3 a aproximadamente 0,60 indica la presencia de una aneuploidía cromosómica fetal (por ejemplo, la presencia de una delección cromosómica). En algunos casos, la aneuploidía fetal determinada es una delección de cromosoma completo.

En algunos casos, una determinación de la presencia o ausencia de una aneuploidía fetal (por ejemplo, de acuerdo con uno o más de los intervalos de una determinación de ploidía anterior) se determina de acuerdo con una zona de identificación. En ciertos casos, se realiza una identificación (por ejemplo, una identificación que determine la presencia o ausencia de una variación genética, por ejemplo, un resultado) cuando un valor (por ejemplo, un valor de ploidía, un valor de fracción fetal, un nivel de incertidumbre) o una colección de valores se encuentra dentro de un intervalo predefinido (por ejemplo, una zona, una zona de identificación). En algunos casos, una zona de identificación se define de acuerdo con una colección de valores que se obtienen de la misma muestra de paciente. En ciertos casos, se define una zona de identificación de acuerdo con una colección de valores que derivan del mismo cromosoma o segmento del mismo. En algunos casos, una zona de identificación basada en una determinación de ploidía se define según un nivel de confianza (por ejemplo, un alto nivel de confianza, por ejemplo, un bajo nivel de incertidumbre) y/o una fracción fetal. En algunos casos, se define una zona de identificación de acuerdo con una determinación de ploidía y una fracción fetal de aproximadamente 2,0% o mayor, aproximadamente 2,5% o mayor, aproximadamente 3 % o mayor, aproximadamente 3,25 % o mayor, aproximadamente 3,5 % o mayor, aproximadamente 3,75 % o mayor o aproximadamente 4,0 % o mayor. Por ejemplo, en algunos casos, se realiza una identificación de que un feto comprende una trisomía 21 basándose en una determinación de ploidía mayor que 1,25 con una determinación de la fracción fetal de 2% o mayor o 4% o mayor para una muestra obtenida de una gestante con un feto. En ciertos casos, por ejemplo, se realiza una identificación de que un feto es euploide basándose en una determinación de ploidía menor que 1,25 con una determinación de la fracción fetal de 2% o mayor o 4% o mayor para una muestra obtenida de una gestante con un feto. En algunos casos, se define una zona de identificación por un nivel de confianza de aproximadamente 99% o mayor, aproximadamente 99,1 % o mayor, aproximadamente 99,2 % o mayor, aproximadamente 99,3 % o mayor, aproximadamente 99,4 % o mayor, aproximadamente 99,5 % o mayor, aproximadamente 99,6 % o mayor, aproximadamente 99,7 % o mayor, aproximadamente 99,8 % o mayor o aproximadamente 99,9 % o mayor. En algunos casos, se realiza una identificación sin usar una zona de identificación. En algunos casos, se realiza una identificación utilizando una zona de identificación y datos o información adicionales. En algunos casos, se realiza una identificación basándose en un valor de ploidía sin el uso de una zona de identificación. En algunos casos, se realiza una identificación sin calcular un valor de ploidía. En algunos casos, se realiza una identificación basándose en la inspección visual de un perfil (por ejemplo, inspección visual de los niveles de sección genómica). Se puede realizar una identificación por cualquier método adecuado basándose en la totalidad, o en parte, sobre determinaciones, valores y/o datos obtenidos por los métodos descritos en el presente documento, cuyos ejemplos no limitantes incluyen una determinación de ploidía fetal, una determinación de fracción fetal, ploidía materna, determinaciones de incertidumbre y/o confianza, niveles de porción, niveles, perfiles, puntuaciones Z, representaciones cromosómicas esperadas, representaciones cromosómicas medidas, recuentos (por ejemplo, recuentos normalizados, recuentos sin procesar), variaciones en el número de copias fetales o maternas (por ejemplo, variaciones en el número de copias clasificadas), niveles significativamente diferentes, niveles ajustados (por ejemplo, *padding*), similares o combinaciones de los mismos.

En algunos casos, una zona de no identificación es donde no se realiza una identificación. En algunos casos, una zona de no identificación se define por un valor o colección de valores que indican baja precisión, alto riesgo, alto error, bajo nivel de confianza, alto nivel de incertidumbre, similares o una combinación de los mismos. En algunos casos, una zona de no identificación se define, en parte, por una fracción fetal de aproximadamente un 5% o menor, aproximadamente un 4% o menor, aproximadamente un 3% o menor, aproximadamente un 2,5% o menor, aproximadamente un 2,0% o menor, aproximadamente un 1,5% o menor o aproximadamente un 1,0% o menor.

En algunos casos, se realiza un método para determinar la presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal) con una precisión de al menos aproximadamente el 90% a aproximadamente el 100%. Por ejemplo, se puede determinar la presencia o ausencia de una variación genética con una precisión de al menos aproximadamente el 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%,

99,7%, 99,8% o 99,9%. En algunos casos, se determina la presencia o ausencia de una variación genética con una precisión que es aproximadamente igual o mayor que la precisión utilizando otros métodos de determinación de variación genética (por ejemplo, análisis de cariotipo). En algunos casos, se determina la presencia o ausencia de una variación genética con una precisión que tenga un intervalo de confianza (IC) de aproximadamente el 80% a aproximadamente el 100%. Por ejemplo, el intervalo de confianza (IC) puede ser de aproximadamente el 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99%.

A veces, se puede determinar el resultado en términos de densidad de marcadores de secuencia. "Densidad de marcadores de secuencia" se refiere al valor normalizado de marcadores o lecturas de secuencia para una sección genómica definida donde se utiliza la densidad de marcadores de secuencia para comparar diferentes muestras y para análisis posteriores. El valor de la densidad de marcadores de secuencia a menudo se normaliza dentro de una muestra. En algunos casos, la normalización se puede realizar contando el número de marcadores que se encuentran dentro de cada sección genómica; obteniendo un valor mediana del recuento total de marcadores de secuencia para cada cromosoma; obteniendo un valor mediana de todos los valores autosómicos; y utilizando este valor como una constante de normalización para tener en cuenta las diferencias en el número total de marcadores de secuencia obtenido para diferentes muestras. Una densidad de marcadores de secuencia a veces es aproximadamente 1 para un cromosoma disómico. Las densidades de marcadores de secuencia pueden variar de acuerdo con los artefactos de secuenciación, principalmente el sesgo G/C, que puede corregirse usando una referencia patrón externa o interna (por ejemplo, derivada de sustancialmente todos los marcadores de secuencia (secuencias genómicas), que puede ser, por ejemplo, un solo cromosoma o un valor calculado de todos los autosomas, en algunos casos). Por lo tanto, el desequilibrio en la dosis de un cromosoma o regiones cromosómicas se puede inferir a partir de la representación porcentual del locus entre otros marcadores secuenciados mapeables de la muestra. Por lo tanto, puede determinarse cuantitativamente y normalizarse el desequilibrio en la dosis de un cromosoma o regiones cromosómicas particulares. Los métodos para la normalización y cuantificación de la densidad de marcadores de secuencia se explican con mayor detalle a continuación.

En algunos casos, una proporción de todas las lecturas de secuencia es de un cromosoma sexual (por ejemplo, cromosoma X, cromosoma Y) o un cromosoma involucrado en una aneuploidía (por ejemplo, cromosoma 13, cromosoma 18, cromosoma 21), y otras lecturas de secuencia son de otros cromosomas. En algunos casos, al tener en cuenta el tamaño relativo del cromosoma sexual o cromosoma involucrado en la aneuploidía (p. ej., "cromosoma diana": cromosoma 21) en comparación con otros cromosomas, se podría obtener una frecuencia normalizada, dentro de un intervalo de referencia, de secuencias específicas del cromosoma diana. Si el feto tiene una aneuploidía, por ejemplo, en un cromosoma diana, entonces, la frecuencia normalizada de las secuencias derivadas del cromosoma diana es estadísticamente mayor que la frecuencia normalizada de las secuencias derivadas de cromosomas no diana, permitiendo de este modo la detección de la aneuploidía. En algunos casos, el grado de cambio en la frecuencia normalizada dependerá de la concentración fraccional de ácidos nucleicos fetales en la muestra analizada.

A veces, se asocia una variación genética con una afección médica. Un resultado determinante de una variación genética es, a veces, un resultado determinante de la presencia o ausencia de una afección (por ejemplo, una afección médica), enfermedad, síndrome o anomalía, o incluye, la detección de una afección, enfermedad, síndrome o anomalía (por ejemplo, los ejemplos no limitantes enumerados en la Tabla 1). En ciertos casos, un diagnóstico comprende la evaluación de un resultado. Un resultado determinante de la presencia o ausencia de una afección (por ejemplo, una afección médica), enfermedad, síndrome o anomalía mediante los métodos descritos en el presente documento a veces puede verificarse independientemente mediante pruebas adicionales (por ejemplo, mediante determinación del cariotipo y/o amniocentesis). El análisis y procesamiento de datos puede proporcionar uno o más resultados. El término "resultado" como se usa en el presente documento puede referirse a un resultado del procesamiento de datos que facilita la determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía, una variación en el número de copias). En ciertos casos, el término "resultado" como se usa en el presente documento se refiere a una conclusión que predice y/o determina la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía, una variación en el número de copias). En ciertos casos, el término "resultado" como se usa en el presente documento se refiere a una conclusión que predice y/o determina un riesgo o probabilidad de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía, una variación en el número de copias) en un sujeto (p.ej., un feto). A veces, un diagnóstico comprende el uso de un resultado. Por ejemplo, un profesional de la salud puede analizar un resultado y proporcionar un diagnóstico basado en, o basado en parte en, el resultado. En algunos casos, la determinación, detección o diagnóstico de una afección, síndrome o anomalía (por ejemplo, enumerados en la Tabla 1) comprende el uso de un resultado determinante de la presencia o ausencia de una variación genética. En algunos casos, un resultado basado en lecturas de secuencias mapeadas contadas o transformaciones de las mismas es determinante de la presencia o ausencia de una variación genética. En ciertos casos, un resultado generado utilizando uno o más métodos (por ejemplo, métodos de procesamiento de datos) descritos en el presente documento, es determinante de la presencia o ausencia de una o más afecciones, síndromes o anomalías enumerados en la Tabla 1. En ciertos casos, un diagnóstico comprende una determinación de la presencia o ausencia de una afección, síndrome o anomalía. Con frecuencia, un diagnóstico comprende una determinación de una variación genética como la naturaleza y/o la causa de una afección, síndrome o anomalía. En ciertos casos, un resultado no es un diagnóstico. Con frecuencia, un resultado comprende uno o más valores numéricos generados usando un método de procesamiento descrito en el presente documento en el contexto de uno o más estudios de probabilidad. Un estudio de riesgo o probabilidad puede incluir, pero sin limitación: un valor de incertidumbre, una medida de variabilidad, nivel

de confianza, sensibilidad, especificidad, desviación estándar, coeficiente de variación (CV) y/o nivel de confianza, puntuaciones Z, valores de Chi, valores de phi, valores de ploidía, fracción fetal ajustada, relaciones de área, niveles mediana, similares o combinaciones de los mismos. Un estudio de probabilidad puede facilitar determinar si un sujeto está en riesgo de tener, o tiene, una variación genética, y un resultado determinante de una presencia o ausencia de un trastorno genético, con frecuencia incluye dicho estudio.

Un resultado a veces es un fenotipo. Un resultado a veces es un fenotipo con un nivel de confianza asociado (por ejemplo, un valor de incertidumbre, por ejemplo, un feto es positivo para la trisomía 21 con un nivel de confianza del 99%; una gestante lleva un feto masculino con un nivel de confianza del 95 %; un sujeto de ensayo es negativo para un cáncer asociado con una variación genética a un nivel de confianza del 95%). A veces, diferentes métodos para generar valores de resultados pueden producir diferentes tipos de conclusiones. En general, hay cuatro tipos de puntuaciones o identificaciones posibles que se pueden realizar en función de los valores de resultados generados utilizando los métodos descritos en el presente documento: positivo verdadero, positivo falso, negativo verdadero y negativo falso. Los términos "puntuación", "puntuaciones", "identificación" e "identificaciones" tal como se usan en el presente documento se refieren al cálculo de la probabilidad de que una variación genética particular esté presente o ausente en un sujeto/muestra. El valor de una puntuación se puede usar para determinar, por ejemplo, una variación, diferencia o proporción de lecturas de secuencias mapeadas que pueden corresponder a una variación genética. Por ejemplo, el cálculo de una puntuación positiva para una variación o porción genética seleccionada de un conjunto de datos, con respecto a un genoma de referencia puede llevar a una identificación de la presencia o ausencia de una variación genética, variación genética que a veces se asocia con una afección médica (por ejemplo, cáncer, preeclampsia, trisomía, monosomía y similares). En algunos casos, un resultado comprende un nivel, un perfil y/o una representación gráfica (por ejemplo, una representación gráfica de perfil). En aquellos casos en que un resultado comprenda un perfil, se puede usar un perfil adecuado o una combinación de perfiles adecuada para un resultado. Ejemplos no limitantes de perfiles que pueden usarse para un resultado incluyen perfiles de puntuación z, perfiles de valor de p, perfiles de valor de chi, perfiles de valor de phi, similares y combinaciones de los mismos.

Un resultado generado para determinar la presencia o ausencia de una variación genética a veces incluye un resultado nulo (por ejemplo, un punto de datos entre dos grupos, un valor numérico con una desviación estándar que abarca valores tanto de la presencia como de la ausencia de una variación genética, un conjunto de datos con una representación gráfica de perfil que no es similar a las representaciones gráficas de perfil para sujetos que tienen o están libres de la variación genética que se está investigando). En algunos casos, un resultado indicativo de un resultado nulo es un resultado determinante, y la determinación puede incluir la necesidad de información adicional y/o una repetición de la generación de datos y/o análisis para determinar la presencia o ausencia de una variación genética.

En algunos casos, se puede generar un resultado después de realizar una o más etapas de procesamiento descritas en el presente documento. En ciertos casos, se genera un resultado como un resultado de una de las etapas de procesamiento descritas en el presente documento, y en algunos casos, se puede generar un resultado después de que se realice cada manipulación estadística y/o matemática de un conjunto de datos. Un resultado relacionado con la determinación de la presencia o ausencia de una variación genética puede expresarse en una forma adecuada, cuya forma comprende, sin limitación, una probabilidad (por ejemplo, razón de probabilidades (*odds ratio*), valor de p), probabilidad, valor dentro o fuera de grupo, valor por encima o por debajo de un valor umbral, valor dentro de un intervalo (por ejemplo, un intervalo umbral), valor con una medida de varianza o confianza, o factor de riesgo, asociado con la presencia o ausencia de una variación genética para un sujeto o muestra. En ciertos casos, la comparación entre muestras permite la confirmación de la identidad de la muestra (por ejemplo, permite la identificación de muestras repetidas y/o muestras que se han mezclado (por ejemplo, etiquetadas incorrectamente, combinadas y similares)).

En algunos casos, un resultado comprende un valor por encima o por debajo de un umbral o valor de corte predeterminado (por ejemplo, superior a 1, inferior a 1) y un nivel de incertidumbre o confianza asociado con el valor. En ciertos casos, un umbral o valor de corte predeterminado es un nivel esperado o un intervalo de nivel esperado. Un resultado también puede describir una suposición utilizada en el procesamiento de datos. En ciertos casos, un resultado comprende un valor que se encuentra dentro o fuera de un intervalo predeterminado de valores (por ejemplo, un intervalo umbral) y la incertidumbre o nivel de confianza asociados para ese valor están dentro o fuera del intervalo. En algunos casos, un resultado comprende un valor que es igual a un valor predeterminado (por ejemplo, igual a 1, igual a cero), o es igual a un valor dentro de un intervalo de valores predeterminado, y su incertidumbre o nivel de confianza asociados para ese valor es igual o está dentro o fuera de un intervalo. Un resultado a veces se representa gráficamente como una representación gráfica (por ejemplo, una representación gráfica de perfil).

Como se ha señalado anteriormente, un resultado puede caracterizarse como un positivo verdadero, negativo verdadero, positivo falso o negativo falso. La expresión "positivo verdadero", como se usa en el presente documento, se refiere a un sujeto diagnosticado correctamente como que tiene una variación genética. La expresión "positivo falso" como se usa en el presente documento, se refiere a un sujeto identificado erróneamente como que tiene una variación genética. La expresión "negativo verdadero" como se usa en el presente documento, se refiere a un sujeto identificado correctamente como que no tiene una variación genética. La expresión "negativo falso", como se usa en el presente documento, se refiere a un sujeto identificado erróneamente como que no tiene una variación genética. Se pueden calcular dos medidas de rendimiento para cualquier método dado basándose en las proporciones de estos casos: (i)

un valor de sensibilidad, que generalmente es la fracción de positivos pronosticados que se identifican correctamente como positivos; y (ii) un valor de especificidad, que en general es la fracción de negativos pronosticados correctamente identificados como negativos.

5 En ciertos casos, uno o más de sensibilidad, especificidad y/o nivel de confianza, se expresa como un porcentaje. En algunos casos, el porcentaje, independientemente de cada variable, es mayor de aproximadamente 90% (por ejemplo, aproximadamente 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99%, o mayor de 99% (por ejemplo, aproximadamente 99,5% o mayor, aproximadamente 99,9% o mayor, aproximadamente 99,95% o mayor, aproximadamente 99,99% o mayor)). En algunos casos, el coeficiente de variación (CV) se expresa como un porcentaje, y algunas veces el porcentaje es
 10 de aproximadamente el 10% o menor (por ejemplo, aproximadamente 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1%, o menor de 1% (por ejemplo, aproximadamente 0,5% o menor, aproximadamente 0,1 % o menor, aproximadamente 0,05% o menor, aproximadamente 0,01% o menor)). En ciertos casos, una probabilidad (por ejemplo, que un resultado particular no se deba al azar) se expresa como una puntuación Z, un valor de p o los resultados de una prueba t. En algunos casos, se puede generar una varianza medida, intervalo de confianza, sensibilidad, especificidad y similares (por ejemplo,
 15 denominados colectivamente como parámetros de confianza) para un resultado utilizando una o más manipulaciones de procesamiento de datos descritas en el presente documento. Ejemplos específicos de generación de resultados y niveles de confianza asociados se describen en la sección de Ejemplos y en la solicitud de patente internacional n.º PCT/US12/59123 (documento WO2013/052913).

20 El término "sensibilidad", como se usa en el presente documento, se refiere al número de positivos verdaderos dividido por el número de positivos verdaderos más el número de negativos falsos, donde la sensibilidad (sens) puede estar dentro del intervalo de $0 \leq \text{sens} \leq 1$. El término "especificidad" como se usa en el presente documento se refiere al número de negativos verdaderos dividido por el número de negativos verdaderos más el número de positivos falsos, donde la especificidad (espec) puede estar dentro del intervalo de $0 \leq \text{espec} \leq 1$. En algunos casos, a veces se
 25 selecciona un método que tenga sensibilidad y especificidad igual a uno, o el 100%, o cerca de uno (por ejemplo, entre aproximadamente el 90% a aproximadamente el 99%). En algunos casos, se selecciona un método que tiene una sensibilidad igual a 1, o el 100%, y en ciertos casos, se selecciona un método que tiene una sensibilidad cercana a 1 (por ejemplo, una sensibilidad de aproximadamente el 90%, una sensibilidad de aproximadamente el 91%, una sensibilidad de aproximadamente el 92%, una sensibilidad de aproximadamente el 93%, una sensibilidad de aproximadamente el 94%, una sensibilidad de aproximadamente el 95%, una sensibilidad de aproximadamente el 96%, una sensibilidad de aproximadamente el 97%, una sensibilidad de aproximadamente el 98% o una sensibilidad de aproximadamente el 99%). En algunos casos, se selecciona un método que tiene una especificidad igual a 1, o el
 30 100%, y en ciertos casos, se selecciona un método que tiene una especificidad cercana a 1 (por ejemplo, una especificidad de aproximadamente el 90%, una especificidad de aproximadamente el 91%, una especificidad de aproximadamente el 92%, una especificidad de aproximadamente el 93%, una especificidad de aproximadamente el 94%, una especificidad de aproximadamente el 95%, una especificidad de aproximadamente el 96%, una especificidad de aproximadamente el 97%, una especificidad de aproximadamente el 98% o una especificidad de aproximadamente el 99%).

40 Idealmente, el número de negativos falsos es igual a cero o cercano a cero, de modo que ningún sujeto se identifique erróneamente como que no tiene al menos una variación genética cuando en realidad tienen al menos una variación genética. Por el contrario, con frecuencia, se realiza una evaluación de la capacidad de un algoritmo de predicción para clasificar correctamente los negativos, una medida complementaria de la sensibilidad. Idealmente, el número de positivos falsos es igual a cero o cercano a cero, de modo que ningún sujeto se identifique erróneamente como que
 45 tiene al menos una variación genética cuando no tiene la variación genética que se está evaluando.

En algunos casos, se determina la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) para un feto. En dichos casos, se determina la presencia o ausencia de una variación genética (por ejemplo, aneuploidía cromosómica fetal).

50 En ciertos casos, se determina la presencia o ausencia de una variación genética (por ejemplo, aneuploidía cromosómica) para una muestra. En dichos casos, se determina la presencia o ausencia de una variación genética en el ácido nucleico de la muestra (por ejemplo, aneuploidía cromosómica). En algunos casos, una variación detectada o no detectada reside en el ácido nucleico de la muestra de una fuente, pero no en el ácido nucleico de la muestra de otra fuente. Ejemplos no limitantes de fuentes incluyen ácido nucleico placentario, ácido nucleico fetal, ácido nucleico materno, ácido nucleico de células cancerosas, ácido nucleico de células no cancerosas, similares y combinaciones de los mismos. En ejemplos no limitantes, una variación genética particular detectada o no detectada (i) se encuentra en el ácido nucleico placentario pero no en el ácido nucleico fetal y no en el ácido nucleico materno; (ii) se encuentra en el ácido nucleico fetal pero no en el ácido nucleico materno; o (iii) se encuentra en el ácido nucleico materno pero
 60 no en el ácido nucleico fetal.

Después de que se hayan generado uno o más resultados, con frecuencia se utiliza un resultado para proporcionar una determinación de la presencia o ausencia de una variación genética y/o afección médica asociada. Por lo general, se proporciona un resultado a un profesional de la salud (por ejemplo, técnico o gerente de laboratorio; médico o
 65 asistente). Con frecuencia, un módulo de resultados proporciona un resultado. En ciertos casos, un módulo de representación gráfica proporciona un resultado. En ciertos casos, se proporciona un resultado en un periférico o

componente de una máquina. Por ejemplo, a veces una impresora o pantalla proporcionan un resultado. En algunos casos, se proporciona un resultado determinante de la presencia o ausencia de una variación genética a un profesional de la salud en forma de un informe, y en ciertos casos, el informe comprende una presentación de un valor de resultado y un parámetro de confianza asociado. En general, un resultado puede mostrarse en un formato adecuado que facilite la determinación de la presencia o ausencia de una variación genética y/o afección médica. Ejemplos no limitantes de formatos adecuados para su uso para informar y/o mostrar conjuntos de datos o informar un resultado incluyen datos digitales, un gráfico, un gráfico 2D, un gráfico 3D y un gráfico 4D, una imagen, un pictograma, un esquema, un gráfico de barras, un gráfico circular, un diagrama, un diagrama de flujo, un diagrama de dispersión, un mapa, un histograma, un esquema de densidad, un gráfico de funciones, un diagrama de circuito, un diagrama de bloques, un mapa de burbujas, un diagrama de constelaciones, un diagrama de contorno, un cartograma, un esquema de arañas, un diagrama de Venn, un nomograma y similares, y una combinación de los anteriores. Varios ejemplos de representaciones de resultados se muestran en los dibujos y se describen en los Ejemplos.

En ciertos casos, la generación de un resultado puede verse como una transformación de las lecturas de secuencias de ácidos nucleicos en una representación del ácido nucleico celular de un sujeto. Una representación del ácido nucleico celular de un sujeto a menudo refleja una dosis o un número de copias para un cromosoma particular o una porción del mismo, y la representación, por lo tanto, con frecuencia es una propiedad del ácido nucleico del sujeto. Convertir una multitud de lecturas de secuencias relativamente pequeñas en una representación de un cromosoma relativamente grande, por ejemplo, puede verse como una transformación. Como ilustración, en un proceso para generar una representación del cromosoma 21, que tiene una longitud de aproximadamente 47 millones de bases, utilizando lecturas de aproximadamente 36 pares de bases de longitud, se transforman muchos miles de lecturas que son al menos 100.000 veces más pequeñas que el cromosoma en una representación del cromosoma significativamente más grande. La generación de tal representación de un cromosoma generalmente implica varias manipulaciones de lecturas (por ejemplo, mapeo, filtrado y/o normalización) para llegar a una representación del cromosoma relativamente grande, como se describe en el presente documento. Con frecuencia, se utilizan múltiples manipulaciones, lo que puede requerir el uso de uno o más ordenadores, con frecuencia, múltiples ordenadores coordinados en paralelo.

Cuando se proporciona una representación de un cromosoma para un cromosoma fetal utilizando una muestra de una gestante, tal transformación adicional es evidente, debido a que, con frecuencia, la mayoría de las lecturas son de ácido nucleico materno y con frecuencia, una minoría de lecturas son de ácido nucleico fetal. Con frecuencia, las lecturas de ácido nucleico materno dominan las lecturas de ácido nucleico fetal, y, con frecuencia, la mayoría de las lecturas de ácido nucleico materno enmascara una representación de un cromosoma fetal. Un fondo típicamente grande de lecturas maternas puede oscurecer las diferencias entre el ácido nucleico de los cromosomas fetales y maternos y obtener una representación de un cromosoma fetal en un fondo de este tipo implica un proceso que contradice la contribución de las lecturas maternas, como se describe en el presente documento.

En algunos casos, un resultado es el resultado de una transformación de las lecturas de secuencia de un sujeto (por ejemplo, una gestante), en una representación de una estructura existente (por ejemplo, un genoma, cromosoma o segmento del mismo) presente en un sujeto (por ejemplo, una madre y/o feto). En algunos casos, un resultado comprende una transformación de las lecturas de secuencia de un primer sujeto (por ejemplo, una gestante), en una representación compuesta de estructuras (por ejemplo, un genoma, un cromosoma o segmento del mismo), y una segunda transformación de la representación compuesta que produce una representación de una estructura presente en un primer sujeto (por ejemplo, una gestante) y/o en un segundo sujeto (por ejemplo, un feto). En algunos casos, un resultado comprende una transformación de las lecturas de secuencia de un primer sujeto (por ejemplo, un sujeto femenino, una gestante), en una representación de estructuras (por ejemplo, un genoma, cromosoma o segmento del mismo) presente en un segundo sujeto (por ejemplo, un feto).

Un método de transformación en el presente documento, a veces comprende determinar la presencia o ausencia de un cromosoma trisómico (es decir, trisomía cromosómica) en un feto (por ejemplo, T21, T18 y/o T13) de las lecturas de ácidos nucleicos en una muestra obtenida de una sujeto gestante que lleva el feto. En algunos casos, un método de transformación, en el presente documento, puede comprender preparar (p. ej., determinar, visualizar, mostrar, proporcionar) una representación de un cromosoma (p. ej., número de copias de cromosomas, dosis de cromosomas) para un feto a partir de lecturas de ácidos nucleicos en una muestra obtenida de una sujeto gestante que lleva el feto. En los últimos casos, con frecuencia, una representación de un cromosoma para un feto es para el cromosoma 13, el cromosoma 18 y/o el cromosoma 21.

Uso de Resultados

Un profesional de la salud u otra persona cualificada, que reciba un informe que comprenda uno o más resultados determinantes de la presencia o ausencia de una variación genética puede usar los datos mostrados en el informe para hacer una identificación sobre el estado del sujeto o paciente de ensayo. En algunos casos, el profesional de la salud puede hacer una recomendación basándose en el resultado proporcionado. Un profesional de la salud o una persona cualificada puede proporcionar a un sujeto o paciente de ensayo una identificación o puntuación en relación con la presencia o ausencia de la variación genética en función del valor o valores del resultado y los parámetros de confianza asociados proporcionados en un informe. En algunos casos, un profesional de la salud o una persona

cualificada puede realizar una puntuación o identificación de forma manual, utilizando la observación visual del informe proporcionado. En ciertos casos, se hace una puntuación o identificación por medio de una rutina automatizada, algunas veces integrada en el software, y revisada por un profesional de la salud o una persona cualificada para verificar su precisión antes de proporcionar información a un sujeto o paciente de ensayo. La expresión "recibir un informe" como se usa en el presente documento se refiere a la obtención, por medios de comunicación, de una representación escrita y/o gráfica que comprende un resultado, que en una revisión permite a un profesional de la salud u otra persona cualificada determinar la presencia o ausencia de una variación genética en un sujeto o paciente de ensayo. El informe se puede generar mediante un ordenador o mediante la entrada de datos por un ser humano, y puede comunicarse utilizando medios electrónicos (por ejemplo, a través de Internet, por ordenador, por fax, desde una ubicación de red a otra ubicación en el mismo sitio físico o en sitios diferentes), o por otro método de envío o recepción de datos (por ejemplo, servicio de correo, servicio de mensajería y similares). En algunos casos, el resultado se transmite a un profesional de la salud en un medio adecuado, incluyendo, sin limitación, en forma verbal, documental o de archivo. El archivo puede ser, por ejemplo, pero sin limitación, un archivo auditivo, un archivo legible por ordenador, un archivo en papel, un archivo de laboratorio o un archivo de registro médico.

La expresión "proporcionar un resultado" y sus equivalentes gramaticales, como se usa en el presente documento, también puede referirse a un método para obtener dicha información, incluyendo, sin limitación, obtener la información de un laboratorio (por ejemplo, un archivo de laboratorio). Un laboratorio que realizó uno o más ensayos o una o más etapas de procesamiento de datos para determinar la presencia o ausencia de la afección médica, puede generar un archivo de laboratorio. El laboratorio puede estar en la misma ubicación o ubicación diferente (por ejemplo, en otro país) que el personal que identifica la presencia o ausencia de la afección médica del archivo de laboratorio. Por ejemplo, el archivo de laboratorio puede generarse en una ubicación y transmitirse a otra ubicación en donde la información que contiene se transmitirá a la sujeto gestante. En ciertos casos, el archivo de laboratorio puede estar en forma tangible o electrónica (por ejemplo, en forma legible por ordenador).

En algunos casos, se puede proporcionar un resultado a un profesional de la salud, médico o persona cualificada de un laboratorio y el profesional de la salud, médico o individuo cualificado puede hacer un diagnóstico basándose en el resultado. En algunos casos, se puede proporcionar un resultado a un profesional de la salud, médico o persona cualificada de un laboratorio y el profesional de la salud, médico o individuo cualificado puede hacer un diagnóstico basándose, en parte, en el resultado junto con datos y/o información adicionales y otros resultados.

Un profesional de la salud o persona cualificada, puede proporcionar una recomendación adecuada basándose en el resultado o los resultados proporcionados en el informe. Ejemplos no limitantes de recomendaciones que pueden proporcionarse basándose en el informe de resultados proporcionado incluyen cirugía, radioterapia, quimioterapia, asesoramiento genético, soluciones para el tratamiento después del nacimiento (por ejemplo, planificación de la vida, atención asistida a largo plazo, medicamentos, tratamientos sintomáticos), interrupción del embarazo, trasplante de órganos, transfusión de sangre, similares o combinaciones de los anteriores. En algunos casos, la recomendación depende de la clasificación proporcionada basada en el resultado (por ejemplo, síndrome de Down, síndrome de Turner, afecciones médicas asociadas con variaciones genéticas en T13, afecciones médicas asociadas con variaciones genéticas en T18).

El personal de laboratorio (por ejemplo, un gerente de laboratorio) puede analizar valores (por ejemplo, recuentos de ensayo, recuentos de referencia, nivel de desviación) subyacentes a la determinación de la presencia o ausencia de una variación genética (o determinación de euploide o no euploide para una región de ensayo). Para identificaciones relacionadas con la presencia o ausencia de una variación genética cercana o cuestionable, el personal de laboratorio puede volver a solicitar la misma prueba y/o solicitar una prueba diferente (por ejemplo, determinación de cariotipo y/o amniocentesis en el caso de determinaciones de aneuploidía fetal), que hace uso del mismo o diferente ácido nucleico de la muestra de un sujeto de ensayo.

Variaciones Genéticas y Afecciones Médicas

La presencia o ausencia de una varianza genética puede determinarse utilizando un método, aparato o máquina descritos en el presente documento. En ciertos casos, la presencia o ausencia de una o más variaciones genéticas se determina de acuerdo con un resultado proporcionado por los métodos y máquinas descritos en el presente documento. Una variación genética generalmente es un fenotipo genético particular presente en ciertos individuos, y a menudo una variación genética está presente en una subpoblación estadísticamente significativa de individuos. En algunos casos, una variación genética es una anomalía cromosómica (p. ej., aneuploidía), anomalía cromosómica parcial o mosaicismo, cada una de las cuales se describe con mayor detalle en el presente documento. Ejemplos no limitantes de variaciones genéticas incluyen una o más deleciones (por ejemplo, microdeleciones), duplicaciones (por ejemplo, microduplicaciones), inserciones, mutaciones, polimorfismos (p. ej., polimorfismos de un solo nucleótido), fusiones, repeticiones (por ejemplo, repeticiones cortas en tándem), distintos sitios de metilación, distintos patrones de metilación, similares y combinaciones de los mismos. Una inserción, repetición, deleción, duplicación, mutación o polimorfismo puede tener cualquier longitud y, en algunos casos, puede tener una longitud de aproximadamente 1 base o par de bases (pb) a aproximadamente 250 megabases (Mb). En algunos casos, una inserción, repetición, deleción, duplicación, mutación o polimorfismo tiene una longitud de aproximadamente 1 base o par de bases (pb) a aproximadamente 1.000 kilobases (kb) (por ejemplo, una longitud de aproximadamente 10 pb, 50 pb, 100 pb, 500 pb,

1 kb, 5 kb, 10 kb, 50 kb, 100 kb, 500 kb o 1000 kb).

Una variación genética a veces es una delección. En ciertos casos, una delección es una mutación (por ejemplo, una aberración genética) en donde falta una parte de un cromosoma o una secuencia de ADN. Con frecuencia, una delección es la pérdida de material genético. Cualquier número de nucleótidos puede estar eliminado. Una delección puede comprender la delección de uno o más cromosomas completos, un segmento de un cromosoma, un alelo, un gen, un intrón, un exón, cualquier región no codificante, cualquier región codificante, un segmento de los mismos o una combinación de los mismos. Una delección puede comprender una microdelección. Una delección puede comprender la delección de una sola base.

Una variación genética a veces es una duplicación genética. En ciertos casos, una duplicación es una mutación (por ejemplo, una aberración genética) en donde una parte de un cromosoma o una secuencia de ADN se copia y se inserta nuevamente en el genoma. En ciertos casos, una duplicación genética (es decir, la duplicación) es cualquier duplicación de una región del ADN. En algunos casos, una duplicación es una secuencia de ácido nucleico que se repite, con frecuencia en tándem, dentro de un genoma o cromosoma. En algunos casos, una duplicación puede comprender la delección de uno o más cromosomas completos, un segmento de un cromosoma, un alelo, un gen, un intrón, un exón, cualquier región no codificante, cualquier región codificante, un segmento de los mismos o una combinación de los mismos. Una duplicación puede comprender una microduplicación. Una duplicación a veces comprende una o más copias de un ácido nucleico duplicado. Una duplicación a veces se caracteriza como una región genética repetida una o más veces (por ejemplo, repetida 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 veces). Las duplicaciones pueden abarcar desde pequeñas regiones (miles de pares de bases) hasta cromosomas completos en algunos casos. Las duplicaciones ocurren con frecuencia como resultado de un error en la recombinación homóloga o debido a un evento de retrotransposón. Las duplicaciones se han asociado con ciertos tipos de enfermedades proliferativas. Las duplicaciones pueden caracterizarse mediante micromatrices genómicas o hibridación genética comparativa (CGH).

Una variación genética a veces es una inserción. Una inserción a veces es la adición de uno o más pares de bases de nucleótidos en una secuencia de ácido nucleico. Una inserción a veces es una microinserción. En ciertos casos, una inserción comprende la adición de un segmento de un cromosoma en un genoma, cromosoma o segmento del mismo. En ciertos casos, una inserción comprende la adición de un alelo, un gen, un intrón, un exón, cualquier región no codificante, cualquier región codificante, un segmento de los mismos o una combinación de los mismos en un genoma o segmento del mismo. En ciertos casos, una inserción comprende la adición (es decir, inserción) de un ácido nucleico de origen desconocido en un genoma, cromosoma o segmento del mismo. En ciertos casos, una inserción comprende la adición (es decir, la inserción) de una sola única.

Como se usa en el presente documento, una "variación en el número de copias" generalmente es una clase o tipo de variación genética o aberración cromosómica. Una variación en el número de copias puede ser una delección (por ejemplo, una microdelección), una duplicación (por ejemplo, una microduplicación) o una inserción (por ejemplo, una microinserción). Con frecuencia, el prefijo "micro", como se usa en el presente documento, a veces es un segmento de ácido nucleico menor de 5 Mb de longitud. Una variación en el número de copias puede incluir una o más delecciones (por ejemplo, microdelección), duplicaciones y/o inserciones (por ej., microduplicación, microinserción de un segmento de un cromosoma). En ciertos casos, una duplicación comprende una inserción. En ciertos casos, una inserción es una duplicación. En ciertos casos, una inserción no es una duplicación. Por ejemplo, con frecuencia, la duplicación de una secuencia en una porción aumenta los recuentos de una porción en donde se encuentra la duplicación. Con frecuencia, una duplicación de una secuencia en una porción aumenta el nivel. En ciertos casos, una duplicación presente en porciones que forman un primer nivel aumenta el nivel en relación con un segundo nivel donde no hay duplicación. En ciertos casos, una inserción aumenta los recuentos de una porción y una secuencia que representa la inserción está presente (es decir, duplicada) en otra ubicación dentro de la misma porción. En ciertos casos, una inserción no aumenta significativamente los recuentos de una porción o nivel y la secuencia que se inserta no es una duplicación de una secuencia dentro de la misma porción. En ciertos casos, una inserción no se detecta o representa como una duplicación y una secuencia duplicada que representa la inserción no está presente en la misma porción.

En algunos casos, una variación en el número de copias es una variación en el número de copias fetales. Con frecuencia, una variación en el número de copias fetales es una variación en el número de copias en el genoma de un feto. En algunos casos, una variación en el número de copias es una variación en el número de copias maternas y/o fetales. En algunos casos, una variación en el número de copias maternas y/o fetales es una variación en el número de copias dentro del genoma de una gestante (por ejemplo, una mujer que tiene un feto), una mujer que dio a luz o una mujer capaz de tener un feto. Una variación en el número de copias puede ser una variación en el número de copias heterocigótica cuando la variación (por ejemplo, una duplicación o delección) está presente en un alelo de un genoma. Una variación en el número de copias puede ser una variación en el número de copias homocigótica cuando la variación está presente en ambos alelos de un genoma. En algunos casos, una variación en el número de copias es una variación en el número de copias fetales heterocigótica u homocigótica. En algunos casos, una variación en el número de copias es una variación en el número de copias materna y/o fetal heterocigótica u homocigótica. Una variación en el número de copias a veces está presente en un genoma materno y en un genoma fetal, en un genoma materno y no en un genoma fetal, o en un genoma fetal y no en un genoma materno.

"Ploidía" es una referencia al número de cromosomas presentes en un feto o madre. En ciertos casos, "ploidía" es lo

mismo que "ploidía cromosómica". En los seres humanos, por ejemplo, los cromosomas autosómicos a menudo se presentan en pares. Por ejemplo, en la ausencia de una variación genética, la mayoría de los seres humanos tienen dos de cada cromosoma autosómico (por ejemplo, los cromosomas 1-22). La presencia del complemento normal de 2 cromosomas autosómicos en un ser humano a menudo se denomina euploide. "Microploidía" es similar en significado a ploidía. Con frecuencia, "Microploidía" se refiere a la ploidía de un segmento de un cromosoma. El término "microploidía" a veces es una referencia a la presencia o ausencia de una variación en el número de copias (por ejemplo, una delección, duplicación y/o una inserción) dentro de un cromosoma (por ejemplo, una delección, duplicación o inserción homocigótica o heterocigótica, similares o combinaciones de las mismas). La "ploidía" y "microploidía" a veces se determinan después de la normalización de los recuentos de un nivel en un perfil. Por lo tanto, un nivel que representa un par de cromosomas autosómicos (por ejemplo, un euploide) a menudo se normaliza a una ploidía de 1. De manera similar, un nivel dentro de un segmento de un cromosoma que representa la ausencia de una duplicación, delección o inserción a menudo se normaliza a una microploidía de 1. La ploidía y la microploidía a menudo son específicas de porción (por ejemplo, específicas de porción) y específicas de muestra. La ploidía a menudo se define como múltiplos enteros de $\frac{1}{2}$, con los valores de 1, $\frac{1}{2}$, 0, $\frac{3}{2}$ y 2 que representan euploide (p. ej., 2 cromosomas), 1 cromosoma presente (p. ej., una delección cromosómica), sin cromosoma presentes, 3 cromosomas (por ejemplo, una trisomía) y 4 cromosomas, respectivamente. De manera análoga, la microploidía a menudo se define como múltiplos enteros de $\frac{1}{2}$, con los valores de 1, $\frac{1}{2}$, 0, $\frac{3}{2}$ y 2 que representan euploide (p. ej., sin variación en el número de copias), una delección heterocigótica, delección homocigótica, duplicación heterocigótica y duplicación homocigótica, respectivamente. En la Tabla 2 se proporcionan algunos ejemplos de valores de ploidía para un feto.

En ciertos casos, la microploidía de un feto coincide con la microploidía de la madre del feto (es decir, la mujer gestante). En ciertos casos, la microploidía de un feto coincide con la microploidía de la madre del feto y tanto la madre como el feto portan la misma variación en el número de copias heterocigótica, variación en el número de copias homocigótica o ambos son euploides. En ciertos casos, la microploidía de un feto es diferente de la microploidía de la madre del feto. Por ejemplo, a veces la microploidía de un feto es heterocigótica para una variación en el número de copias, la madre es homocigótica para una variación en el número de copias y la microploidía del feto no coincide (por ejemplo, no es igual) con la microploidía de la madre para la variación en el número de copias.

Con frecuencia, se asocia una microploidía con un nivel esperado. Por ejemplo, a veces un nivel (por ejemplo, un nivel en un perfil, a veces un nivel que no incluye sustancialmente ninguna variación en el número de copias) se normaliza a un valor de 1 (por ejemplo, una ploidía de 1, una microploidía de 1) y la microploidía de una duplicación homocigótica es 2, una duplicación heterocigótica es 1,5, una delección heterocigótica es 0,5 y una delección homocigótica es cero.

En ciertos casos, una variación genética para la cual se identifica la presencia o ausencia para un sujeto está asociada con una afección médica. Por lo tanto, la tecnología descrita, en el presente documento, se puede usar para identificar la presencia o ausencia de una o más variaciones genéticas asociadas con una afección médica o estado médico. Ejemplos no limitantes de afecciones médicas incluyen aquellas asociadas con discapacidad intelectual (por ejemplo, Síndrome de Down), proliferación celular aberrante (por ejemplo, cáncer), presencia de un ácido nucleico de microorganismos (por ejemplo, virus, bacteria, hongos, levaduras), y preeclampsia.

Se describen, a continuación, ejemplos no limitantes de variaciones genéticas, afecciones y estados médicos.

Sexo Fetal

En algunos casos, la predicción de un sexo fetal o de un trastorno relacionado con el sexo fetal (por ejemplo, la aneuploidía de cromosomas sexuales), se puede determinar mediante un método o una máquina descrito en el presente documento. La determinación del sexo generalmente se basa en un cromosoma sexual. En los seres humanos, hay dos cromosomas sexuales, los cromosomas X e Y. El cromosoma Y contiene un gen, SRY, que desencadena el desarrollo embrionario como varón. Los cromosomas Y de los seres humanos y otros mamíferos también contienen otros genes necesarios para la producción normal de esperma. Los individuos con XX son mujeres y con XY son varones y variaciones no limitantes, a menudo denominadas aneuploidías de cromosomas sexuales, incluyen X0, XYY, XXX y XXY. En ciertos casos, los varones tienen dos cromosomas X y un cromosoma Y (XXY; síndrome de Klinefelter), o un cromosoma X y dos cromosomas Y (síndrome XYY; síndrome de Jacobs), y algunas mujeres tienen tres cromosomas X (XXX; síndrome de Triple X) o un solo cromosoma X en lugar de dos (X0; síndrome de Turner). En ciertos casos, solo una porción de las células en un individuo se ve afectada por una aneuploidía de los cromosomas sexuales que se puede denominar mosaicismo (por ejemplo, mosaicismo de Turner). Otros casos incluyen aquellos en los que SRY está dañado (lo que da lugar a una mujer XY), o está copiado en el X (lo que da lugar a un varón XX).

En algunos casos, un método en donde se determina el sexo fetal también puede comprender determinar la fracción fetal y/o la presencia o ausencia de una variación genética fetal (por ejemplo, aneuploidía cromosómica fetal). La determinación de la presencia o ausencia de una variación genética fetal se puede realizar de una manera adecuada, cuyos ejemplos no limitantes incluyen análisis del cariotipo, amniocentesis, análisis de ácido nucleico circulante libre de células, análisis de ADN fetal libre de células, análisis de secuencia de nucleótidos, cuantificación de lecturas de secuencias, aproximaciones dirigidas, aproximaciones basadas en amplificación, aproximaciones basadas en espectrometría de masas, aproximaciones basadas en metilación diferencial, aproximaciones basadas en digestión

diferencial, aproximaciones basadas en polimorfismo, aproximaciones basadas en hibridación (por ejemplo, utilizando sondas) y similares.

En determinados casos, puede ser beneficioso determinar el sexo de un feto en el útero. Por ejemplo, una paciente (por ejemplo, una gestante) con antecedentes familiares de uno o más trastornos ligados al sexo puede desear determinar el sexo del feto que lleva para ayudar a evaluar el riesgo de que el feto herede tal trastorno. Los trastornos ligados al sexo incluyen, sin limitación, trastornos ligados a X y ligados a Y. Los trastornos ligados a X incluyen trastornos recesivos ligados a X y trastornos dominantes ligados a X. Ejemplos de trastornos recesivos ligados a X incluyen, sin limitación, trastornos inmunitarios (por ejemplo, enfermedad granulomatosa crónica (CYBB), síndrome de Wiskott-Aldrich, inmunodeficiencia combinada grave ligada a X, agammaglobulinemia ligada a X, síndrome de hiper-IgM tipo 1, IPEX, Enfermedad linfoproliferativa ligada a X, deficiencia de Properdin), trastornos hematológicos (por ejemplo, Hemofilia A, Hemofilia B, anemia sideroblástica ligada a X), trastornos endocrinos (por ejemplo, síndrome de insensibilidad a los andrógenos/enfermedad de Kennedy, síndrome de Kallmann KAL1, hipoplasia suprarrenal congénita ligada a X), trastornos metabólicos (por ejemplo, deficiencia de ornitina transcarbamilasa, síndrome de oculocerebrorenal, adrenoleucodistrofia, deficiencia de glucosa-6-fosfato deshidrogenasa, deficiencia de piruvato deshidrogenasa, enfermedad de Danon/enfermedad por almacenamiento de glucógeno Tipo IIb, enfermedad de Fabry, síndrome de Hunter, síndrome de Lesch-Nyhan, enfermedad de Menkes/síndrome del cuerno occipital), trastornos del sistema nervioso (por ejemplo, síndrome de Coffin-Lowry, Síndrome MASA, síndrome alfa-talasemia con retraso mental ligado a X, síndrome de Siderio con retraso mental ligado a X, daltonismo, albinismo ocular, enfermedad de Norrie, coroideremia, Enfermedad de Charcot-Marie-Tooth (CMTX2-3), Enfermedad de Pelizaeus-Merzbacher, SMAX2), trastornos de la piel y tejidos relacionados (p. ej., disqueratosis congénita, displasia ectodérmica hipohidrótica (EDA), ictiosis ligada a X, distrofia corneal endotelial ligada a X), trastornos neuromusculares (por ejemplo, distrofia muscular de Becker/Duchenne, miopatía centronuclear (MTM1), síndrome de Conradi-Hünemann, distrofia muscular de Emery-Dreifuss 1), trastornos urológicos (por ejemplo, síndrome de Alport, enfermedad de Dent, Diabetes insípida nefrogénica ligada a X), trastornos óseos/dentales (por ejemplo, AMELX Amelogenesis imperfecta) y otros trastornos (por ejemplo, síndrome de Barth, síndrome de McLeod, Síndrome de Smith-Fineman-Myers, síndrome de Simpson-Golabi-Behmel, síndrome de Mohr-Tranebjaerg, síndrome Nasodigitoacústico). Ejemplos de trastornos dominantes ligados a X incluyen, sin limitación, hipofosfatemia ligada a X, Hipoplasia dérmica focal, síndrome de X frágil, síndrome de Aicardi, Incontinencia pigmentaria, síndrome de Rett, Síndrome CHILD, síndrome de Lujan-Fryns y síndrome orofaciadigital 1. Ejemplos de trastornos ligados a Y incluyen, sin limitación, infertilidad masculina, retinitis pigmentosa, y azoospermia.

Anomalías Cromosómicas

En algunos casos, la presencia o ausencia de una anomalía cromosómica fetal puede determinarse utilizando un método o una máquina descrito en el presente documento. Las anomalías cromosómicas incluyen, sin limitación, una ganancia o pérdida de un cromosoma completo o una región de un cromosoma que comprende uno o más genes. Las anomalías cromosómicas incluyen monosomías, trisomías, polisomías, pérdida de heterocigosidad, translocaciones, deleciones y/o duplicaciones de una o más secuencias de nucleótidos (por ejemplo, uno o más genes), incluidas las deleciones y duplicaciones causadas por translocaciones desequilibradas. La expresión "anomalía cromosómica", "aneuploidía" y/o "aneuploide", como se usa en el presente documento, se refiere a una desviación entre la estructura del cromosoma en cuestión y un cromosoma homólogo normal. El término "normal" se refiere al cariotipo predominante o al patrón de bandas que se encuentra en individuos sanos de una especie en particular, por ejemplo, un genoma euploide (en seres humanos, 46,XX o 46,XY). Como diferentes organismos tienen complementos cromosómicos muy variados, el término "aneuploidía" y "aneuploide" no se refiere a un número concreto de cromosomas, sino a la situación en donde el contenido de cromosomas dentro de una célula o células dadas de un organismo es anormal. En algunos casos, el término "aneuploidía" y "aneuploide" en el presente documento se refiere a un desequilibrio del material genético causado por una pérdida o ganancia de un cromosoma completo, o parte de un cromosoma. Una "aneuploidía" puede referirse a una o más deleciones y/o inserciones de un segmento de un cromosoma. El término "euploide" puede referirse a un complemento normal de cromosomas.

El término "monosomía" como se usa en el presente documento se refiere a la carencia de un cromosoma del complemento normal. La monosomía parcial puede producirse en translocaciones o deleciones desequilibradas, en las que solo un segmento del cromosoma está presente en una sola copia. La monosomía de los cromosomas sexuales (45, X) causa el síndrome de Turner, por ejemplo. El término "disomía" se refiere a la presencia de dos copias de un cromosoma. Para organismos como los seres humanos que tienen dos copias de cada cromosoma (aquellos que son diploides o "euploides"), la disomía es la condición normal. Para los organismos que normalmente tienen tres o más copias de cada cromosoma (aquellos que son triploides o superiores), la disomía es un estado cromosómico aneuploide. En la disomía uniparental, ambas copias de un cromosoma provienen del mismo progenitor (sin la contribución del otro progenitor).

El término "trisomía", como se usa en el presente documento, se refiere a la presencia de tres copias, en lugar de dos copias, de un cromosoma particular. La presencia de un cromosoma 21 adicional, que se encuentra en el síndrome de Down humano, se conoce como "Trisomía 21". La Trisomía 18 y la Trisomía 13 son otras dos trisomías autosómicas humanas. La trisomía de los cromosomas sexuales se puede ver en mujeres (por ejemplo, 47, XXX en el síndrome de Triple X) o en varones (por ejemplo, 47, XXY en el síndrome de Klinefelter; o 47, XYY en el síndrome de Jacobs). En

algunos casos, una trisomía es una duplicación de la mayoría o la totalidad de un autosoma. En ciertos casos, una trisomía es una aneuploidía de un cromosoma completo que da como resultado tres apariciones (por ejemplo, tres copias) de un tipo particular de cromosoma (por ejemplo, en lugar de dos apariciones (es decir, un par) de un tipo particular de cromosoma para un euploide).

5 Los términos "tetrasomía" y "pentasomía" como se usan en el presente documento se refieren a la presencia de cuatro o cinco copias de un cromosoma, respectivamente. Aunque rara vez se observa con autosomas, se ha informado de tetrasomía y pentasomía de cromosomas sexuales en seres humanos, incluyendo XXXX, XXXY, XXYY, XYYY, XXXXX, XXXXY, XXXYY, XYYYY y XYYYYY.

10 Las anomalías cromosómicas pueden estar causadas por una variedad de mecanismos. Los mecanismos incluyen, pero sin limitación (i) no disyunción que se produce como resultado de un punto de control mitótico debilitado, (ii) puntos de control mitóticos inactivos que causan la no disyunción en múltiples cromosomas, (iii) unión merotética que se produce cuando un cinetocoro está unido a ambos polos de los husos mitóticos, (iv) un huso multipolar que se forma cuando se forman más de dos polos de husos, (v) un huso monopolar que se forma cuando solo se forma un polo de huso único, y (vi) un intermedio tetraploide que se produce como resultado final del mecanismo del huso monopolar.

20 El término "monosomía parcial" y "trisomía parcial" como se usa en el presente documento, se refiere a un desequilibrio del material genético causado por una pérdida o ganancia de parte de un cromosoma. Una monosomía parcial o trisomía parcial puede ser el resultado de una translocación desequilibrada, donde un individuo porta un cromosoma derivado formado a través de la rotura y fusión de dos cromosomas diferentes. En esta situación, el individuo tendría tres copias de parte de un cromosoma (dos copias normales y el segmento que existe en el cromosoma derivado) y solo una copia de parte del otro cromosoma involucrado en el cromosoma derivado.

25 El término "mosaicismo", como se usa en el presente documento, se refiere a aneuploidía en algunas células, pero no en todas las células, de un organismo. Ciertas anomalías cromosómicas pueden existir como anomalías cromosómicas en mosaico y no mosaico. Por ejemplo, ciertos individuos con trisomía 21 tienen síndrome de Down en mosaico y algunos tienen síndrome de Down no mosaico. Diferentes mecanismos pueden conducir al mosaicismo. Por ejemplo, (i) un cigoto inicial puede tener tres cromosomas 21, lo que normalmente daría como resultado una trisomía 21 simple, pero durante el curso de la división celular una o más líneas celulares perdieron uno de los cromosomas 21; y (ii) un cigoto inicial puede tener dos cromosomas 21, pero durante el curso de la división celular, se duplicó uno de los cromosomas 21. El mosaicismo somático probablemente se produce a través de mecanismos distintos de los típicamente asociados con los síndromes genéticos que involucran la aneuploidía completa o en mosaico. El mosaicismo somático se ha identificado en ciertos tipos de cáncer y en neuronas, por ejemplo. En ciertos casos, se identificó la trisomía 12 en la leucemia linfocítica crónica (CLL) y se identificó la trisomía 8 en la leucemia mieloide aguda (AML). Además, los síndromes genéticos en los que un individuo está predispuesto a la ruptura de los cromosomas (síndromes de inestabilidad cromosómica) se asocian con frecuencia con un mayor riesgo de diversos tipos de cáncer, destacando, de este modo, el papel de la aneuploidía somática en la carcinogénesis. Los métodos y protocolos descritos en el presente documento, pueden identificar la presencia o ausencia de anomalías cromosómicas en no mosaico y en mosaico.

45 Las tablas 1A y 1B presentan una lista no limitante de afecciones cromosómicas, síndromes y/o anomalías que pueden identificarse potencialmente mediante los métodos y las máquinas descritos en el presente documento. La tabla 1B es de la base de datos DECIPHER al 6 de octubre de 2011 (por ejemplo, versión 5.1, basada en las posiciones mapeadas en GRCh37; disponible en el localizador uniforme de recursos (URL) decipher.sanger.ac.uk).

Tabla 1A

Cromosoma	Anomalía	Asociación de Enfermedad
X	XO	Síndrome de Turner
Y	XXY	Síndrome de Klinefelter
Y	XYY	Síndrome de doble Y
Y	XXX	Síndrome de trisomía X
Y	XXXX	Síndrome de cuatro X
Y	Delección Xp21	Síndrome de Duchenne/Becker, hiperplasia suprarrenal congénita, enfermedad granulomatosa crónica
Y	Delección Xp22	deficiencia de sulfatasa esteroidea

(continuación)

Cromosoma	Anomalía	Asociación de Enfermedad
Y	Deleción Xq26	enfermedad linfoproliferativa ligada a X
1	monosomía trisomía 1 p (somática)	neuroblastoma
2	monosomía trisomía 2q	retraso del crecimiento, retraso mental y del desarrollo, y anomalías físicas menores
3	monosomía trisomía (somática)	Linfoma no-Hodgkin
4	monosomía trisomía (somática)	Leucemia no-linfocítica aguda (ANLL)
5	5p	Mauilido de gato; síndrome de Lejeune
5	monosomía trisomía 5q (somática)	síndrome mielodisplásico
6	monosomía trisomía (somática)	sarcoma de células claras
7	Deleción 7q11.23	síndrome de William
7	monosomía trisomía	síndrome de monosomía 7 de la infancia; somática: adenomas corticales renales; síndrome mielodisplásico
8	Deleción 8q24.1	síndrome de Langer-Giedon
8	monosomía trisomía	síndrome mielodisplásico; síndrome de Warkany; somática: leucemia mielógena crónica
9	monosomía 9p	síndrome de Alfi
9	monosomía trisomía parcial 9p	síndrome de Rethoré
9	trisomía	síndrome de trisomía 9 completa; síndrome de trisomía 9 en mosaico
10	Monosomía trisomía (somática)	ALL o ANLL
11	11p-	Aniridia; tumor de Wilms
11	11q-	Síndrome de Jacobsen
11	monosomía (somática) trisomía	linajes mieloides afectados (ANLL, MDS)
12	monosomía trisomía (somática)	CLL, Tumor de células granulosas juvenil (TCGJ)
13	13q-	síndrome 13q; síndrome de Orbeli
13	Deleción 13q14	retinoblastoma
13	monosomía trisomía	síndrome de Patau
14	monosomía trisomía (somática)	trastornos mieloides (MDS, ANLL, CML atípico)
15	monosomía de la deleción 15q11-q13	síndrome de Prader-Willi, síndrome de Angelman
15	trisomía (somática)	linajes mieloides y linfoides afectados, por ejemplo, MDS, ANLL, ALL, CLL)
16	Deleción 16q13.3	Rubenstein-Taybi
3	monosomía trisomía (somática)	carcinomas de células renales papilares (malignos)
17	17p-(somática)	síndrome 17p en neoplasias mieloides
17	Deleción 17q11.2	Smith-Magenis
17	17q13.3	Miller-Dieker
17	monosomía trisomía (somática)	adenomas corticales renales

(continuación)

Cromosoma	Anomalía	Asociación de Enfermedad
17	trisomía 17p11.2 - 12	Enfermedad de Charcot-Marie-Tooth tipo 1; HNPP
18	18p-	Síndrome de monosomía parcial 18p o síndrome de Grouchy Lamy Thieffry
18	18q-	síndrome de Grouchy Lamy Salmon Landry
18	monosomía trisomía	Síndrome de Edwards
19	monosomía trisomía	
20	20p-	Síndrome de trisomía 20p
20	delección 20p11.2 - 12	Alagille
20	20q-	somática: MDS, ANLL, policitemia vera, leucemia neutrofilica crónica
20	monosomía trisomía (somática)	carcinomas de células renales papilares (malignos)
21	monosomía trisomía	síndrome de Down
22	Delección 22q11.2	síndrome de DiGeorge, síndrome velocardiofacial, síndrome de la cara con anomalía conotruncal, síndrome de Opitz G/BBB autosómico dominante, síndrome cardiofacial de Caylor
22	monosomía trisomía	síndrome de trisomía 22 completa

Tabla 1B

Síndrome	Cromosoma	Inicio	Final	Intervalo (Mb)	Grado
Síndrome de microdelección 12q14	12	65.071.919	68.645.525	3,57	
Síndrome de microdelección 15q13.3	15	30.769.995	32.701.482	1,93	
Síndrome de microdelección 15q24 recurrente	15	74.377.174	76.162.277	1,79	
Síndrome de sobrecrecimiento 15q26	15	99.357.970	102.521.392	3,16	
Síndrome de microduplicación 16p11.2	16	29.501.198	30.202.572	0,70	
Síndrome de microdelección 16p11.2-p12.2	16	21.613.956	29.042.192	7,43	
Microdelección 16p13.11 recurrente (locus de susceptibilidad al trastorno neurocognitivo)	16	15.504.454	16.284.248	0,78	
Microduplicación 16p13.11 recurrente (locus de susceptibilidad al trastorno neurocognitivo)	16	15.504.454	16.284.248	0,78	
Síndrome de microdelección 17q21.3 recurrente	17	43.632.466	44.210.205	0,58	1
Síndrome de microdelección 1p36	1	10.001	5.408.761	5,40	1
Microdelección 1q21.1 recurrente (locus de susceptibilidad para trastornos del desarrollo neurológico)	1	146.512.930	147.737.500	1,22	3
Microduplicación 1q21.1 recurrente (posible locus de susceptibilidad para trastornos del desarrollo neurológico)	1	146.512.930	147.737.500	1,22	3
Locus 1q21.1 de susceptibilidad para el síndrome de trombocitopenia con ausencia de radios (TAR)	1	145.401.253	145.928.123	0,53	3
Síndrome de delección 22q11 (síndrome velocardiofacial/de DiGeorge)	22	18.546.349	22.336.469	3,79	1
Síndrome de duplicación 22q11	22	18.546.349	22.336.469	3,79	3
Síndrome de delección 22q11.2 distal	22	22.115.848	23.696.229	1,58	
Síndrome de delección 22q13 (síndrome de Phelan-Mcdermid)	22	51.045.516	51.187.844	0,14	1
Síndrome de microdelección 2p15-16.1	2	57.741.796	61.738.334	4,00	

(continuación)

Síndrome	Cromosoma	Inicio	Final	Intervalo (Mb)	Grado
Síndrome de delección 2q33.1	2	196.925.089	205.206.940	8,28	1
Monosomía 2q37	2	239.954.693	243.102.476	3,15	1
Síndrome de microdelección 3q29	3	195.672.229	197.497.869	1,83	
Síndrome de microduplicación 3q29	3	195.672.229	197.497.869	1,83	
Síndrome de duplicación 7q11.23	7	72.332.743	74.616.901	2,28	
Síndrome de delección 8p23.1	8	8.119.295	11.765.719	3,65	
Síndrome de delección 9q subtelomérica	9	140.403.363	141.153.431	0,75	1
Leucodistrofia autosómica dominante (ADLD) de inicio en el adulto	5	126.063.045	126.204.952	0,14	
Síndrome de Angelman (Tipo 1)	15	22.876.632	28.557.186	5,68	1
Síndrome de Angelman (Tipo 2)	15	23.758.390	28.557.186	4,80	1
Síndrome ATR-16	16	60.001	834.372	0,77	1
AZFa	Y	14.352.761	15.154.862	0,80	
AZFb	Y	20.118.045	26.065.197	5,95	
AZFb+AZFc	Y	19.964.826	27.793.830	7,83	
AZFc	Y	24.977.425	28.033.929	3,06	
Síndrome de ojo de gato (Tipo I)	22	1	16.971.860	16,97	
Síndrome de Charcot-Marie-Tooth tipo 1A (CMT1A)	17	13.968.607	15.434.038	1,47	1
Síndrome de maullido de gato (delección 5p)	5	10.001	11.723.854	11,71	1
Enfermedad de Alzheimer de inicio temprano con angiopatía amiloide cerebral	21	27.037.956	27.548.479	0,51	
Poliposis Adenomatosa Familiar	5	112.101.596	112.221.377	0,12	
Neuropatía Hereditaria con Susceptibilidad a la Parálisis por Presión (HNPP)	17	13.968.607	15.434.038	1,47	1
Discondroostosis de Leri-Weill (LWD)- delección SHOX	X	751.878	867.875	0,12	
Discondroostosis de Leri-Weill (LWD)- delección SHOX	X	460.558	753.877	0,29	
Síndrome de Miller-Dieker (MDS)	17	1	2.545.429	2,55	1
Síndrome de microdelección NF1	17	29.162.822	30.218.667	1,06	1
Enfermedad de Pelizaeus-Merzbacher	X	102.642.051	103.131.767	0,49	
Síndrome de Potocki-Lupski (síndrome de duplicación 17p11.2)	17	16.706.021	20.482.061	3,78	
Síndrome de Potocki-Shaffer	11	43.985.277	46.064.560	2,08	1
Síndrome de Prader-Willi (Tipo 1)	15	22.876.632	28.557.186	5,68	1
Prader-Willi (Tipo 2)	15	23.758.390	28.557.186	4,80	1
RCAD (quistes renales y diabetes)	17	34.907.366	36.076.803	1,17	
Rubinstein-Taybi	16	3.781.464	3.861.246	0,08	1
Smith-Magenis	17	16.706.021	20.482.061	3,78	1
Síndrome de Sotos	5	175.130.402	177.456.545	2,33	1
Malformación 1 de mano/pie hendidos (SHFM1)	7	95.533.860	96.779.486	1,25	
Deficiencia de sulfatasa esteroidea (STS)	X	6.441.957	8.167.697	1,73	
Síndrome de WARG de delección 11p13	11	31.803.509	32.510.988	0,71	
Williams-Beuren (WBS)	7	72.332.743	74.616.901	2,28	1
Wolf-Hirschhorn	4	10.001	2.073.670	2,06	1
Duplicación Xq28 (MECP2)	X	152.749.900	153.390.999	0,64	

Con frecuencia, las afecciones de grado 1 tienen una o más de las siguientes características: anomalía patógena; fuerte acuerdo entre los genetistas; altamente penetrante; aún puede tener un fenotipo variable pero algunas

características comunes; todos los casos en la literatura tienen un fenotipo clínico; no hay casos de individuos sanos con la anomalía; no se informa en las bases de datos de DVG ni se encuentra en una población sana; datos funcionales que confirman el efecto de la dosis de un solo gen o de múltiples genes; genes candidatos confirmados o sólidos; implicaciones de manejo clínico definidas; riesgo conocido de cáncer con implicación para la supervisión; múltiples fuentes de información (OMIM, GeneReviews, Orphaned, Unique, Wikipedia); y/o están disponibles para uso diagnóstico (asesoramiento reproductivo).

Con frecuencia, las afecciones de grado 2 tienen una o más de las siguientes características: probable anomalía patógena; altamente penetrante; fenotipo variable sin características consistentes distintas de la DD; pequeño número de casos/informes en la literatura; todos los casos informados tienen un fenotipo clínico; no hay datos funcionales ni genes patógenos confirmados; múltiples fuentes de información (OMIM, GeneReviews, Orphanet, Unique, Wikipedia); y/o se pueden utilizar para fines de diagnóstico y asesoramiento reproductivo.

Con frecuencia, las afecciones de grado 3 tienen una o más de las siguientes características: locus de susceptibilidad; individuos sanos o progenitores no afectados de un probando descrito; presente en poblaciones de control; no penetrante; fenotipo leve y no específico; características menos consistentes; no hay datos funcionales ni genes patógenos confirmados; fuentes de datos más limitadas; la posibilidad de un segundo diagnóstico sigue siendo una posibilidad para los casos que se desvían de la mayoría o si se encuentran nuevos hallazgos clínicos; y/o precaución cuando se utilizan con fines diagnósticos y consejos precisos para el asesoramiento reproductivo.

Preeclampsia

En algunos casos, la presencia o ausencia de preeclampsia se determina utilizando un método o una máquina descrito en el presente documento. La preeclampsia es una afección en donde surge la hipertensión en el embarazo (es decir, la hipertensión inducida por el embarazo) y se asocia con cantidades significativas de proteínas en la orina. En ciertos casos, la preeclampsia también se asocia con niveles elevados de ácido nucleico extracelular y/o alteraciones en los patrones de metilación. Por ejemplo, se ha observado una correlación positiva entre los niveles de RASSF1A extracelular hipermetilado derivado del feto y la gravedad de la preeclampsia. En algunos ejemplos, se observa un aumento de la metilación del ADN para el gen H19 en placentas preeclámpticas en comparación con los controles normales.

La preeclampsia es una de las principales causas de mortalidad y morbilidad materna y fetal/neonatal en todo el mundo. Los ácidos nucleicos libres de células circulantes en plasma y suero son nuevos biomarcadores con aplicaciones clínicas prometedoras en diferentes campos médicos, incluyendo el diagnóstico prenatal. Se han informado cambios cuantitativos del ADN (cff) fetal libre de células en el plasma materno como indicador de la preeclampsia inminente en diferentes estudios, por ejemplo, utilizando la PCR cuantitativa en tiempo real para los loci SRY o DYS 14 específicos de varones. En los casos de preeclampsia de inicio temprano, se pueden observar niveles elevados en el primer trimestre. El aumento de los niveles de ADNcff antes del inicio de los síntomas puede deberse a la hipoxia/reoxigenación en el espacio intervelloso que da lugar al estrés oxidativo tisular y al aumento de la apoptosis y la necrosis placentaria. Además de la evidencia de un aumento de evacuación de ADNcff en la circulación materna, también hay evidencia de una reducción del aclaramiento renal de ADNcff en la preeclampsia. Como la cantidad de ADN fetal se determina actualmente mediante la cuantificación de secuencias específicas del cromosoma Y, las aproximaciones alternativas tales como la medición del ADN libre de células total o el uso de marcadores epigenéticos fetales independientes del sexo, como la metilación del ADN, ofrecen una alternativa. El ARN libre de células de origen placentario es otro biomarcador alternativo que se puede usar para detectar y diagnosticar la preeclampsia en la práctica clínica. El ARN fetal está asociado con partículas placentarias subcelulares que lo protegen de la degradación. Los niveles de ARN fetal a veces son diez veces más altos en las gestantes con preeclampsia en comparación con los controles, y por lo tanto es un biomarcador alternativo que se puede usar para detectar y diagnosticar la preeclampsia en la práctica clínica.

Patógenos

En algunos casos, la presencia o ausencia de una afección patógena se determina mediante un método o una máquina descrito en el presente documento. Una afección patógena puede estar causada por la infección de un hospedador por un patógeno que incluye, pero sin limitación, una bacteria, virus y hongos. Debido a que los patógenos generalmente poseen ácido nucleico (por ejemplo, ADN genómico, ARN genómico, ARNm) que puede distinguirse del ácido nucleico del hospedador, los métodos y las máquinas que se proporcionan en el presente documento pueden usarse para determinar la presencia o ausencia de un patógeno. Con frecuencia, los patógenos poseen ácido nucleico con características únicas para un patógeno particular tal como, por ejemplo, el estado epigenético y/o una o más variaciones de secuencia, duplicaciones y/o deleciones. Por lo tanto, los métodos proporcionados en el presente documento se pueden usar para identificar un patógeno particular o una variante de patógeno (por ejemplo, una cepa).

Cánceres

En algunos casos, la presencia o ausencia de un trastorno de proliferación celular (por ejemplo, cáncer) se determina

utilizando un método o una máquina descrito en el presente documento. Por ejemplo, los niveles de ácido nucleico libre de células en suero pueden ser elevados en pacientes con varios tipos de cáncer en comparación con pacientes sanos. Los pacientes con enfermedades metastásicas, por ejemplo, a veces pueden tener niveles de ADN en suero aproximadamente dos veces más altos que los pacientes no metastásicos. Los pacientes con enfermedades metastásicas también pueden identificarse mediante marcadores específicos del cáncer y/o ciertos polimorfismos de un solo nucleótido o repeticiones cortas en tándem, por ejemplo. Ejemplos no limitantes de tipos de cáncer que pueden estar correlacionados positivamente con niveles elevados de ADN circulante incluyen cáncer de mama, cáncer colorrectal, cáncer gastrointestinal, cáncer hepatocelular, cáncer de pulmón, melanoma, linfoma no-Hodgkin, leucemia, mieloma múltiple, cáncer de vejiga, hepatoma, cáncer de cuello uterino, cáncer de esófago, cáncer pancreático y cáncer de próstata. Diversos cánceres pueden poseer, y algunas veces liberar, en el torrente sanguíneo, ácidos nucleicos con características que se distinguen de los ácidos nucleicos de las células sanas no cancerosas, tal como, por ejemplo, el estado epigenético y/o variaciones de secuencia, duplicaciones y/o deleciones. Tales características pueden, por ejemplo, ser específicas de un tipo particular de cáncer. Por lo tanto, se contempla además que un método proporcionado en el presente documento puede usarse para identificar un tipo particular de cáncer.

Se puede utilizar el software para realizar una o más etapas en los procesos descritos en el presente documento, incluyendo, pero sin limitación: recuento, procesamiento de datos, generar un resultado y/o proporcionar una o más recomendaciones basadas en los resultados generados, como se describe con mayor detalle a continuación.

20 *Máquinas, Software e Interfaces*

Con frecuencia, ciertos procesos y métodos descritos en el presente documento (p. ej., cuantificación, mapeo, normalización, configuración de intervalos, ajuste, clasificación, recuento y/o determinación de lecturas de secuencia, recuentos, niveles (p. ej., niveles) y/o perfiles) no se pueden realizar sin un ordenador, procesador, software, módulo u otra máquina. Los métodos descritos en el presente documento, son, generalmente, métodos implementados por ordenador, y una o más partes de un método a veces se realizan mediante uno o más procesadores (por ejemplo, microprocesadores), ordenadores o máquinas controladas por microprocesadores. Los casos relacionados con los métodos descritos en este documento generalmente son aplicables a los mismos procesos o procesos relacionados implementados por instrucciones en sistemas, máquinas y productos de programas informáticos descritos en el presente documento. En algunos casos, los procesos y métodos descritos en el presente documento (p. ej., cuantificación, recuento y/o determinación de lecturas de secuencia, recuentos, niveles y/o perfiles) se llevan a cabo mediante métodos automatizados. En algunos casos, se llevan a cabo una o más etapas y un método descrito en el presente documento mediante un procesador y/u ordenador, y/o se llevan a cabo junto con la memoria. En algunos casos, un método automatizado se incorpora en el software, módulos, procesadores, periféricos y/o una máquina que comprende el mismo, que determinan las lecturas de secuencia, recuentos, mapeo, marcadores de secuencias mapeadas, niveles, perfiles, normalizaciones, comparaciones, configuración de intervalos, clasificación, ajustes, representación gráfica, resultados, transformaciones e identificaciones. Como se usa en el presente documento, software se refiere a instrucciones de programa legibles por ordenador que, cuando son ejecutadas por un procesador, realizan operaciones de ordenador, como se describe en el presente documento.

Las lecturas de secuencias, recuentos, niveles y perfiles derivados de un sujeto de ensayo (por ejemplo, un paciente, una gestante) y/o de un sujeto de referencia pueden analizarse y procesarse adicionalmente para determinar la presencia o ausencia de una variación genética. Las lecturas de secuencias, recuentos, los niveles y/o los perfiles a veces se denominan "datos" o "conjuntos de datos". En algunos casos, los datos o conjuntos de datos se pueden caracterizar por una o más características o variables (p. ej., basados en la secuencia [p. ej., contenido de GC, secuencia de nucleótidos específica, similares], función específica [por ejemplo, genes expresados, genes de cáncer, similares], basados en la ubicación [específicos del genoma, específicos del cromosoma, porción o específicos de porción], similares y combinaciones de las mismas). En ciertos casos, los datos o conjuntos de datos se pueden organizar en una matriz que tiene dos o más dimensiones basándose en una o más características o variables. Los datos organizados en matrices se pueden organizar utilizando cualquier característica o variable adecuada. Un ejemplo no limitante de datos en una matriz incluye datos que están organizados por edad materna, ploidía materna y contribución fetal. En ciertos casos, los conjuntos de datos caracterizados por una o más características o variables a veces se procesan después del recuento.

Pueden utilizarse máquinas, software e interfaces para realizar los métodos descritos en el presente documento. Usando máquinas, software e interfaces, un usuario puede introducir, solicitar, consultar o determinar opciones para usar información, programas o procesos particulares (por ejemplo, mapear lecturas de secuencias, procesar datos mapeados y/o proporcionar un resultado), lo que puede implicar la implementación de algoritmos de análisis estadísticos, algoritmos de significación estadística, algoritmos estadísticos, pasos iterativos, algoritmos de validación y representaciones gráficas, por ejemplo. En algunos casos, un usuario puede introducir un conjunto de datos como información de entrada, un usuario puede descargar uno o más conjuntos de datos mediante medios de hardware adecuados (por ejemplo, una unidad flash) y/o un usuario puede enviar un conjunto de datos de un sistema a otro para procesar posteriormente y/o proporcionar un resultado (p. ej., enviar datos de lecturas de secuencias desde un secuenciador a un sistema informático para el mapeo de lecturas de secuencia; enviar datos de secuencia mapeados en un sistema informático para procesar y generar un resultado y/o informe).

Un sistema comprende, generalmente, una o más máquinas. Cada máquina comprende una o más memorias, uno o más procesadores e instrucciones. Cuando un sistema incluye dos o más máquinas, algunas o todas las máquinas pueden estar ubicadas en la misma ubicación, algunas o todas las máquinas pueden estar ubicadas en diferentes ubicaciones, todas las máquinas pueden estar ubicadas en una ubicación y/o todas las máquinas pueden estar
 5 ubicadas en diferentes ubicaciones. Cuando un sistema incluye dos o más máquinas, algunas o todas las máquinas pueden estar ubicadas en la misma ubicación que un usuario, algunas o todas las máquinas pueden estar ubicadas en una ubicación diferente a la de un usuario, todas las máquinas pueden estar ubicadas en la misma ubicación que el usuario, y/o todas las máquinas pueden estar ubicadas en una o más ubicaciones diferentes a las del usuario.

10 Un sistema comprende a veces un ordenador y una máquina de secuenciación, donde la máquina de secuenciación está configurada para recibir ácido nucleico físico y generar lecturas de secuencias, y el ordenador está configurado para procesar las lecturas desde la máquina de secuenciación. El ordenador a veces se configura para determinar la presencia o ausencia de una variación genética (por ejemplo, la variación en el número de copias; aneuploidía cromosómica fetal) a partir de las lecturas de secuencia.

15 Un usuario puede, por ejemplo, realizar una consulta a un software que puede adquirir, a continuación, un conjunto de datos a través de acceso a Internet, y en ciertos casos, se le puede pedir a un procesador programable que adquiera un conjunto de datos adecuado basándose en los parámetros dados. Un procesador programable también puede pedirle a un usuario que seleccione una o más opciones de conjuntos de datos seleccionadas por el procesador basándose en los parámetros dados. Un procesador programable puede pedir a un usuario que seleccione una o más
 20 opciones de conjuntos de datos seleccionadas por el procesador basándose en la información encontrada a través de Internet, otra información interna o externa, o similares. Se pueden elegir opciones para seleccionar una o más selecciones de características de datos, uno o más algoritmos estadísticos, uno o más algoritmos de análisis estadístico, uno o más algoritmos de significación estadística, pasos iterativos, uno o más algoritmos de validación y
 25 una o más representaciones gráficas de métodos, máquinas o programas informáticos.

Los sistemas abordados en el presente documento pueden comprender componentes generales de sistemas informáticos, tal como, por ejemplo, servidores de red, sistemas de ordenadores portátiles, sistemas de ordenadores de escritorio, sistemas portátiles, asistentes digitales personales, quioscos informáticos y similares. Un sistema
 30 informático puede comprender uno o más medios de entrada, como un teclado, pantalla táctil, ratón, reconocimiento por voz u otros medios para permitir que el usuario introduzca datos en el sistema. Un sistema puede comprender adicionalmente una o más salidas, incluyendo, pero sin limitación, una pantalla de visualización (p. ej., CRT o LCD), altavoz, máquina de fax, impresora (p. ej., impresora láser, de chorro de tinta, de impacto, en blanco y negro o en color), u otra salida útil para proporcionar una salida visual, auditiva y/o impresa de información (por ejemplo, resultado
 35 y/o informe).

En un sistema, los medios de entrada y salida se pueden conectar a una unidad central de procesamiento que puede comprender, entre otros componentes, un microprocesador para ejecutar las instrucciones del programa y la memoria para almacenar el código y los datos del programa. En algunos casos, los procesos pueden implementarse como un
 40 sistema de usuario único ubicado en un sitio geográfico único. En ciertos casos, los procesos pueden implementarse como un sistema multiusuario. En el caso de una implementación multiusuario, se pueden conectar múltiples unidades centrales de procesamiento por medio de una red. La red puede ser local, abarcando un solo departamento en una parte de un edificio, un edificio completo, abarcar múltiples edificios, abarcar una región, abarcar un país entero o ser mundial. La red puede ser privada, ser propiedad y estar controlado por un proveedor, o puede implementarse como
 45 un servicio basado en Internet donde el usuario accede a una página web para introducir y recuperar información. Por consiguiente, en ciertos casos, un sistema incluye una o más máquinas, que pueden ser locales o remotas con respecto a un usuario. Un usuario puede acceder a más de una máquina en una o varias ubicaciones, y los datos se pueden mapear y/o procesar en serie y/o en paralelo. Por lo tanto, se puede utilizar una configuración y control adecuados para mapear y/o procesar datos utilizando múltiples máquinas, como en redes locales, redes remotas y/o
 50 plataformas informáticas en la "nube".

En algunos casos, un sistema puede incluir una interfaz de comunicaciones. Una interfaz de comunicaciones permite la transferencia de software y datos entre un sistema informático y uno o más dispositivos externos. Ejemplos no
 55 limitantes de interfaces de comunicaciones incluyen un módem, una interfaz de red (como una tarjeta Ethernet), un puerto de comunicaciones, una ranura y tarjeta PCMCIA y similares. El software y los datos transferidos a través de una interfaz de comunicaciones generalmente tienen la forma de señales, que pueden ser electrónicas, electromagnéticas, ópticas y/u otras señales capaces de ser recibidas por una interfaz de comunicaciones. Las señales a menudo se proporcionan a una interfaz de comunicaciones a través de un canal. Un canal a menudo transporta señales y puede implementarse utilizando red o cable, fibra óptica, una línea telefónica, un enlace de teléfono móvil, un enlace de RF y/u otros canales de comunicaciones. Por lo tanto, en un ejemplo, se puede utilizar una interfaz de
 60 comunicaciones para recibir información de señal que puede detectarse mediante un módulo de detección de señal.

Los datos pueden introducirse por un dispositivo y/o método adecuado, incluyendo, pero sin limitación, dispositivos de entrada manual o dispositivos de entrada directa de datos (DDE). Ejemplos no limitantes de dispositivos manuales
 65 incluyen teclados, teclados de conceptos, pantallas sensibles al tacto, rotuladores, ratón, bolas de seguimiento, joysticks, tabletas gráficas, escáneres, cámaras digitales, digitalizadores de vídeo y dispositivos de reconocimiento por

voz. Ejemplos no limitantes de DDE incluyen lectores de códigos de barras, códigos de banda magnética, tarjetas inteligentes, reconocimiento de caracteres con tinta magnética, reconocimiento óptico de caracteres, reconocimiento óptico de marcas y documentos de respuesta.

- 5 En algunos casos, la salida de una máquina de secuenciación puede servir como datos que pueden introducirse a través de un dispositivo de entrada. En ciertos casos, las lecturas de secuencia mapeadas pueden servir como datos que pueden introducirse a través de un dispositivo de entrada. En ciertos casos, el tamaño del fragmento de ácido nucleico (por ejemplo, la longitud) puede servir como datos que pueden introducirse a través de un dispositivo de entrada. En ciertos casos, la salida de un proceso de captura de ácidos nucleicos (por ejemplo, datos de origen de la región genómica) puede servir como datos que se pueden introducir a través de un dispositivo de entrada. En ciertos casos, una combinación de tamaño de fragmentos de ácidos nucleicos (por ejemplo, longitud) y salida de un proceso de captura de ácidos nucleicos (por ejemplo, datos de origen de región genómica) puede servir como datos que pueden introducirse a través de un dispositivo de entrada. En ciertos casos, los datos simulados se generan mediante un proceso *in silico* y los datos simulados sirven como datos que pueden introducirse a través de un dispositivo de entrada. La expresión "*in silico*" se refiere a investigaciones y experimentos realizados con un ordenador. Los procesos *in silico* incluyen, pero sin limitación, mapear lecturas de secuencias y procesar las lecturas de secuencias mapeadas de acuerdo con los procesos descritos en el presente documento.

20 Un sistema puede incluir software útil para realizar un proceso descrito en el presente documento, y el software puede incluir uno o más módulos para realizar tales procesos (por ejemplo, módulo de secuenciación, módulo de procesamiento lógico, módulo de organización de visualización de datos). El término "software" se refiere a instrucciones de programa legibles por ordenador que, cuando son ejecutadas por un ordenador, realizan operaciones de ordenador. Las instrucciones ejecutables por uno o más procesadores a veces se proporcionan como código ejecutable que, cuando se ejecuta, puede hacer que uno o más procesadores implementen un método descrito en el presente documento. Un módulo descrito en el presente documento puede existir como software, y las instrucciones (por ejemplo, procesos, rutinas, subrutinas) incorporadas en el software se pueden implementar o realizar por un procesador. Por ejemplo, un módulo (por ejemplo, un módulo de software) puede ser parte de un programa que realiza un proceso o tarea particular. El término "módulo" se refiere a una unidad funcional autónoma que se puede usar en una máquina más grande o en un sistema de software. Un módulo puede comprender un conjunto de instrucciones para llevar a cabo una función del módulo. Un módulo puede transformar datos y/o información. Los datos y/o información pueden estar en una forma adecuada. Por ejemplo, los datos y/o información pueden ser digitales o analógicos. En ciertos casos, los datos y/o información a veces pueden ser paquetes, bytes, caracteres o bits. Los datos y/o información pueden ser datos o información recopilados, ensamblados o utilizables. Ejemplos no limitantes de datos y/o información incluyen medios adecuados, imágenes, vídeo, sonido (por ejemplo, frecuencias, audibles o no audibles), números, constantes, un valor, objetos, tiempo, funciones, instrucciones, mapas, referencias, secuencias, lecturas, lecturas mapeadas, niveles, intervalos, umbrales, señales, pantallas, representaciones o transformaciones de los mismos. Un módulo puede aceptar o recibir datos y/o información, transformar los datos y/o información en una segunda forma, y proporcionar o transferir la segunda forma a una máquina, periférico, componente u otro módulo. Un módulo puede realizar una o más de las siguientes funciones no limitantes: mapear lecturas de secuencias, proporcionar recuentos, ensamblar porciones, proporcionar o determinar un nivel, proporcionar un perfil de recuento, normalizar (por ejemplo, normalizar lecturas, normalizar recuentos y similares), proporcionar un perfil de recuentos normalizados o niveles de recuentos normalizados, comparar dos o más niveles, proporcionar valores de incertidumbre, proporcionar o determinar niveles esperados e intervalos esperados (por ejemplo, intervalos de nivel esperados, intervalos de umbral y niveles umbral), proporcionar ajustes a los niveles (por ejemplo, ajustar un primer nivel, ajustar un segundo nivel, ajustar un perfil de un cromosoma o un segmento del mismo, *y/o padding*), proporcionar identificación (por ejemplo, identificar una variación en el número de copias, variación genética o aneuploidía), clasificar, representar gráficamente y/o determinar un resultado, por ejemplo. En ciertos casos, un procesador puede llevar a cabo las instrucciones en un módulo. En ciertos casos, se pueden requerir uno o más procesadores para llevar a cabo las instrucciones en un módulo o grupo de módulos. Un módulo puede proporcionar datos y/o información a otro módulo, máquina o fuente y puede recibir datos y/o información de otro módulo, máquina o fuente.

55 A veces, se incorpora un programa informático en un medio tangible legible por ordenador, y, a veces, se materializa de manera tangible en un medio legible por ordenador no transitorio. A veces, se almacena un módulo en un medio legible por ordenador (por ejemplo, disco, unidad) o en memoria (por ejemplo, memoria de acceso aleatorio). Un módulo y un procesador capaces de implementar instrucciones desde un módulo pueden ubicarse en una máquina o en una máquina diferente. Un módulo y/o procesador capaz de implementar una instrucción para un módulo puede ubicarse en la misma ubicación que un usuario (por ejemplo, red local) o en una ubicación diferente a la de un usuario (por ejemplo, red remota, sistema en la nube). En los casos en que se lleva a cabo un método junto con dos o más módulos, los módulos pueden ubicarse en la misma máquina, uno o más módulos pueden ubicarse en una máquina diferente en la misma ubicación física, y uno o más módulos pueden ubicarse en diferentes máquinas en diferentes ubicaciones físicas.

65 En algunos casos, una máquina, comprende al menos un procesador para llevar a cabo las instrucciones en un módulo. A veces, se accede a los recuentos de lecturas de secuencias mapeadas en porciones de un genoma de referencia mediante un procesador que ejecuta instrucciones configuradas para llevar a cabo un método descrito en el presente documento. Los recuentos a los que se accede mediante un procesador pueden estar dentro de la memoria

de un sistema, y se puede acceder a los recuentos y colocarlos en la memoria del sistema una vez que se obtienen. En algunos casos, una máquina incluye un procesador (por ejemplo, uno o más procesadores) cuyo procesador puede realizar y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) desde un módulo. En algunos casos, una máquina incluye múltiples procesadores, como procesadores coordinados y que trabajan en paralelo. En algunos casos, una máquina opera con uno o más procesadores externos (por ejemplo, una red interna o externa, servidor, dispositivo de almacenamiento y/o red de almacenamiento (por ejemplo, una nube)). En algunos casos, una máquina comprende un módulo. En ciertos casos, una máquina comprende uno o más módulos. Una máquina que comprende un módulo, con frecuencia, puede recibir y transferir uno o más datos y/o información hacia y desde otros módulos. En ciertos casos, una máquina comprende periféricos y/o componentes. En ciertos casos, una máquina comprende uno o más periféricos o componentes que pueden transferir datos y/o información hacia y desde otros módulos, periféricos y/o componentes. En ciertos casos, una máquina interactúa con un periférico y/o componente que proporciona datos y/o información. En ciertos casos, los periféricos y los componentes ayudan a una máquina a realizar una función o a interactuar directamente con un módulo. Ejemplos no limitantes de periféricos y componentes incluyen un periférico informático adecuado, I/O o método o dispositivo de almacenamiento que incluye, pero sin limitación, escáneres, impresoras, pantallas (por ejemplo, monitores, LED, LCT o CRT), cámaras, micrófonos, pads (por ejemplo, iPads, tabletas), pantallas táctiles, teléfonos inteligentes, teléfonos móviles, Dispositivos USB de E/S, dispositivos de almacenamiento masivo USB, teclados, un ratón de ordenador, bolígrafos digitales, módems, discos duros, unidades de salto, unidades flash, un procesador, un servidor, CD, DVD, tarjetas gráficas, Dispositivos de E/S especializados (por ejemplo, secuenciadores, células fotoeléctricas, tubos multiplicadores de fotos, lectores ópticos, sensores, etc.), una o más celdas de flujo, componentes de manejo de fluidos, controladores de interfaz de red, ROM, RAM, métodos y dispositivos de transferencia inalámbrica (Bluetooth, WiFi y similares), la red mundial (www), internet, un ordenador y/u otro módulo.

Con frecuencia, el software se proporciona en un programa que contiene instrucciones del programa grabadas en un medio legible por ordenador, incluyendo, pero sin limitación, medios magnéticos que incluyen disquetes, discos duros y cintas magnéticas; y medios ópticos que incluyen discos CD-ROM, discos DVD, discos magnetoópticos, unidades flash, RAM, disco flexible, similares y otros medios similares en los que se pueden grabar las instrucciones del programa. En la implementación en línea, pueden configurarse un servidor y un sitio web mantenido por una organización para proporcionar descargas de software a usuarios remotos, o los usuarios remotos pueden acceder a un sistema remoto mantenido por una organización para acceder de manera remota al software. El software puede obtener o recibir información de entrada. El software puede incluir un módulo que específicamente obtiene o recibe datos (por ejemplo, un módulo de recepción de datos que recibe datos de lecturas de secuencias y/o datos de lecturas mapeadas) y puede incluir un módulo que procesa específicamente los datos (por ejemplo, un módulo de procesamiento que procesa datos recibidos (por ejemplo, filtra, normaliza, proporciona un resultado y/o informe)). Los términos "obtener" y "recibir" información de entrada se refieren a recibir datos (por ejemplo, lecturas de secuencias, lecturas mapeadas) por medios de comunicación por ordenador de un sitio local o remoto, entrada de datos por seres humanos o cualquier otro método para recibir datos. La información de entrada puede generarse en la misma ubicación en donde se recibe, o puede generarse en una ubicación diferente y transmitirse a la ubicación de recepción. En algunos casos, la información de entrada se modifica antes de que se procese (por ejemplo, se coloca en un formato susceptible de procesamiento (por ejemplo, tabulado)).

En algunos casos, se proporcionan productos de programas informáticos, tales como, por ejemplo, un medio de almacenamiento legible por ordenador no transitorio con un programa ejecutable almacenado en el mismo, en el que el programa da instrucciones a un microprocesador para realizar lo siguiente: (a) obtener lecturas de secuencias de ácidos nucleicos de muestra de un sujeto de ensayo; (b) mapear las lecturas de secuencias obtenidas en (a) a un genoma conocido, cuyo genoma conocido se ha dividido en porciones; (c) hacer el recuento de las lecturas de secuencias mapeadas dentro de las porciones; (d) generar un perfil de recuentos normalizados de la muestra normalizando los recuentos para las porciones obtenidas en (c); y (e) determinar la presencia o ausencia de una variación genética del perfil de recuentos normalizados de la muestra en (d).

En algunos casos, en el presente documento se proporciona un sistema que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende lecturas de secuencia de nucleótidos mapeadas en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas en las que una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que las de una longitud de un fragmento seleccionado se mapea, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) determinar la cantidad de lecturas de secuencia mapeadas en el subconjunto de secciones genómicas y (b) determinar la presencia o ausencia de una variación genética de acuerdo con la cantidad determinada en (a). En algunos casos, un sistema comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende lecturas de secuencias de nucleótidos de ácido nucleico circulante libre de células mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de una gestante y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para seleccionar un subconjunto de secciones genómicas que consisten esencialmente en secciones genómicas a las que una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que la longitud de un

fragmento seleccionado se mapean. En algunos casos, un sistema comprende uno o más procesadores y memoria, cuya memoria comprende una lista a la que pueden acceder uno o más procesadores, cuya lista comprende un subconjunto de secciones genómicas que consisten esencialmente en secciones genómicas a las cuales una cantidad significativa de las lecturas de secuencia de fragmentos de ácido nucleico que tienen longitudes menores que la longitud de un fragmento seleccionado se mapean, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de una gestante que tiene un feto.

En algunos casos, en el presente documento se presenta un sistema que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácido nucleico que tienen longitudes que son menores que la longitud de un fragmento seleccionado y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencia mapeadas en las secciones genómicas y (b) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

En algunos casos, en el presente documento se presenta una máquina que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende lecturas de secuencia de nucleótidos mapeadas en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas en las que una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que una longitud de un fragmento seleccionado se mapea, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) determinar la cantidad de lecturas de secuencia mapeadas en el subconjunto de secciones genómicas y (b) determinar la presencia o ausencia de una variación genética de acuerdo con la cantidad determinada en (a). En algunos casos, una máquina comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende lecturas de secuencias de nucleótidos de ácido nucleico circulante libre de células mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de una gestante y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para seleccionar un subconjunto de secciones genómicas que consisten esencialmente en secciones genómicas a las que se mapea una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que una longitud de fragmento seleccionada. En ciertos casos, una máquina comprende uno o más procesadores y memoria, cuya memoria comprende una lista a la que se puede acceder mediante uno o más procesadores, cuya lista comprende un subconjunto de secciones genómicas que consisten esencialmente en secciones genómicas en las que una cantidad significativa de lecturas de secuencias de fragmentos de ácido nucleico que tienen longitudes menores que una longitud del fragmento seleccionado se mapea, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de una gestante que tiene un feto.

En algunos casos, en el presente documento se presenta una máquina que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende recuentos de lecturas de secuencia de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante y (ii) lecturas de fragmentos de ácido nucleico que tienen longitudes que son menores que una longitud de un fragmento seleccionado, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencia mapeadas en las secciones genómicas y (b) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

En algunos casos, en el presente documento se presenta un medio de almacenamiento no transitorio legible por ordenador con un programa ejecutable almacenado en él, en el que el programa da instrucciones a un microprocesador para que realice lo siguiente: (a) obtener lecturas de secuencias de nucleótidos mapeadas en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas a las que una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que una longitud de fragmento seleccionada se mapea (b) determinar la cantidad de lecturas de secuencias mapeadas en el subconjunto de secciones genómicas y (c) determinar la presencia o ausencia de una variación genética de acuerdo con la cantidad determinada en (b). En algunos casos, en el presente documento se presenta un medio de almacenamiento no transitorio legible por ordenador con un programa ejecutable almacenado en él, en el que el programa da instrucciones a un microprocesador para que realice lo siguiente: (a) obtener lecturas de secuencias de nucleótidos mapeadas en un subconjunto de secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, en donde el subconjunto consiste esencialmente en secciones genómicas a las que una cantidad significativa de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que una longitud de un fragmento seleccionado se mapean y (b) seleccionar un subconjunto de secciones genómicas que consisten

esencialmente en secciones genómicas a las que se mapea la cantidad de lecturas de fragmentos de ácido nucleico que tienen longitudes menores que la longitud de un fragmento seleccionado. En ciertos casos, en el presente documento se presenta un medio de almacenamiento no transitorio legible por ordenador con un programa ejecutable almacenado en él, que comprende una lista a la que pueden acceder uno o más procesadores, cuya lista comprende

5 un subconjunto de secciones genómicas que consisten esencialmente en secciones genómicas en las que una cantidad significativa de lecturas de secuencias de fragmentos de ácido nucleico que tienen longitudes menores que una longitud de un fragmento seleccionado se mapean, cuyas lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra de una gestante que tiene un feto

10 En el presente documento también se presenta un medio de almacenamiento no transitorio legible por ordenador con un programa ejecutable almacenado en el mismo, en el que el programa da instrucciones a un microprocesador para que realice lo siguiente: (a) obtener recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácido nucleico que tienen

15 longitudes que son menores que una longitud de un fragmento seleccionado (b) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas y (c) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados

En ciertos casos, el software puede incluir uno o más algoritmos. Se puede usar un algoritmo para procesar datos y/o proporcionar un resultado o informe de acuerdo con una secuencia finita de instrucciones. Con frecuencia, un algoritmo es una lista de instrucciones definidas para completar una tarea. A partir de un estado inicial, las instrucciones pueden describir un cálculo que se realiza a través de una serie definida de estados sucesivos, terminando finalmente en un estado de finalización final. La transición de un estado al siguiente no es necesariamente determinista (por ejemplo, algunos algoritmos incorporan aleatoriedad). A modo de ejemplo y sin limitación, un algoritmo puede ser un algoritmo

20 de búsqueda, algoritmo de clasificación, algoritmo de combinación, algoritmo numérico, algoritmo gráfico, algoritmo de cadena, algoritmo de modelado, algoritmo genómico computacional, algoritmo combinatorio, algoritmo de aprendizaje automático, algoritmo de criptografía, algoritmo de compresión de datos, algoritmo de análisis y similares. Un algoritmo puede incluir un algoritmo o dos o más algoritmos que trabajan en combinación. Un algoritmo puede ser de cualquier clase de complejidad y/o complejidad parametrizada adecuadas. Para el cálculo y/o el procesamiento de datos se puede usar un algoritmo y, en algunos casos, se puede usar en una aproximación determinista o probabilística/predictiva. Se puede implementar un algoritmo en un entorno informático usando un lenguaje de programación adecuado, cuyos ejemplos no limitantes son C, C++, Java, Perl, Python, Fortran y similares. En algunos casos, un algoritmo puede configurarse o modificarse para incluir un margen de errores, análisis estadístico, significación estadística y/o comparación con otra información o conjuntos de datos (por ejemplo, aplicable cuando se

25 utiliza una red neuronal o un algoritmo de agrupamiento).

En ciertos casos, se pueden implementar varios algoritmos para su uso en el software. En algunos casos, estos algoritmos se pueden entrenar con datos sin procesar. Para cada nueva muestra de datos sin procesar, los algoritmos capacitados pueden producir un conjunto de datos procesados o resultados representativos. Un conjunto de datos procesados a veces es de complejidad reducida en comparación con el conjunto de datos matriz que se procesó. En algunos casos, basándose en un conjunto procesado, el rendimiento de un algoritmo capacitado puede evaluarse según la sensibilidad y la especificidad. Se puede identificar y utilizar un algoritmo con la mayor sensibilidad y/o

30 especificidad.

En ciertos casos, los datos simulados (o simulación) pueden ayudar al procesamiento de datos, por ejemplo, capacitando un algoritmo o probando un algoritmo. En algunos casos, los datos simulados incluyen varias muestras hipotéticas de diferentes agrupaciones de lecturas de secuencias. Los datos simulados pueden basarse en lo que se podría esperar de una población real o se pueden sesgar para probar un algoritmo y/o asignar una clasificación correcta. Los datos simulados también se denominan en el presente documento, datos "virtuales". En ciertos casos,

35 las simulaciones pueden realizarse con un programa informático. Una etapa posible en el uso de un conjunto de datos simulados es evaluar la confianza de resultados identificados, por ejemplo, cómo de bien coincide o representa mejor un muestreo aleatorio los datos originales. Una aproximación consiste en calcular un valor de probabilidad (valor de p), que estima la probabilidad de que una muestra aleatoria tenga una mejor puntuación que las muestras seleccionadas. En algunos casos, se puede evaluar un modelo empírico, en donde se supone que al menos una muestra coincide con una muestra de referencia (con o sin variaciones resueltas). En algunos casos, otra distribución, tal como una distribución de Poisson, por ejemplo, puede usarse para definir la distribución de probabilidad.

En ciertos casos, un sistema puede incluir uno o más procesadores. Un procesador se puede conectar a un bus de comunicación. Un sistema informático puede incluir una memoria principal, con frecuencia, una memoria de acceso aleatorio (RAM), y también puede incluir una memoria secundaria. En algunos casos, la memoria comprende un medio de almacenamiento legible por ordenador no transitorio. La memoria secundaria puede incluir, por ejemplo, una unidad de disco duro y/o una unidad de almacenamiento extraíble, que representa una unidad de disquete, una unidad de cinta magnética, una unidad de disco óptico, una tarjeta de memoria y similares. Con frecuencia, una unidad de almacenamiento extraíble lee de y/o escribe en una unidad de almacenamiento extraíble. Ejemplos no limitantes de

40 unidades de almacenamiento extraíbles incluyen un disquete, cinta magnética, disco óptico y similares, que pueden leerse y escribirse mediante, por ejemplo, una unidad de almacenamiento extraíble. Una unidad de almacenamiento

extraíble puede incluir un medio de almacenamiento utilizable por ordenador que tiene almacenado en el mismo, software y/o datos informáticos.

5 Un procesador puede implementar software en un sistema. En algunos casos, un procesador puede programarse para realizar automáticamente una tarea descrita en el presente documento que un usuario podría realizar. Por consiguiente, un procesador, o algoritmo conducido por tal procesador, puede requerir poca o ninguna supervisión o entrada por parte de un usuario (por ejemplo, el software puede programarse para implementar una función automáticamente). En algunos casos, la complejidad de un proceso es tan grande que una sola persona o grupo de personas no podría realizar el proceso en un período de tiempo suficientemente corto para determinar la presencia o ausencia de una variación genética.

15 En algunos casos, la memoria secundaria puede incluir otros medios similares para permitir que los programas informáticos u otras instrucciones se carguen en un sistema informático. Por ejemplo, un sistema puede incluir una unidad de almacenamiento extraíble y un dispositivo de interfaz. Ejemplos no limitantes de dichos sistemas incluyen un cartucho de programa y una interfaz de cartucho (como la que se encuentra en los dispositivos de videojuegos), un chip de memoria extraíble (como una EPROM o PROM) y un conector asociado, y otras unidades de almacenamiento extraíbles e interfaces que permite que el software y los datos se transfieran desde la unidad de almacenamiento extraíble a un sistema informático.

20 Una entidad puede generar recuentos de lecturas de secuencia, mapear las lecturas de secuencia a porciones, contar las lecturas mapeadas y, en algunos casos, utilizar las lecturas mapeadas contabilizadas en un método, sistema, máquina o programa informático descrito en el presente documento. En ciertos casos, los recuentos de lecturas de secuencia mapeadas en porciones a veces se transfieren por una entidad a una segunda entidad para utilizarse por la segunda entidad en un método, sistema, máquina o programa informático descrito en el presente documento.

25 En algunos casos, una entidad genera lecturas de secuencia y, en algunos casos, una segunda entidad mapea esas lecturas de secuencia a porciones en un genoma de referencia. La segunda entidad a veces cuenta las lecturas mapeadas y utiliza las lecturas mapeadas contadas en un método, sistema, máquina o programa informático descrito en el presente documento. En ciertos casos, la segunda entidad transfiere las lecturas mapeadas en una tercera entidad, y la tercera entidad cuenta las lecturas mapeadas y utiliza las lecturas mapeadas en un método, sistema, máquina o programa informático descrito en el presente documento. En ciertos casos, la segunda entidad cuenta las lecturas mapeadas y transfiere las lecturas mapeadas contadas a una tercera entidad, y la tercera entidad utiliza las lecturas mapeadas contadas en un método, sistema, máquina o programa informático descrito en el presente documento. En los casos que implican una tercera entidad, la tercera entidad a veces es la misma que la primera entidad. Es decir, la primera entidad a veces transfiere las lecturas de secuencia a una segunda entidad, cuya segunda entidad puede mapear lecturas de secuencia a porciones en un genoma de referencia y/o contar las lecturas mapeadas, y la segunda entidad puede transferir las lecturas mapeadas y/o contadas a una tercera entidad. Una tercera entidad a veces puede utilizar las lecturas mapeadas y/o contadas en un método, sistema, máquina o programa informático descrito en el presente documento, en donde la tercera entidad a veces es la misma que la primera entidad, y algunas veces la tercera entidad es diferente de la primera o segunda entidad.

45 En algunos casos, una entidad obtiene sangre de una gestante, opcionalmente aísla el ácido nucleico de la sangre (por ejemplo, del plasma o suero), y transfiere la sangre o el ácido nucleico a una segunda entidad que genera lecturas de secuencia del ácido nucleico.

50 La FIG. 24 ilustra un ejemplo no limitante de un entorno informático 510 en donde se implementan varios sistemas, métodos algoritmos y estructuras de datos descritos en el presente documento. El entorno informático 510 es solo un ejemplo de un entorno informático adecuado y no pretende sugerir ninguna limitación en cuanto al alcance del uso o la funcionalidad de los sistemas, métodos y estructuras de datos descritos en el presente documento. El entorno informático 510 tampoco debe interpretarse como que tiene una dependencia o requisito relacionado con ninguno o con una combinación de componentes ilustrados en el entorno informático 510. En ciertos casos, puede utilizarse un subconjunto de sistemas, métodos y estructuras de datos mostrados en la FIG. 24. Los sistemas, métodos y estructuras de datos descritos en el presente documento son operativos con muchos otros entornos o configuraciones de sistemas informáticos de uso general o de uso especial. Ejemplos de sistemas informáticos, entornos y/o configuraciones conocidos que pueden ser adecuados incluyen, pero sin limitación, ordenadores personales, ordenadores servidores, clientes ligeros, clientes pesados, dispositivos de mano o portátiles, sistemas multiprocesador, sistemas basados en microprocesadores, decodificadores, dispositivos electrónicos de consumo programables, PC de red, miniordenadores, ordenadores centrales, entornos informáticos distribuidos que incluyen cualquiera de los sistemas o dispositivos mencionados anteriormente, y similares.

60 El entorno operativo 510 de la FIG. 24 incluye un dispositivo informático de uso general en forma de un ordenador 520, que incluye una unidad de procesamiento 521, una memoria del sistema 522 y un bus del sistema 523 que acopla operativamente varios componentes del sistema, incluida la memoria del sistema 522, a la unidad de procesamiento 521. Puede haber solo una o puede haber más de una unidad de procesamiento 521, de manera que el procesador del ordenador 520 incluya una sola unidad central de procesamiento (CPU), o una pluralidad de unidades de procesamiento, comúnmente conocidas como un entorno de procesamiento paralelo. El ordenador 520 puede ser un

ordenador convencional, un ordenador distribuido o cualquier otro tipo de ordenador.

El bus del sistema 523 puede ser cualquiera de los varios tipos de estructuras de bus que incluyen un bus de memoria o un controlador de memoria, un bus periférico y un bus local que utiliza cualquiera de una variedad de arquitecturas de bus. La memoria del sistema también puede denominarse simplemente memoria, e incluye memoria de solo lectura (ROM) 524 y memoria de acceso aleatorio (RAM). Un sistema básico de entrada/salida (BIOS) 526, que contiene las rutinas básicas que ayudan a transferir información entre elementos dentro del ordenador 520, como durante el inicio, se almacena en la ROM 524. El ordenador 520 puede incluir además una interfaz de disco duro 527 para leer de y escribir en un disco duro, no se muestra, una unidad de disco magnético 528 para leer de o escribir en un disco magnético extraíble 529, y una unidad de disco óptico 530 para leer de o escribir en un disco óptico extraíble 531 como un CD ROM u otros medios ópticos.

La unidad de disco duro 527, la unidad de disco magnético 528 y la unidad de disco óptico 530 están conectadas al bus del sistema 523 mediante una interfaz de unidad de disco duro 532, una interfaz de unidad de disco magnético 533 y una interfaz de unidad de disco óptico 534, respectivamente. Las unidades y sus medios asociados legibles por ordenador proporcionan un almacenamiento no volátil de instrucciones legibles por ordenador, estructuras de datos, módulos de programas y otros datos para el ordenador 520. Se puede utilizar en el entorno operativo, cualquier tipo de medio legible por ordenador que pueda almacenar datos a los que pueda acceder un ordenador, como casetes magnéticos, tarjetas de memoria flash, discos de vídeo digital, cartuchos Bernoulli, memorias de acceso aleatorio (RAM), memorias de solo lectura (ROM) y similares.

Se pueden almacenar varios módulos de programas en el disco duro, disco magnético 529, disco óptico 531, ROM 524 o RAM, incluido un sistema operativo 535, uno o más programas de aplicación 536, otros módulos de programa 537 y datos de programa 538. Un usuario puede introducir comandos e información en el ordenador personal 520 a través de dispositivos de entrada como un teclado 540 y un dispositivo señalador 542. Otros dispositivos de entrada (no mostrados) pueden incluir un micrófono, joystick, control de juegos, antena parabólica, escáner o similares. Con frecuencia, estos y otros dispositivos de entrada se conectan a la unidad de procesamiento 521 a través de una interfaz de puerto en serie 546 que está acoplada al bus del sistema, pero pueden estar conectadas por otras interfaces, como un puerto paralelo, puerto de juegos o un bus universal en serie (USB). Un monitor 547 u otro tipo de dispositivo de visualización también se conecta al bus del sistema 523 a través de una interfaz, como un adaptador de vídeo 548. Además del monitor, los ordenadores suelen incluir otros dispositivos de salida periféricos (no mostrados), como altavoces e impresoras.

El ordenador 520 puede operar en un entorno de red usando conexiones lógicas a uno o más ordenadores remotos, como el ordenador remoto 549. Estas conexiones lógicas se pueden lograr mediante un dispositivo de comunicación acoplado a o a una parte del ordenador 520, o de otra manera. El ordenador remoto 549 puede ser otro ordenador, un servidor, un enrutador, un PC de red, un cliente, un dispositivo de par u otro nodo de red común, y generalmente incluye muchos o todos los elementos descritos anteriormente en relación con el ordenador 520, aunque solo se ha ilustrado un dispositivo de almacenamiento de memoria 550 en la FIG. 24. Las conexiones lógicas representadas en la FIG. 24 incluyen una red de área local (LAN) 551 y una red de área amplia (WAN) 552. Dichos entornos de red son comunes en las redes de oficinas, redes de ordenadores de todas las empresas, intranets e Internet, que son todos tipos de redes.

Cuando se usa en un entorno de red LAN, el ordenador 520 está conectado a la red local 551 a través de una interfaz de red o adaptador 553, que es un tipo de dispositivo de comunicaciones. Cuando se usa en un entorno de red WAN, el ordenador 520 a menudo incluye un módem 554, un tipo de dispositivo de comunicaciones o cualquier otro tipo de dispositivo de comunicaciones para establecer comunicaciones a través de la red de área amplia 552. El módem 554, que puede ser interno o externo, se conecta al bus del sistema 523 a través de la interfaz de puerto en serie 546. En un entorno de red, los módulos de programa representados en relación con el ordenador personal 520, o partes de los mismos, pueden almacenarse en el dispositivo de almacenamiento de memoria remota. Se aprecia que las conexiones de red mostradas son ejemplos no limitantes y se pueden usar otros dispositivos de comunicaciones para establecer un enlace de comunicaciones entre ordenadores.

Módulos

Se pueden utilizar uno o más módulos pueden ser utilizados en un método descrito en el presente documento, cuyos ejemplos no limitantes incluyen un módulo de procesamiento lógico, módulo de organización de visualización de datos, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de filtrado, módulo de ponderación, módulo de normalización, módulo de sesgo de GC, módulo de nivel, módulo de comparación, módulo de configuración de intervalos, módulo de clasificación, módulo de ajuste, módulo de representación gráfica, módulo de representación, módulo de relación, módulo de resultados y/o módulo de organización de visualización de datos, módulo de relación, módulo de resultados y/o módulo de organización de visualización de datos, similares o combinaciones de los mismos. Los módulos a veces están controlados por un microprocesador. En ciertos casos, un módulo o una máquina que comprende uno o más módulos, recopila, ensambla, recibe, obtiene, accede, recupera, proporciona y/o transfiere datos y/o información a, o desde, otro módulo, máquina, componente, periférico u operador de una máquina. En algunos casos, los datos y/o la información (por ejemplo, lecturas de secuenciación) se proporcionan a un módulo

mediante una máquina que comprende uno o más de los siguientes: una o más celdas de flujo, una cámara, un detector (por ejemplo, un fotodetector, una célula fotoeléctrica, un detector eléctrico (por ejemplo, un detector de modulación de amplitud, un detector de modulación de frecuencia y fase, un detector de bucle de bloqueo de fase), un contador, un sensor (por ejemplo, un sensor de presión, la temperatura, volumen, flujo, peso), un dispositivo de
 5 manejo de fluidos, una impresora, una pantalla (por ejemplo, una LED, LCT o CRT), similares o combinaciones de los mismos. Por ejemplo, a veces, un operador de una máquina proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo. Con frecuencia, se configura un módulo para transferir datos y/o información a o desde otro módulo o máquina. Un módulo puede recibir datos y/o información de otro módulo, cuyos ejemplos no limitantes incluyen un módulo de procesamiento lógico, módulo de secuenciación, módulo de mapeo, módulo de
 10 recuento, módulo de filtrado, módulo de ponderación, módulo de normalización, módulo de sesgo de GC, módulo de nivel, módulo de comparación, módulo de configuración de intervalos, módulo de clasificación, módulo de representación gráfica, módulo de representación, módulo de relación, módulo de resultados y/o módulo de organización de visualización de datos, módulo de relación, módulo de resultados y/o módulo de organización de visualización de datos, similares o combinaciones de los mismos. Un módulo puede manipular y/o transformar datos
 15 y/o información. Los datos y/o la información derivada de o transformada por un módulo se pueden transferir a otra máquina y/o módulo adecuado, cuyos ejemplos no limitantes incluyen un módulo de procesamiento lógico, módulo de organización de visualización de datos, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de filtrado, módulo de ponderación, módulo de normalización, módulo de sesgo de GC, módulo de nivel, módulo de comparación, módulo de configuración de intervalos, módulo de clasificación, módulo de ajuste, módulo de
 20 representación gráfica, módulo de representación, módulo de relación, módulo de resultados y/o módulo de organización de visualización de datos, similares o combinaciones de los mismos. Una máquina que comprende un módulo puede comprender al menos un procesador. En algunos casos, una máquina que comprende un módulo recibe y/o proporciona datos y/o información. Una máquina que comprende un módulo puede incluir un procesador (por ejemplo, uno o más procesadores) cuyo procesador puede llevar a cabo y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) de un módulo. En algunos casos, un módulo opera con uno o más procesadores
 25 externos (por ejemplo, una red interna o externa, servidor, dispositivo de almacenamiento y/o red de almacenamiento (por ejemplo, una nube)).

30 *Módulo de Procesamiento Lógico*

En ciertos casos, un módulo de procesamiento lógico orquesta, controla, limita, organiza, ordena, distribuye, particiona, transforma y/o regula datos y/o información o transfiere datos y/o información hacia, y desde, uno o más módulos,
 35 periféricos o dispositivos.

Módulo de Organización de Visualización de Datos

En ciertos casos, un módulo de organización de visualización de datos procesa y/o transforma datos y/o información en un medio visual adecuado, cuyos ejemplos no limitantes incluyen imágenes, vídeo y/o texto (por ejemplo, números,
 40 letras y símbolos). En algunos casos, un módulo de organización de visualización de datos procesa, transforma y/o transfiere datos y/o información para su presentación en una pantalla adecuada (por ejemplo, un monitor, LED, LCD, CRT, similares o combinaciones de los mismos), una impresora, un periférico o dispositivo adecuado. En algunos casos, un módulo de organización de visualización de datos procesa, transforma datos y/o información en una representación visual de un genoma, cromosoma o parte de los mismos fetal o materno.

45 *Módulo de Secuenciación*

En algunos casos, un módulo de secuenciación obtiene, genera, recopila, ensambla, manipula, transforma, procesa, y/o transfiere lecturas de secuencias. Un "módulo de recepción de secuencia" como se usa en el presente documento es lo mismo que un "módulo de secuenciación". Una máquina que comprende un módulo de secuenciación puede ser
 50 cualquier máquina que determine la secuencia de un ácido nucleico utilizando una tecnología de secuenciación conocida en la técnica. En algunos casos, un módulo de secuenciación puede alinear, ensamblar, fragmentar, complementar, complementar de manera inversa, verificar errores o corregir errores en las lecturas de secuencias.

55 *Módulo de Mapeo*

Las lecturas de secuencias se pueden mapear por un módulo de mapeo o por una máquina que comprende un módulo de mapeo, cuyo módulo de mapeo generalmente mapea las lecturas a un genoma de referencia o un segmento del
 60 mismo. Un módulo de mapeo puede mapear lecturas de secuencias mediante un método adecuado conocido en la técnica. En algunos casos, se requiere un módulo de mapeo o una máquina que comprende un módulo de mapeo para proporcionar lecturas de secuencias mapeadas.

Módulo de Recuento

65 Los recuentos se pueden proporcionar por un módulo de recuento o por una máquina que comprende un módulo de recuento. En algunos casos, un módulo de recuento cuenta las lecturas de secuencia mapeadas en un genoma de

referencia. En algunos casos, un módulo de recuento genera, ensambla y/o proporciona recuentos de acuerdo con un método de recuento conocido en la técnica. En algunos casos, se requiere un módulo de recuento o una máquina que comprende un módulo de recuento para proporcionar recuentos.

5 *Módulo de Filtrado*

Las porciones de filtrado (por ejemplo, porciones de un genoma de referencia) pueden proporcionarse por un módulo de filtrado (por ejemplo, por una máquina que comprende un módulo de filtrado). En algunos casos, se requiere un módulo de filtrado para proporcionar datos de porciones filtradas (por ejemplo, porciones filtradas) y/o para eliminar porciones del estudio. En ciertos casos, un módulo de filtrado elimina los recuentos mapeados en una porción del estudio. En ciertos casos, un módulo de filtrado elimina los recuentos mapeados en una porción de una determinación de un nivel o un perfil. Un módulo de filtrado puede filtrar datos (por ejemplo, recuentos, recuentos mapeados en porciones, porciones, niveles de porción, recuentos normalizados, recuentos sin procesar y similares) mediante uno o más métodos de filtrado conocidos en la técnica o descritos en el presente documento.

15 *Módulo de Ponderación*

Las porciones de ponderación (por ejemplo, porciones de un genoma de referencia) pueden proporcionarse por un módulo de ponderación (por ejemplo, por una máquina que comprende un módulo de ponderación). En algunos casos, se requiere un módulo de ponderación para ponderar secciones de genómica y/o proporcionar valores de porciones ponderados. Un módulo de ponderación puede ponderar porciones mediante uno o más métodos de ponderación conocidos en la técnica o descritos en el presente documento.

25 *Módulo de Normalización*

Los datos normalizados (por ejemplo, recuentos normalizados) pueden proporcionarse por un módulo de normalización (por ejemplo, por una máquina que comprende un módulo de normalización). En algunos casos, se requiere un módulo de normalización para proporcionar datos normalizados (por ejemplo, recuentos normalizados) obtenidos de lecturas de secuenciación. Un módulo de normalización puede normalizar los datos (por ejemplo, recuentos, recuentos filtrados, recuentos sin procesar) mediante uno o más métodos de normalización descritos en el presente documento (por ejemplo, PERUN, normalización híbrida, similares o combinaciones de los mismos) o conocidos en la técnica.

35 *Módulo de Sesgo de GC*

La determinación del sesgo de GC (por ejemplo, la determinación del sesgo de GC para cada una de las porciones de un genoma de referencia (por ejemplo, porciones, porciones de un genoma de referencia)) puede proporcionarse por un módulo de sesgo de GC (por ejemplo, por una máquina que comprende un módulo de sesgo de GC). En algunos casos, se requiere un módulo de sesgo de GC para proporcionar una determinación del sesgo de GC. En algunos casos, un módulo de sesgo de GC proporciona una determinación del sesgo de GC a partir de una relación ajustada (por ejemplo, una relación lineal ajustada) entre los recuentos de lecturas de secuencia mapeadas en cada una de las porciones de un genoma de referencia y el contenido de GC de cada porción. Un módulo de sesgo de GC a veces es parte de un módulo de normalización (por ejemplo, el módulo de normalización PERUN).

45 *Módulo de Nivel*

Un módulo de nivel (por ejemplo, una máquina que comprende un módulo de nivel) puede proporcionar niveles de determinación (por ejemplo, niveles) y/o calcular niveles de sección genómica para porciones de un genoma de referencia. En algunos casos, se requiere un módulo de nivel para proporcionar un nivel o un nivel de sección genómica calculado (por ejemplo, de acuerdo con la Ecuación A, B, L, M, N, O y/o Q). En algunos casos, un módulo de nivel proporciona un nivel a partir de una relación ajustada (por ejemplo, una relación lineal ajustada) entre un sesgo de GC y los recuentos de lecturas de secuencia mapeadas en cada una de las porciones de un genoma de referencia. En algunos casos, un módulo de nivel calcula un nivel de sección genómica como parte de PERUN. En algunos casos, un módulo de nivel proporciona un nivel de sección genómica (es decir, L_i) de acuerdo con la ecuación $L_i = (m_i - G_i)S$ I^{-1} en donde G_i es el sesgo de GC, m_i es los recuentos medidos mapeados en cada porción de un genoma de referencia, i es un muestra, e I es la intersección y S es la pendiente de una relación ajustada (por ejemplo, una relación lineal ajustada) entre un sesgo de GC y los recuentos de lecturas de secuencias mapeadas en cada una de las porciones de un genoma de referencia.

60 *Módulo de Comparación*

Un primer nivel puede identificarse como significativamente diferente de un segundo nivel por un módulo de comparación o por una máquina que comprende un módulo de comparación. En algunos casos, se requiere un módulo de comparación o una máquina que comprenda un módulo de comparación para proporcionar una comparación entre dos niveles.

Módulo de Configuración de Intervalos

Los intervalos esperados (p. ej., intervalos de nivel esperados) para diversas variaciones en el número de copias (p. ej., duplicaciones, inserciones y/o deleciones) o intervalos para la ausencia de una variación en el número de copias pueden proporcionarse por un módulo de configuración de intervalos o por una máquina que comprenda un módulo de configuración de intervalos. En ciertos casos, un módulo de configuración de intervalos o una máquina que comprende un módulo de configuración de intervalos, proporcionan niveles esperados. En algunos casos, se requiere un módulo de configuración de intervalos o una máquina que comprenda un módulo de configuración de intervalos para proporcionar los niveles y/o intervalos esperados.

Módulo de Clasificación

Una variación en el número de copias (p. ej., una variación en el número de copias maternas y/o fetales, una variación en el número de copias fetales, una duplicación, inserción, deleción) se pueden clasificar por un módulo de clasificación o por una máquina que comprende un módulo de clasificación. En ciertos casos, un módulo de clasificación clasifica una variación en el número de copias (p. ej., una variación en el número de copias maternas y/o fetales). En ciertos casos, un módulo de clasificación identifica como representativo de una variación en el número de copias, un nivel (por ejemplo, un primer nivel) determinado para ser significativamente diferente que otro nivel (por ejemplo, un segundo nivel). En ciertos casos, un módulo de clasificación determina la ausencia de una variación en el número de copias. En algunos casos, una máquina que comprende un módulo de clasificación determina una determinación de una variación en el número de copias. Un módulo de clasificación puede estar especializado para clasificar una variación en el número de copias maternas y/o fetales, una variación en el número de copias fetales, una duplicación, deleción o inserción o carencia de las mismas o una combinación de las anteriores. Por ejemplo, un módulo de clasificación que identifica una eliminación materna puede ser diferente a y/o distinto de un módulo de clasificación que identifica una duplicación fetal. En algunos casos, se requiere un módulo de clasificación o una máquina que comprenda un módulo de clasificación para identificar una variación en el número de copias o un resultado determinante de una variación en el número de copias.

Módulo de Ajuste

En algunos casos, los ajustes de un nivel (por ejemplo, ajustes a los niveles de la sección genómica, un nivel de un perfil, un nivel de variación en el número de copias, un nivel de una o más porciones, similares o combinaciones de los mismos) se realizan mediante un módulo de ajuste o mediante una máquina que comprende un módulo de ajuste. En algunos casos, se requiere requerir un módulo de ajuste o una máquina que comprenda un módulo de ajuste para ajustar un nivel. Un nivel ajustado por los métodos descritos en el presente documento puede verificarse y/o ajustarse independientemente mediante pruebas adicionales (por ejemplo, mediante secuenciación dirigida de ácido nucleico materno o fetal).

Módulo de Representación Gráfica

En algunos casos, un módulo de representación gráfica procesa y/o transforma datos y/o información en un medio visual adecuado, cuyos ejemplos no limitantes incluyen un esquema, representación gráfica, gráfico, similares o combinaciones de los mismos. En algunos casos, un módulo de representación gráfica procesa, transforma y/o transfiere datos y/o información para su presentación en una pantalla adecuada (por ejemplo, un monitor, LED, LCD, CRT, similares o combinaciones de los mismos), una impresora, un periférico o dispositivo adecuado. En ciertos casos, un módulo de representación gráfica proporciona una visualización visual de un recuento, un nivel y/o un perfil. En algunos casos, un módulo de organización de visualización de datos procesar, transforma datos y/o información en una representación visual de un genoma, cromosoma o parte de los mismos fetal o materno. En algunos casos se requiere un módulo de representación gráfica o una máquina que comprenda un módulo de representación gráfica para representar gráficamente un recuento, un nivel o un perfil.

Módulo de Relación

En ciertos casos, un módulo de relación procesa y/o transforma datos y/o información en una relación. En ciertos casos, una relación se generar por y/o se transferirse desde un módulo de relación.

Módulo de Resultados

En algunos casos, la presencia o ausencia de una variación genética (una aneuploidía, una aneuploidía fetal, una variación en el número de copias) se identifica por un módulo de resultados o por una máquina que comprende un módulo de resultados. En ciertos casos, un módulo de resultados identifica una variación genética. Con frecuencia, la determinación de la presencia o ausencia de una aneuploidía se identifica mediante un módulo de resultados. En algunos casos, un resultado determinante de una variación genética (una aneuploidía, una variación en el número de copias) puede identificarse mediante un módulo de resultados o mediante una máquina que comprenda un módulo de resultados. Un módulo de resultados puede especializarse para determinar una variación genética específica (por ejemplo, trisomía, una trisomía 21, una trisomía 18). Por ejemplo, un módulo de resultados que identifica una trisomía

21 puede ser diferente a y/o distinto de un módulo de resultados que identifica una trisomía 18. En algunos casos, se requiere un módulo de resultados o una máquina que comprende un módulo de resultados para identificar una variación genética o un resultado determinante de una variación genética (por ejemplo, una aneuploidía, una variación en el número de copias). Una variación genética o un resultado determinante de una variación genética identificada por los métodos descritos en el presente documento puede verificarse independientemente mediante pruebas adicionales (por ejemplo, mediante secuenciación dirigida de ácido nucleico materno y/o fetal).

Transformaciones

10 Como se ha señalado anteriormente, los datos a veces se transforman de una forma a otra forma. Los términos "transformado", "transformación", y derivaciones gramaticales o equivalentes de los mismos, como se usan en el presente documento, se refieren a una alteración de los datos de un material de partida físico (por ejemplo, ácido nucleico del sujeto de ensayo y/o del sujeto de referencia) en una representación digital del material de partida físico (por ejemplo, datos de lecturas de secuencias), y en algunos casos, incluye una transformación adicional en uno o más valores numéricos o representaciones gráficas de la representación digital que puede utilizarse para proporcionar un resultado. En ciertos casos, el uno o más valores numéricos y/o representaciones gráficas de datos representados digitalmente se utilizan para representar la apariencia del genoma físico de un sujeto de ensayo (por ejemplo, representar virtualmente o visualmente la presencia o ausencia de una inserción, duplicación o delección genómica; representar la presencia o ausencia de una variación en la cantidad física de una secuencia asociada con afecciones médicas). Una representación virtual a veces se transforma adicionalmente en uno o más valores numéricos o representaciones gráficas de la representación digital del material de partida. Estos métodos pueden transformar el material de partida físico en un valor numérico o una representación gráfica, o una representación de la apariencia física del genoma de un sujeto de ensayo.

25 En algunos casos, la transformación de un conjunto de datos facilita proporcionar un resultado al reducir la complejidad de los datos y/o la dimensionalidad de los datos. La complejidad del conjunto de datos a veces se reduce durante el proceso de transformación de un material de partida físico en una representación virtual del material de partida (por ejemplo, las lecturas de secuencias representativas del material de partida físico). Se puede utilizar una característica o variable adecuada para reducir la complejidad y/o dimensionalidad del conjunto de datos. Ejemplos no limitantes de características que pueden elegirse para su uso como una característica diana para el procesamiento de datos incluyen contenido de GC, predicción del sexo fetal, tamaño del fragmento (por ejemplo, longitud de fragmentos CCF, lecturas o una representación adecuada de las mismas (por ejemplo, FRS)), secuencia de fragmentos, identificación de aneuploidía cromosómica, identificación de genes o proteínas particulares, identificación de cáncer, enfermedades, genes/rasgos heredados, anomalías cromosómicas, una categoría biológica, una categoría química, una categoría bioquímica, una categoría de genes o proteínas, una ontología genética, una ontología proteica, genes corregulados, genes de señalización celular, genes del ciclo celular, proteínas pertenecientes a los genes anteriores, variantes de genes, variantes de proteínas, genes corregulados, proteínas correguladas, secuencia de aminoácidos, secuencia de nucleótidos, datos de estructura de proteínas y similares y combinaciones de los anteriores. Ejemplos no limitantes de la reducción de la complejidad y/o de la dimensionalidad del conjunto de datos incluyen; reducción de una pluralidad de lecturas de secuencias para representaciones gráficas de perfil, reducción de una pluralidad de lecturas de secuencias a valores numéricos (por ejemplo, valores normalizados, puntuaciones Z, valores de p); reducción de múltiples métodos de análisis a representaciones gráficas de probabilidad o puntos únicos; análisis de componentes principales de cantidades derivadas; y similares o combinaciones de los mismos.

45 *Ciertos sistemas, máquinas y productos de programas informáticos*

En ciertos aspectos, se proporciona un método implementado por ordenador para determinar la presencia o ausencia de una variación genética, que comprende (a) obtener recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son: (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácidos nucleicos que tienen longitudes que son menores que la longitud de un fragmento seleccionado; (b) normalizar los recuentos, generando, de este modo, recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas; y (c) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

55 También se proporciona, en ciertos aspectos, un sistema que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por el uno o más procesadores y cuya memoria comprende recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas son (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácidos nucleicos que tienen longitudes que son menores que la longitud de un fragmento seleccionado; y cuyas instrucciones ejecutables por uno o más microprocesadores están configuradas para (a) normalizar los recuentos, generando, de este modo, recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas; y (b) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

65 También se proporciona, en ciertos aspectos, una máquina que comprende uno o más procesadores y memoria, cuya memoria comprende instrucciones ejecutables por uno o más procesadores y cuya memoria comprende recuentos de

lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácidos nucleicos que tienen longitudes que son menores que una longitud de un fragmento seleccionado; y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencia mapeadas en las secciones genómicas; y (b) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

También se proporciona, en ciertos casos, un producto de programa informático materializado de forma tangible en un medio legible por ordenador, que comprende instrucciones que cuando se ejecutan por uno o más procesadores están configuradas para (a) acceder a recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencia son: (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y (ii) lecturas de fragmentos de ácidos nucleicos que tienen longitudes que son menores que una longitud de un fragmento seleccionado (b) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas; y (c) determinar la presencia o ausencia de una variación genética de acuerdo con los recuentos normalizados.

En ciertos casos, un sistema, máquina y/o producto de programa de computadora comprenden un módulo de recuento configurado para contar lecturas mapeadas en secciones genómicas de un genoma de referencia o parte del mismo (por ejemplo, un subconjunto de secciones genómicas, un conjunto seleccionado de secciones genómicas). Un módulo de recuento a menudo está configurado para contar lecturas de fragmentos de ácido nucleico que tienen longitudes que son menores que una longitud de un fragmento seleccionado. Los recuentos a veces son recuentos sin procesar, filtrados, normalizados o una combinación de los anteriores. En algunos casos, un módulo de recuento puede normalizar los recuentos, por ejemplo, usando cualquier proceso de normalización adecuado descrito en el presente documento o conocido en la técnica.

En ciertos casos, un sistema, máquina y/o producto de programa informático comprenden un módulo de comparación de recuentos. Con frecuencia, un módulo de comparación de recuentos se configura para comparar el número de recuentos de lecturas contadas por un módulo de recuento, haciendo, de este modo, una comparación de recuentos. Con frecuencia, un módulo de comparación de recuentos se configura para acceder, recibir, utilizar, almacenar, buscar y/o alinear recuentos de lecturas (por ejemplo, desde un módulo de recuento o un módulo de normalización). Con frecuencia, un módulo de comparación de recuentos se configura para proporcionar una comparación adecuada entre recuentos, cuyos ejemplos no limitantes incluyen una comparación simple (por ejemplo, coincidencia o no coincidencia entre recuentos de lecturas mapeadas en un primer conjunto de secciones genómicas en comparación con un segundo conjunto de secciones genómicas), comparación matemática (por ejemplo, proporción, porcentaje), comparación estadística (por ejemplo, comparaciones múltiples, pruebas múltiples, estandarización (por ejemplo, análisis de puntuación z)), similares y combinaciones de los mismos. Se puede proporcionar un valor de comparación de recuentos adecuada mediante un módulo de comparación de recuentos, cuyos ejemplos no limitantes incluyen la presencia o ausencia de una coincidencia entre recuentos, una relación, porcentaje, puntuación z, un valor acoplado con una medida de varianza o incertidumbre (por ejemplo, desviación estándar, desviación absoluta media, intervalo de confianza), similares y combinaciones de los mismos. Un módulo de comparación de recuentos a veces se configura para transmitir un valor de comparación a otro módulo o máquina, tal como un módulo de variación genética, una máquina de visualización o una impresora, por ejemplo.

En ciertos casos, un sistema, máquina y/o producto de programa informático comprenden un módulo de variación genética. Con frecuencia, un módulo de variación genética se configura para proporcionar una determinación de la presencia o ausencia de una variación genética de acuerdo con los recuentos de lecturas mapeadas en secciones genómicas de un genoma de referencia. Con frecuencia, un módulo de variación genética se configura para proporcionar una determinación de la presencia o ausencia de una variación genética de acuerdo con una comparación de recuentos. Con frecuencia, un módulo de variación genética se configura para acceder, recibir, utilizar, almacenar, buscar y/o alinear una o más comparaciones de un módulo de comparación de recuentos y/o recuentos de un módulo de recuento. Un módulo de variación genética puede determinar la presencia o ausencia de una variación genética de una o más comparaciones o de recuentos de una manera adecuada. Un módulo de variación genética a veces determina si existe una diferencia significativa entre los recuentos para diferentes conjuntos de secciones genómicas en un genoma de referencia. Se puede determinar, de una manera adecuada, la importancia de una diferencia mediante un módulo de variación genética (por ejemplo, diferencia porcentual, análisis de puntuación z). Un módulo de variación genética a veces determina si una determinación de recuentos o una comparación de recuentos se encuentra en una categoría particular. Por ejemplo, un módulo de variación genética puede clasificar una comparación particular a un umbral de relación particular o un intervalo de relaciones asociadas con una determinación euploide, o un umbral de relación particular o intervalo de relaciones asociadas con una determinación aneuploide. En otro ejemplo no limitante, un módulo de variación genética puede clasificar una determinación de recuentos particular a un umbral de recuentos particular o un intervalo de recuentos asociados con una determinación euploide, o un umbral de recuentos particular o un intervalo de recuentos asociados con una determinación aneuploide. Un módulo de variación genética puede proporcionar un resultado en un formato adecuado, que a veces es una identificación relacionada con una variación genética opcionalmente asociada con una medida de varianza o incertidumbre (por ejemplo, desviación estándar, desviación absoluta media, precisión (por ejemplo, dentro de un intervalo de confianza particular). Un módulo

de variación genética a veces se configura para transmitir una determinación de la presencia o ausencia de una variación genética da otro módulo o máquina, tal como una máquina de visualización o una impresora, por ejemplo.

5 Una máquina o sistema que comprende un módulo descrito en el presente documento (por ejemplo, un módulo de comparación de referencia) puede comprender uno o más procesadores. En algunos casos, una máquina o sistema puede incluir múltiples procesadores, tales como procesadores coordinados y que trabajan en paralelo. Un procesador (por ejemplo, uno o más procesadores) en un sistema o máquina puede realizar y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) en un módulo descrito en el presente documento. Un módulo descrito, en el presente documento, a veces se encuentra en la memoria o está asociado con una máquina o sistema. En algunos casos, un módulo descrito en el presente documento, opera con uno o más procesadores externos (por ejemplo, una red interna o externa, servidor, dispositivo de almacenamiento y/o red de almacenamiento (por ejemplo, una nube)). En algunos casos, un módulo descrito en el presente documento, se configura para acceder, recopilar, ensamblar y/o recibir datos y/o información de otro módulo, máquina o sistema (por ejemplo, componente, periférico). En algunos casos, un módulo descrito en el presente documento, se configura para proporcionar y/o transferir datos y/o información a otro módulo, máquina o sistema (por ejemplo, componente, periférico). En algunos casos, un módulo descrito en el presente documento se configura para acceder, aceptar, recibir y/o recopilar datos de entrada y/o información de un operador de una máquina o sistema (es decir, usuario). Por ejemplo, a veces un usuario proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo. Un módulo descrito en el presente documento, a veces, se configura para transformar datos y/o información a los que accede, recibe, recopila y/o ensambla.

En ciertos casos, un sistema, máquina y/o producto de programa informático comprende (i) un módulo de secuenciación configurado para obtener y/o acceder a lecturas de secuencias de ácidos nucleicos y/o lecturas de secuencias de nucleótidos parciales; (ii) un módulo de mapeo configurado para mapear las lecturas de secuencias de ácidos nucleicos a porciones de un genoma de referencia; (iii) un módulo de recuento configurado para proporcionar recuentos de lecturas de secuencias de ácidos nucleicos mapeadas en porciones de un genoma de referencia; (iv) un módulo de normalización configurado para proporcionar recuentos normalizados; (v) un módulo de comparación configurado para proporcionar una identificación de una primera elevación que es significativamente diferente de una segunda elevación; (vi) un módulo de configuración de intervalos configurado para proporcionar uno o más intervalos de nivel esperados; (vii) un módulo de clasificación configurado para identificar una elevación representativa de una variación en el número de copias; (viii) un módulo de ajuste configurado para ajustar un nivel identificado como una variación en el número de copias; (ix) módulo de representación gráfica configurado para representar gráficamente y mostrar un nivel y/o un perfil; (x) un módulo de resultados configurado para determinar la presencia o ausencia de una variación genética, o determinar un resultado (por ejemplo, un resultado determinante de la presencia o ausencia de una aneuploidía fetal); (xi) un módulo de organización de visualización de datos configurado para mostrar una determinación de variación genética; (xii) un módulo de procesamiento lógico configurado para realizar una o más lecturas de secuencias mapa, contar las lecturas de secuencias mapeadas, normalizar los recuentos y generar un resultado; (xiii) un módulo de comparación de recuentos, (xiv) módulo de fracción fetal configurado para proporcionar una determinación de fracción fetal; (xv) un módulo de variación genética configurado para proporcionar una determinación de la presencia o ausencia de una variación genética; o (xvi) combinación de dos o más de los anteriores.

En algunos casos, se configura un módulo de secuenciación y un módulo de mapeo para transferir las lecturas de secuencias desde el módulo de secuenciación al módulo de mapeo. El módulo de mapeo y el módulo de recuento a veces están configurados para transferir lecturas de secuencias mapeadas desde el módulo de mapeo al módulo de recuento. En algunos casos, el módulo de normalización y/o el módulo de comparación se configuran para transferir recuentos normalizados al módulo de comparación y/o al módulo de configuración de intervalos. En algunos casos, el módulo de comparación, el módulo de configuración de intervalos y/o el módulo de clasificación están configurados de manera independiente para transferir (i) una identificación de una primera elevación que es significativamente diferente a una segunda elevación y/o (ii) un intervalo de nivel esperado del módulo de comparación y/o del módulo de configuración de intervalos al módulo de clasificación. En ciertos casos, el módulo de clasificación y el módulo de ajuste se configuran para transferir una elevación clasificada como una variación en el número de copias del módulo de clasificación al módulo de ajuste. En algunos casos, el módulo de ajuste, módulo de representación gráfica y el módulo de resultados se configuran para transferir uno o más niveles ajustados del módulo de ajuste al módulo de representación gráfica o módulo de resultados. El módulo de normalización a veces se configura para transferir los recuentos de lecturas de secuencias normalizadas mapeadas en uno o más del módulo de comparación, módulo de configuración de intervalos, módulo de clasificación, módulo de ajuste, módulo de resultados o módulo de representación gráfica.

60 Ejemplos

Los siguientes ejemplos se proporcionan solo como ilustración y no como limitación. Por lo tanto, los ejemplos que se exponen a continuación ilustran ciertas realizaciones y no limitan la tecnología.

65

Ejemplo 1: PERUN y métodos generales para detectar afecciones asociadas con variaciones genéticas.

Los métodos y la teoría subyacente descritos en el presente documento pueden utilizarse para detectar varias afecciones asociadas con la variación genética y proporcionar un resultado determinante de, o determinar la presencia o ausencia de una variación genética.

Eliminación de porciones no informativas de un genoma de referencia

Los intentos múltiples de eliminar porciones no informativas de un genoma de referencia han indicado que la selección de porciones tiene el potencial de mejorar la clasificación.

Ecuación A:

$$M = LI + GS \tag{A}$$

Los diversos términos en la ec. A tiene los siguientes significados:

- **M:** recuentos medidos, que representan la información primaria contaminada por una variación no deseada.
- **L:** nivel cromosómico: este es el resultado deseado del procedimiento de procesamiento de datos. *L* indica aberraciones fetales y/o maternas de euploides. Esta es la cantidad que se enmascara tanto por errores estocásticos como por los sesgos sistemáticos. El nivel cromosómico *L* es tanto específico de muestra como específico de porción.
- **G:** Coeficiente de sesgo de GC medido utilizando un modelo lineal, LOESS, o cualquier aproximación equivalente. *G* representa información secundaria, extraída de *M* y de un conjunto de valores de contenido de GC específico de porción, generalmente derivados del genoma de referencia (pero también pueden derivar de los contenidos de GC observados en realidad). *G* es específico de la muestra y no varía a lo largo de la posición genómica. Engloba una porción de la variación no deseada.
- **I:** Intersección del modelo lineal. Este parámetro del modelo es fijo para una configuración experimental dada, independiente de la muestra, y específico de porción.
- **S:** Pendiente del modelo lineal. Este parámetro del modelo es fijo para una configuración experimental dada, independiente de la muestra, y específico de porción.

Se miden las cantidades *M* y *G*. Inicialmente, los valores específicos de porción *I* y *S* son desconocidos. Para evaluar *I* y *S* desconocidos, debemos asumir que *L* = 1 para todas las porciones de un genoma de referencia en muestras euploides. El supuesto no siempre es cierto, pero es razonable esperar que cualquier muestra con deleciones/duplicaciones se vea superada por muestras con niveles cromosómicos normales. Un modelo lineal aplicado a las muestras euploides extrae los valores de los parámetros *I* y *S* específicos para la porción seleccionada (suponiendo que *L* = 1). El mismo procedimiento se aplica a todas las porciones de un genoma de referencia en el genoma humano, produciendo un conjunto de intersecciones/y pendientes *S* para cada ubicación genómica. La validación cruzada selecciona aleatoriamente un conjunto de trabajo que contiene el 90% de todos los euploides LDTv2CE y utiliza ese subconjunto para entrenar el modelo. La selección aleatoria se repite 100 veces, produciendo un conjunto de 100 pendientes y 100 intersecciones por cada porción.

Extracción del Nivel Cromosómico de los Recuentos Medidos

Suponiendo que los valores de los parámetros del modelo *I* y *S* están disponibles para cada porción, las mediciones *M* recogidas en una nueva muestra de ensayo se utilizan para evaluar el nivel cromosómico de acuerdo con la siguiente Ecuación B:

$$L = (M - GS)/I \tag{B}$$

Como en la Ec. A, el coeficiente *G* de sesgo de GC se evalúa como la pendiente de la regresión entre los recuentos *M* sin procesar medidos por porciones y el contenido de GC del genoma de referencia. El nivel cromosómico *L* se usa a continuación, para análisis adicionales (valores *Z*, deleciones/duplicaciones maternas, microdeleciones/microduplicaciones fetales, sexo fetal, aneuploidías sexuales, etc.). El procedimiento englobado por la Ec. B se denomina Eliminación de Error Parametrizado y Normalización Imparcial (PERUN).

Ejemplo 2: Ejemplos de fórmulas

A continuación se proporcionan ejemplos no limitantes de fórmulas matemáticas y/o estadísticas que se pueden usar en los métodos descritos en el presente documento.

Las puntuaciones Z y los valores de p calculados a partir de las puntuaciones Z asociadas con desviaciones del nivel esperado de 1 pueden evaluarse a la luz de la estimación de incertidumbre en el nivel promedio. Los valores de p se basan en una distribución t cuyo orden está determinado por el número de porciones de un genoma de referencia en un máximo. Dependiendo del nivel de confianza deseado, un valor de corte puede suprimir el ruido y permitir la detección inequívoca de la señal real.

Ecuación 1:

$$Z = \frac{\Delta_1 - \Delta_2}{\sqrt{\sigma_1^2 \left(\frac{1}{N_1} + \frac{1}{n_1} \right) + \sigma_2^2 \left(\frac{1}{N_2} + \frac{1}{n_2} \right)}} \quad (1)$$

La ecuación 1 se puede usar para comparar directamente el nivel máximo de dos muestras diferentes, donde N y n se refieren a los números de porciones de un genoma de referencia en todo el cromosoma y dentro de la aberración, respectivamente. El orden de la prueba t que dará un valor de p que mide la similitud entre dos muestras se determina por el número de porciones de un genoma de referencia en el más corto de los dos tramos desviados.

La ecuación 8 se puede utilizar para incorporar los recuentos de fracción fetal, ploidía materna y mediana de referencia en un esquema de clasificación para determinar la presencia o ausencia de una variación genética con respecto a la aneuploidía fetal.

Ecuación 8:

$$y_i = (1 - F)M_i f_i + F X f_i \quad (8)$$

donde Y_i representa los recuentos medidos para una porción en la muestra de ensayo correspondiente a la porción en el perfil de recuentos mediana, F representa la fracción fetal, X representa la ploidía fetal y M_i representa ploidía materna asignada a cada porción. Los posibles valores utilizados para X en la ecuación (8) son: 1 si el feto es euploide; 3/2, si el feto es triploide; y, 5/4, si hay fetos gemelos y uno se ve afectado y el otro no. 5/4 se usa en el caso de gemelos en los que un feto está afectado y el otro no, porque el término F en la ecuación (8) representa el ADN fetal total, por lo tanto, se debe tener en cuenta todo el ADN fetal. En algunos casos, las grandes deleciones y/o duplicaciones en el genoma materno pueden explicarse asignando ploidía materna, M_i , a cada porción o porción. Con frecuencia, la ploidía materna se asigna como un múltiplo de 1/2, y en algunos casos puede estimarse utilizando la normalización por porciones. Debido a que, con frecuencia, la ploidía materna es un múltiplo de 1/2, la ploidía materna se puede explicar fácilmente, y por lo tanto no se incluirá en ecuaciones adicionales para simplificar derivaciones.

Al evaluar la ecuación (8) en $X = 1$, (por ejemplo, suposición euploide), la fracción fetal se cancela y se obtiene la siguiente ecuación para la suma de los restos al cuadrado.

Ecuación 9:

$$\varphi_E = \sum_{i=1}^N \frac{1}{\sigma_i^2} (y_i - f_i)^2 = \sum_{i=1}^N \frac{y_i^2}{\sigma_i^2} - 2 \sum_{i=1}^N \frac{y_i f_i}{\sigma_i^2} + \sum_{i=1}^N \frac{f_i^2}{\sigma_i^2} = E_{yy} - 2E_{fy} + E_{ff} \quad (9)$$

Para simplificar la ecuación (9) y los cálculos subsiguientes, se utilizan las siguientes ecuaciones.

Ecuación 10:

$$E_{yy} = \sum_{i=1}^N \frac{y_i^2}{\sigma_i^2} \quad (10)$$

Ecuación 11:

$$E_{ff} = \sum_{i=1}^N \frac{f_i^2}{\sigma_i^2} \quad (11)$$

Ecuación 12:

$$E_{fy} = \sum_{i=1}^N \frac{y_i f_i}{\sigma_i^2} \quad (12)$$

Al evaluar la ecuación (8) en $X = 3/2$ (por ejemplo, suposición triploide), se obtiene la siguiente ecuación para la suma de los restos al cuadrado.

Ecuación 13:

$$\varphi_T = \sum_{i=1}^N \frac{1}{\sigma_i^2} (y_i - f_i - \frac{1}{2} F f_i)^2 = \mathcal{E}_{yy} - 2\mathcal{E}_{fy} + \mathcal{E}_{ff} + F(\mathcal{E}_{ff} - \mathcal{E}_{fy}) + \frac{1}{4} F^2 \mathcal{E}_{ff} \quad (13)$$

La diferencia entre las ecuaciones (9) y (13) forma el resultado funcional (por ejemplo, phi) que se puede usar para probar la hipótesis nula (por ejemplo, euploide, X= 1) en comparación con la hipótesis alternativa (por ejemplo, trisomía de feto único, X = 3/2):

Ecuación 14:

$$\varphi = \varphi_E - \varphi_T = F(\mathcal{E}_{fy} - \mathcal{E}_{ff}) - \frac{1}{4} F^2 \mathcal{E}_{ff} \quad (14)$$

10 Ecuación 18:

$$\begin{aligned} \varphi &= \sum_{i=1}^N \frac{1}{\sigma_i^2} [y_i - (1 - F)M_i f_i - F X f_i]^2 \\ &= \sum_{i=1}^N \frac{1}{\sigma_i^2} [y_i^2 - 2(1 - F)M_i f_i y_i - 2 F X f_i y_i + (1 - F)^2 M_i^2 f_i^2 + 2F(1 - F) X M_i f_i^2 + F^2 X^2 f_i^2] \end{aligned} \quad (18)$$

15

El valor óptimo de ploidía a veces viene dado por la Ecuación 20:

$$X = \frac{\sum_{i=1}^N \frac{f_i y_i}{\sigma_i^2} - (1 - F) \sum_{i=1}^N \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^N \frac{f_i^2}{\sigma_i^2}} \quad (20)$$

20

El término para ploidía materna, M_i , se puede omitir de algunas derivaciones matemáticas. La expresión resultante para X corresponde con el caso especial relativamente simple, y con frecuencia que se produce de manera más frecuente, en donde la madre no tiene deleciones o duplicaciones en el cromosoma o cromosomas que se están evaluando.

25

Ecuación 21:

$$X = \frac{\mathcal{E}_{fy} - (1 - F)\mathcal{E}_{ff}}{F\mathcal{E}_{ff}} = \frac{\mathcal{E}_{fy}}{F\mathcal{E}_{ff}} - \frac{1 - F}{F} = 1 + \frac{1}{F} \left(\frac{\mathcal{E}_{fy}}{\mathcal{E}_{ff}} - 1 \right) \quad (21)$$

30

X_{if} y X_{if_y} están dados por las ecuaciones (11) y (12), respectivamente. En los casos en que todos los errores experimentales son despreciables, resolver la ecuación (21) da como resultado un valor de 1 para los euploides donde $X_{if} = X_{if_y}$. En ciertos casos donde todos los errores experimentales son despreciables, resolver la ecuación (21) da como resultado un valor de 3/2 para los triploides (véase la ecuación (15) para la relación triploide entre X_{if} y X_{if_y}).

35

Tabla 2

Estado Gestación	de	Cr21 fetal	Cr18 fetal	Cr13 fetal	CrX fetal	CrY fetal
Mujer T21		$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Mujer T18		$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Mujer T13		$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
Varón T21		$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Varón T18		$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Varón T13		$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
Varón Euploide		$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$

Estado de Gestación	Cr21 fetal	Cr18 fetal	Cr13 fetal	CrX fetal	CrY fetal
Turner	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 0$
Jacobs	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1$
Klinefelter	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$
TripleX	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 0$

Ejemplo 3: Selección de porciones usando FRS.

5 Las porciones del genoma de referencia humano designado HG19 primero se prefiltraron utilizando un método basado en PERUN que elimina las porciones con alta variabilidad, baja capacidad de mapeo y se une con un gran porcentaje de elementos repetitivos. Se excluyeron las porciones (seleccionadas para LDTv2) con alta variabilidad, baja capacidad de mapeo y una gran fracción de secuencias repetitivas. Para cada porción de 50 kb (por ejemplo, porción), se calculó un estadístico de relación fetal para lecturas de secuencia final apareadas de fragmentos CCF de menos de 150 bases y de fragmentos CCF de menos de 600 bases. La FRS se promedió luego a través de 264 muestras no agrupadas procesadas utilizando la preparación de la biblioteca bioquímica TruSeq con limpieza automática de las perlas. Las porciones con FRS > mediana (FRS) se seleccionaron y se muestran en la TABLA 4 con referencia a las posiciones de inicio y final específicas del cromosoma. Las posiciones de inicio y final específicas de los cromosomas en la TABLA 4 ponen referencias de posiciones de base de nucleótidos en el genoma de referencia humano HG19.

15 Todas las porciones con FRS > mediana (FRS) se representaron gráficamente de manera simultánea con el número de posiciones de inicio únicas de exones en cada porción respectiva. Se mostró una correlación significativa para las regiones de los genes que contienen una representación excesiva de pequeños fragmentos (FIG. 1 - 9). Se mostró una correlación significativamente más fuerte con el contenido de GC (porcentaje de bases de GC en una porción de 50 kb) y FRS (Tabla 3).

20 La selección de la porción se restringió aún más a las porciones (es decir, las porciones) del genoma donde FRS > mediana (FRS) para la detección de trisomía cromosómica. La aplicación de esta aproximación en un conjunto de datos preliminares de 264 muestras proporcionó márgenes de clasificación consistentes a pesar de descartar el 50% de los datos. Por el contrario, al restringir las porciones en las que FRS < mediana (FRS), el margen de clasificación se redujo drásticamente, sugiriendo una dilución del ADN fetal para los análisis (FIG. 10 - 11).

30 En la FIG. 10 y la FIG. 11 hay dos líneas de regresión, una para muestras que no son T21 solamente (línea de puntos) y la otra para muestras de T21 (línea de guiones y puntos). La línea de regresión para muestras de T21 basada en porciones de FRS alta estaba por encima de la línea de regresión para muestras no T21 basadas en FRS alta (FIG. 11). Por el contrario, esta regresión similar fue más baja que las muestras no T21 cuando se compararon las puntuaciones Z calculadas en porciones de FRS baja (FIG. 10). Esto sugiere que el uso de porciones de FRS alta puede mejorar la precisión de las determinaciones de los resultados, ya que las puntuaciones Z tienden a ser mayores para las muestras de T21.

35

TABLA 3
Correlación parcial de Spearman [Corr (X,Y|Z)]

	Correlación Parcial		
	Truseq_FRS	recuentodegenesporventana	contenidogc
Truseq_FRS	1,000		
recuentodegenesporventana	0,350	1,000	
contenidogc	0,543	0,115	1,000
valordeP			
	Truseq_FHS	recuentodegenesporventana	contenidogc
Truseq_FHS	0		
recuentodegenesporventana	0	0	
contenidogc	0	2,66E-148	0

Ejemplo 4: Detección de trisomía 21 usando una combinación de separación basada en la secuencia y análisis basado en la longitud

40 Las muestras de plasma que contienen ADN libre de células circulante obtenidas de gestantes se analizan para

determinar la trisomía 21 utilizando el siguiente método.

Separación basada en la secuencia

5 Agilent obtiene una biblioteca de captura personalizada SURESELECT que incluye un conjunto de ARN de captura biotinilados diseñados a medida. Los ARN de captura están diseñados de acuerdo con secuencias de nucleótidos específicas del cromosoma 21 (cromosoma de ensayo) y específicas del cromosoma 14 (cromosoma de referencia) y se identifican mediante la herramienta de diseño basada en la web EARRAY de Agilent. Se han diseñado 100 ARN de captura independientes para el cromosoma 14 y el cromosoma 21. Se seleccionan las secuencias de nucleótidos de copia única en el intervalo de 40 a 60 pares de bases que son exclusivas del cromosoma 14 o 21 y son ricas en AT para el diseño de ARN de captura personalizado.

15 El ácido nucleico de muestra, que es ácido nucleico del plasma circulante libre de células de una hembra gestante en el primer trimestre de gestación, se divide en dos tubos y se incuba con el ARN de captura del cromosoma 21 o el ARN de captura del cromosoma 14 durante 24 horas a 65°C, de acuerdo con las instrucciones del fabricante. Tras la hibridación, los fragmentos diana capturados y los fragmentos de referencia capturados (denominados conjuntamente fragmentos capturados) se seleccionan mediante la eliminación de los híbridos ARN biotinilado/fragmento usando perlas magnéticas recubiertas con estreptavidina (DYNAL DYNAMAG-2, Invitrogen, Carlsbad, CA), y purifican con el Kit de purificación por PCR MINELUTE (Qiagen, Germantown, MD). El ARN de captura se digiere y los fragmentos de ADN restantes se amplifican de acuerdo con las instrucciones del fabricante.

Análisis basado en la longitud

25 Las muestras que contienen fragmentos de ácidos nucleicos separados anteriormente se hibridan en condiciones de hibridación no rigurosas a sondas de poliinosina que comprenden inosina biotinilada, cuyas sondas son más largas que los fragmentos de ADN con los que hibridan y tienen una longitud de 500 pares de bases. En algunos casos, la hibridación se realiza durante la noche a 65 °C en 6xSSC y SDS al 1%. La hibridación se puede realizar durante la noche a 43 °C en NaCl 1,0 M, tampón fosfato de sodio 50 mM (pH 7,4), EDTA 1,0 mM, dodecil sulfato sódico al 2% (p/v), gelatina al 0,1% (p/v), 50 µg/ml de ARNt y formamida al 30% (v/v). Se realizan cuatro lavados de 30 minutos a 55 °C en 1.2X de SSC (1X de SSC es NaCl 0,15 M más citrato de sodio 0,015 M), fosfato de sodio 10 mM (pH 7.4), EDTA 1,0 mM y dodecil sulfato de sodio al 0,5% (p/v) sulfato. Tras la hibridación, las porciones de la sonda no hibridadas se digieren usando Exonucleasa I (New England Biolabs, Ipswich, MA) y Fosfodiesterasa II (Worthington Biochemical Corp., Lakewood, NJ). Los dúplex de fragmentos-sonda se desnaturalizan a 95 °C durante dos minutos y las sondas se separan de los fragmentos (es decir, se eliminan) utilizando perlas magnéticas recubiertas con estreptavidina (DYNAL DYNAMAG-2, Invitrogen, Carlsbad, CA), y purifican con el Kit de purificación por PCR MINELUTE (Qiagen, Germantown, MD). Las sondas de poliinosina recortadas, aisladas y purificadas se miden para determinar la masa utilizando espectrometría de masas MALDI. La longitud de la sonda y, por lo tanto, la correspondiente longitud del fragmento, se extrapola a partir de los máximos de masa para cada especie de longitud de sonda en comparación con los máximos de masa para patrones de poli-inosina biotinilados de longitud conocida.

40

Determinación de trisomía 21

45 La cantidad relativa de cada especie de longitud de fragmento se determina en función de la amplitud de los máximos de masa para cada especie de longitud de sonda. Los fragmentos de 150 pares de bases o menos se cuantifican para el cromosoma 14 y el cromosoma 21. Las muestras con cantidades sustancialmente iguales de fragmentos del cromosoma 14 y el cromosoma 21 se determinan como euploides para el cromosoma 21. Las muestras con una cantidad estadísticamente significativamente mayor de fragmentos del cromosoma 21 en comparación con el cromosoma 14 (por ejemplo, una elevación del 2% en los fragmentos del cromosoma 21 en comparación con el cromosoma 14) se determinan como triploides para el cromosoma 21.

50

Ejemplo 5: Detección de trisomía mediante filtrado de longitud de fragmentos y representación cromosómica

55 En este ejemplo, las muestras maternas que contenían ácido nucleico libre de células se clasificaron como portadoras de un feto euploide o un feto que tenía una aneuploidía (es decir, trisomía 13, trisomía 18, trisomía 21) basándose en los recuentos de lecturas de secuencias de nucleótidos de un subconjunto de fragmentos que tienen ciertos parámetros de longitud. Las muestras se obtuvieron del Hospital Materno Infantil (estudio WI; Palomaki et al. (2011) Genet. Med. 13(11):913-20). Las lecturas de secuencias de nucleótidos (lecturas de 36 bases) para cada muestra se obtuvieron utilizando una plataforma de secuenciación de extremo pareado de Illumina (Illumina, Inc., San Diego, CA). Las lecturas de las secuencias de nucleótidos de extremo pareado se alinearon con un genoma de referencia (compilación 37 (hg19)) utilizando el programa del alineador BOWTIE 2 beta 3 y se determinó la longitud del fragmento basándose en las alineaciones de las lecturas de extremos pareados.

60

65 Ciertas lecturas de secuencias de nucleótidos se filtraron de acuerdo con los siguientes parámetros de longitud de fragmentos de ácidos nucleicos: 1) fragmentos que tienen longitudes mayores o iguales a 120 bases; 2) fragmentos que tienen longitudes mayores o iguales a 130 bases; 3) fragmentos que tienen longitudes mayores o iguales a 140 bases; 4) fragmentos que tienen longitudes mayores o iguales a 150 bases; 5) fragmentos que tienen longitudes

mayores o iguales a 160 bases; o 6) fragmentos que tienen longitudes mayores o iguales a 170 bases. Por lo tanto, se filtraron las lecturas de extremos pareados correspondientes a fragmentos iguales o más largos que un umbral de longitud dado (por ejemplo, 120 bases, 130 bases, 140 bases, 150 bases, 160 bases, 170 bases) y se conservaron las lecturas de extremos pareados correspondientes a fragmentos más cortos que un umbral de longitud dado para el análisis.

Las representaciones cromosómicas para el cromosoma 13, el cromosoma 18 y el cromosoma 21 se calcularon para los conjuntos de datos presentados en la FIG. 23 utilizando 1) lecturas de secuencias sin filtrar y 2) lecturas de secuencias filtradas por longitud en un umbral de fragmentos de 150 bases. La representación cromosómica para cada uno de los cromosomas 13, 18 y 21 se calculó de acuerdo con lo siguiente:

Representación del cromosoma 13 (Cr 13) = Σ de recuentos de lecturas de secuencias del Cr 13 (sin filtrar) / Σ de todos los recuentos de lecturas de secuencias autosómicas (sin filtrar)

Representación del cromosoma 13 (Cr 13) = Σ de recuentos de lecturas de secuencias del Cr 13 (filtradas) / Σ de todos los recuentos de lecturas de secuencias autosómicas (filtradas)

Representación del cromosoma 18 (Cr 18) = Σ de recuentos de lecturas de secuencias del Cr 18 (sin filtrar) / Σ de todos los recuentos de lecturas de secuencias autosómicas (sin filtrar)

Representación del cromosoma 18 (Cr 18) = Σ de recuentos de lecturas de secuencias del Cr 18 (filtradas) / Σ de todos los recuentos de lecturas de secuencias autosómicas (filtradas)

Representación del cromosoma 21 (Cr 21) = Σ de recuentos de lecturas de secuencias del Cr 21 (sin filtrar) / Σ de todos los recuentos de lecturas de secuencias autosómicas (sin filtrar)

Representación del cromosoma 21 (Cr 21) = Σ de recuentos de lecturas de secuencias del Cr 21 (filtradas) / Σ de todos los recuentos de lecturas de secuencias autosómicas (filtradas)

Las figuras 14, 16 y 18 muestran representaciones cromosómicas para los cromosomas 13, 18 y 21, respectivamente, utilizando lecturas de secuencias sin filtrar. Las figuras 15, 17 y 19 muestran representaciones cromosómicas para los cromosomas 13, 18 y 21, respectivamente, utilizando lecturas de secuencias filtradas por longitud. Para los conjuntos de datos filtrados, la representación cromosómica aumentó para las muestras de trisomía debido en parte a un aumento en los datos de secuencia fetal aportada. Aunque este aumento en la representación cromosómica puede aumentar el poder para detectar anomalías cromosómicas, la varianza de la representación cromosómica para muestras sin trisomía aumentó debido a una reducción aproximada del 63-82% en los recuentos de lecturas. Las distribuciones de ejemplo de recuentos de lecturas en varios valores umbral de longitud de fragmento se ilustran en la FIG. 13 y se presenta en la Tabla 5 a continuación.

TABLA 5	
Umbral (longitudes de fragmento)	AUC media (% de lecturas menores que el umbral)
120	0,027
130	0,049
140	0,092
150	0,175
160	0,294
170	0,508
Todas	1

Se determinaron los valores del área media bajo la curva (AUC) para lecturas de fragmentos de menos de una cierta longitud para ilustrar la reducción general de las lecturas (es decir, la cobertura de secuencia) observada en promedio. Para un ensayo dado que genera aproximadamente 15 millones de lecturas de secuencias (o cobertura de 0,2X del genoma humano), la exclusión de lecturas de más de 150 bases, por ejemplo, es equivalente a una cobertura de aproximadamente 0,035X.

Para determinar un umbral de tamaño de fragmento óptimo para la representación cromosómica, el umbral de tamaño de fragmento se varió de 120 a 170 bases, en incrementos de 10 bases. La representación cromosómica (es decir, para los cromosomas 13, 18 y 21) se calculó después de la normalización de los recuentos de lecturas de secuencias (es decir, ajuste PERUN con LOESS) para cada conjunto de datos de longitud filtrado (lecturas de extremos pareados) y para un conjunto de datos sin filtrar (lecturas de un solo extremo; también denominadas "todas"). Las representaciones de los cromosomas 13, 18 y 21 se presentan en las Figs. 20, 21 y 22, respectivamente. La representación cromosómica de los conjuntos de datos filtrados en los umbrales de 150, 160 y 170 bases fue bastante consistente con el conjunto de datos sin filtrar. Las siguientes tablas presentan la especificidad y sensibilidad

5 observadas para la detección de trisomía de los cromosomas 13, 18 y 21 en los valores de corte de puntuación Z respectivos (es decir, 3,95 para el cromosoma 13, 3,95 para el cromosoma 18 y 3 para el cromosoma 21). Los valores de puntuación Z se basaron en la media específica de la celda de flujo y los valores de MAD del histórico específico de los conjuntos de datos y de la población. Adicionalmente, se llevó a cabo una validación cruzada de 10 veces de los análisis de característica operativa del receptor (ROC) (es decir, una validación cruzada estratificada 10 veces, repetida 100 veces) y el área promedio bajo la curva (AUC; es decir, una medida de precisión) para cada análisis (calculada resumiendo todos los valores de los tiempos de sensibilidad (1-especificidad) y se implementó usando el paquete R ROCR) que se presenta en las Tablas 6, 7 y 8 a continuación.

10

TABLA 6			
Umbral (longitudes de fragmento)	CROMOSOMA 13 (z=3,95)		
	AUC ROC	Especificidad	Sensibilidad
120	0,85	1,00	0,00
130	0,99	1,00	0,67
140	1,00	1,00	1,00
150	1,00	0,99	1,00
160	1,00	0,99	1,00
170	1,00	0,99	1,00
Todas	1,00	0,99	1,00

TABLA 7			
Umbral (longitudes de fragmento)	CROMOSOMA 18 (z=3,95)		
	AUC ROC	Especificidad	Sensibilidad
120	0,77	1,00	0,04
130	0,98	1,00	0,26
140	1,00	1,00	0,91
150	1,00	1,00	1,00
160	1,00	1,00	1,00
170	1,00	1,00	1,00
Todas	1,00	1,00	0,91

TABLA 8			
Umbral (longitudes de fragmento)	CROMOSOMA 21 (z=3)		
	AUC ROC	Especificidad	Sensibilidad
120	0,83	1,00	0,08
130	0,88	1,00	0,44
140	0,97	1,00	0,88
150	1,00	1,00	0,92
160	1,00	1,00	1,00
170	1,00	1,00	0,96
Todas	1,00	1,00	1,00

15 Los datos muestran que, a pesar de una reducción significativa en la cobertura de secuencia para las muestras filtradas por longitud, las trisomías se pueden identificar utilizando muestras filtradas con ciertos umbrales de longitud de fragmentos (por ejemplo, 150 bases, 160 bases) con una precisión, sensibilidad y especificidad similares en comparación con las muestras sin filtrar.

REIVINDICACIONES

1. Un método para determinar la presencia o la ausencia de una aneuploidía cromosómica, que comprende:
 - 5 (a) obtener recuentos de lecturas de secuencias de nucleótidos mapeadas en secciones genómicas de un genoma de referencia, cuyas lecturas de secuencias son:
 - (i) lecturas de ácido nucleico circulante libre de células de una muestra de ensayo de una gestante, y
 - (ii) lecturas de fragmentos de ácido nucleico que son más cortos que 160 bases, en donde la obtención de recuentos comprende eliminar lecturas de secuencias de nucleótidos que son de fragmentos que tienen una longitud igual a o mayor que 160 bases;
 - 10 (b) normalizar los recuentos, generando así recuentos normalizados de lecturas de secuencias mapeadas en las secciones genómicas; y
 - (c) determinar la presencia o la ausencia de una aneuploidía cromosómica de acuerdo con los recuentos normalizados.
- 15 2. El método de la reivindicación 1, en donde la normalización en (b) comprende:
 - (I) determinar un coeficiente de sesgo de guanina y citosina (GC) para la muestra de ensayo basándose en una relación ajustada entre (i) los recuentos de las lecturas de secuencias mapeadas para cada una de las secciones genómicas en (a) y (ii) el contenido de GC para cada una de las secciones genómicas;
 - 20 (II) calcular una elevación de sección genómica para cada una de las secciones genómicas basándose en los recuentos de (a), del coeficiente de sesgo de GC de (I) y de una relación ajustada, para cada una de las secciones genómicas, entre (i) un coeficiente de sesgo de GC para cada una de las múltiples muestras de múltiples gestantes y
 - (ii) los recuentos de las lecturas de secuencias mapeadas en cada una de las secciones genómicas para las múltiples muestras, proporcionando así elevaciones de secciones genómicas calculadas, por lo que el sesgo en los recuentos de las lecturas de secuencias mapeadas en cada una de las secciones genómicas se reduce en las elevaciones de secciones genómicas calculadas.
3. El método de la reivindicación 1 o 2, en donde el método comprende comparar los recuentos normalizados, o derivados de los mismos, con un valor de umbral, proporcionando así una comparación.
- 30 4. El método de la reivindicación 3, en donde la determinación de la presencia o ausencia de una aneuploidía cromosómica es según la comparación.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en donde los recuentos de las lecturas de secuencias de nucleótidos normalizadas en (b) se reducen en relación con los recuentos de las lecturas de secuencias de nucleótidos no restringidos por la longitud del fragmento.
- 35 6. El método de la reivindicación 5, en donde los recuentos de las lecturas de secuencias de nucleótidos normalizadas en (b) se reducen en al menos aproximadamente 50 %, preferentemente en aproximadamente de 63 % a aproximadamente 82 %.
7. El método de una cualquiera de las reivindicaciones 1 a 6, que comprende determinar la longitud de un fragmento de ácido nucleico.
- 40 8. El método de la reivindicación 7, en donde la determinación de la longitud de un fragmento de ácido nucleico comprende un método de secuenciación de extremo pareado.
9. El método de una cualquiera de las reivindicaciones 1 a 8, en donde la presencia o la ausencia de una aneuploidía cromosómica se determina con una sensibilidad de al menos aproximadamente 0,92 y una especificidad de al menos aproximadamente 0,99.
- 45 10. El método de una cualquiera de las reivindicaciones 1 a 9, en donde la presencia o la ausencia de una aneuploidía cromosómica se determina con una sensibilidad de aproximadamente 1,00 y una especificidad de al menos aproximadamente 0,99.
11. El método de una cualquiera de las reivindicaciones 1 a 10, en donde la muestra es de sangre, suero o plasma.
12. El método de una cualquiera de las reivindicaciones 1 a 11, en donde la aneuploidía cromosómica es una trisomía.
- 50 13. El método de la reivindicación 12, en donde la trisomía es trisomía 21, trisomía 18 o trisomía 13.
14. El método de una cualquiera de las reivindicaciones 1 a 13, en donde las lecturas de las secuencias de nucleótidos que son de fragmentos más largos de 160 bases, se eliminan *in silico*.

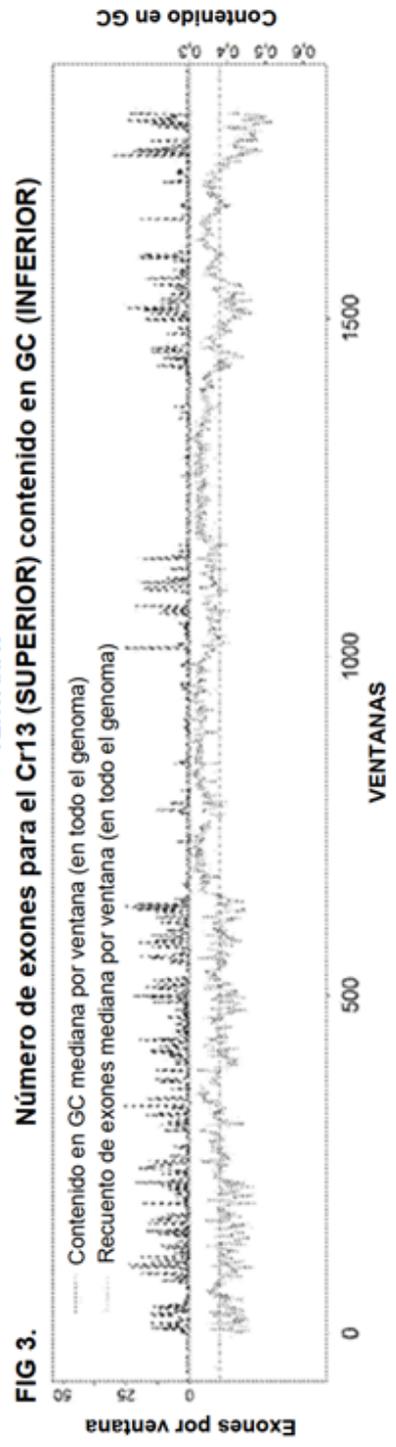
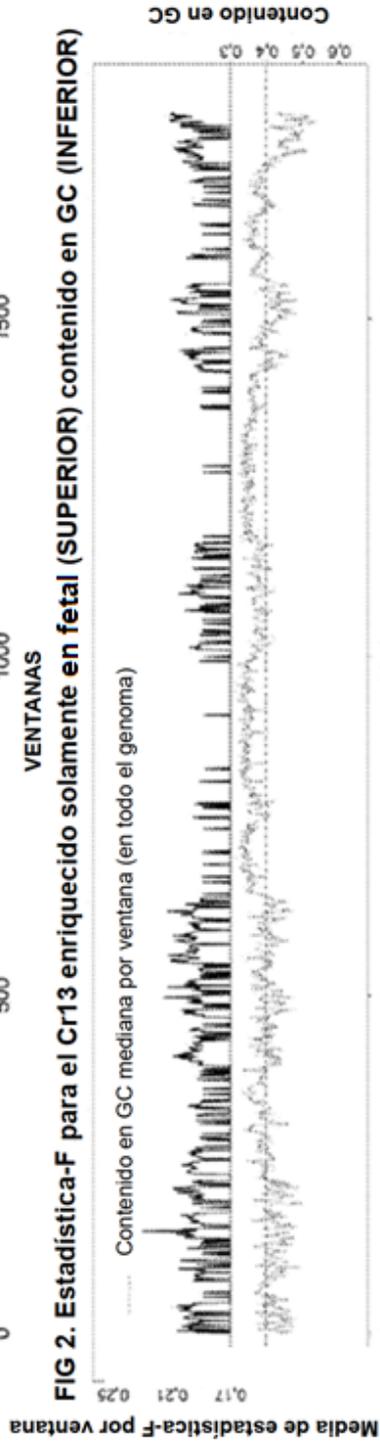
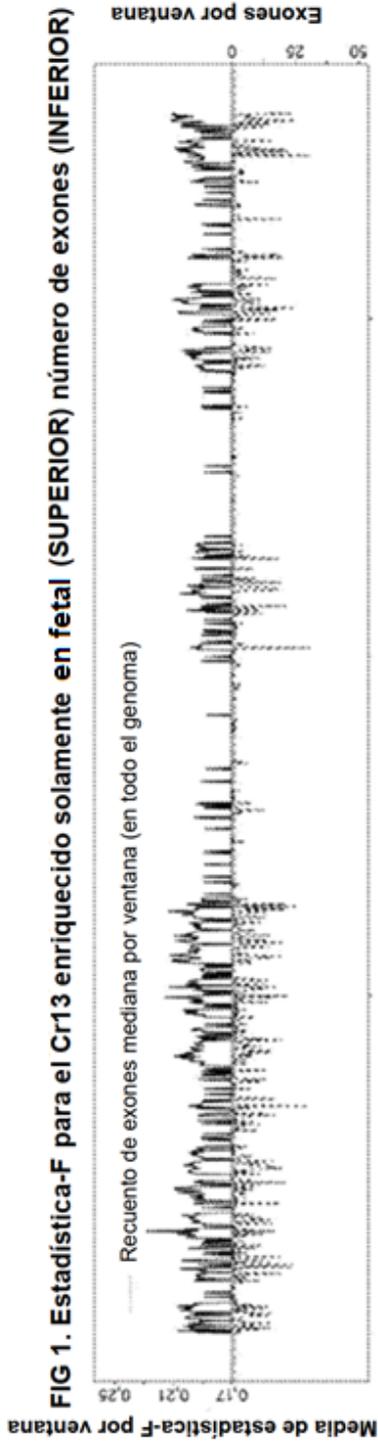


FIG 4. Estadística-F para el Cr18 enriquecido solamente en fetal (SUPERIOR) número de exones (INFERIOR)

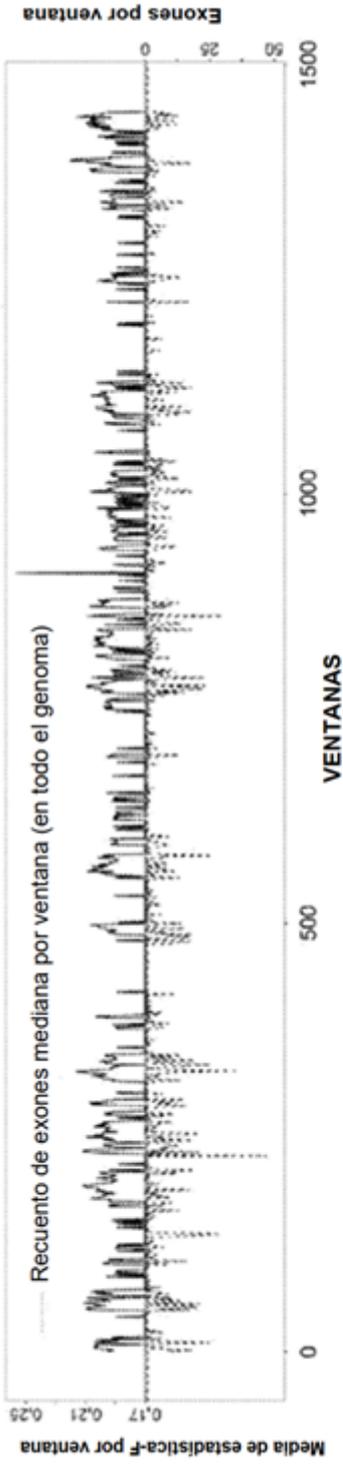


FIG 5. Estadística-F para el Cr18 enriquecido solamente en fetal (SUPERIOR) contenido en GC (INFERIOR)

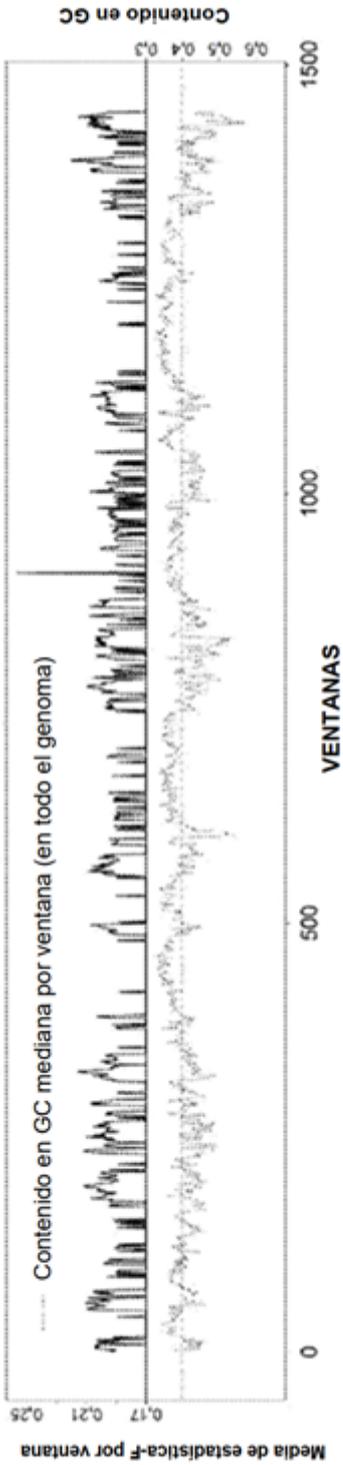


FIG 6. Número de exones para el Cr18 (SUPERIOR) contenido en GC (INFERIOR)

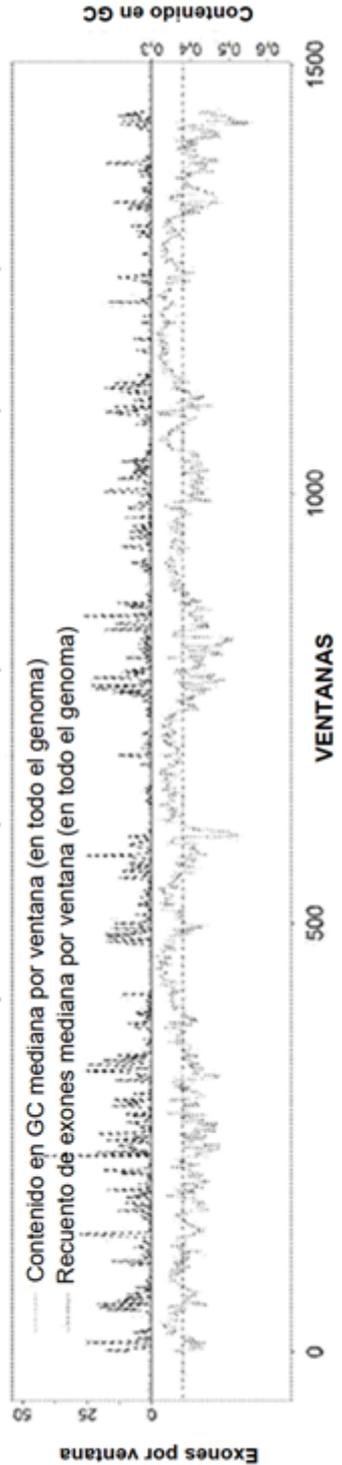


FIG 7. Estadística-F para el Cr21 enriquecido solamente en fetal (SUPERIOR) número de exones (INFERIOR)

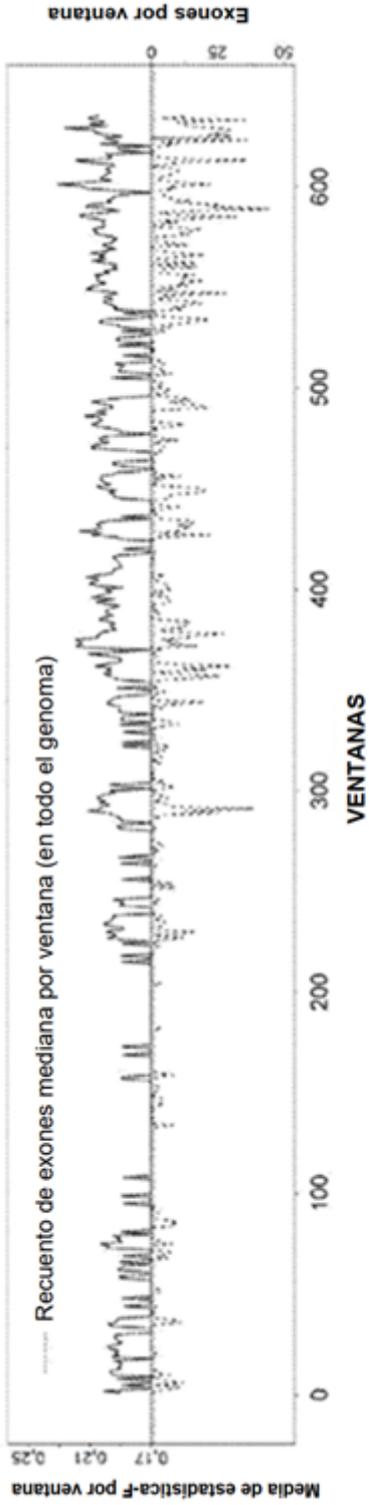


FIG 8. Estadística-F para el Cr21 enriquecido solamente en fetal (SUPERIOR) contenido en GC (INFERIOR)

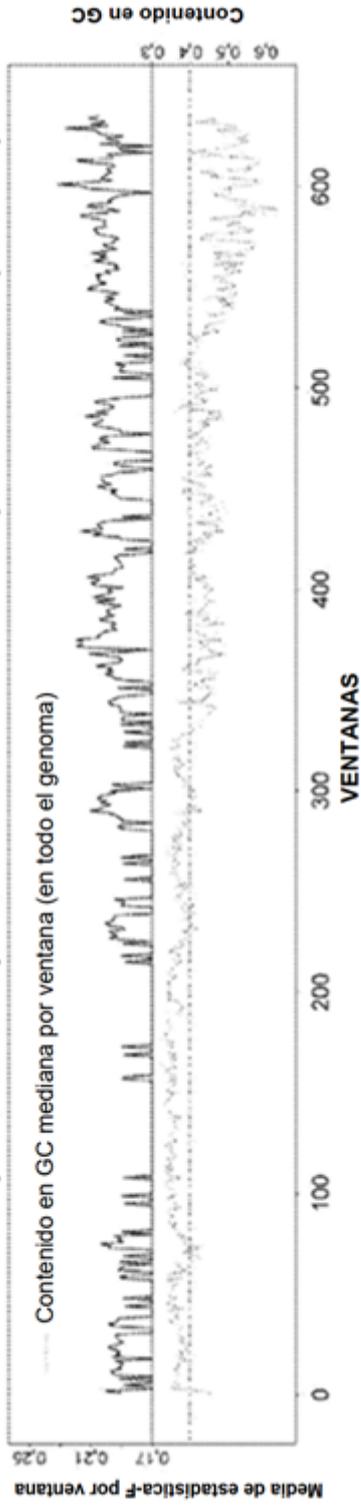


FIG 9. Número de exones para el Cr21 (SUPERIOR) contenido en GC (INFERIOR)

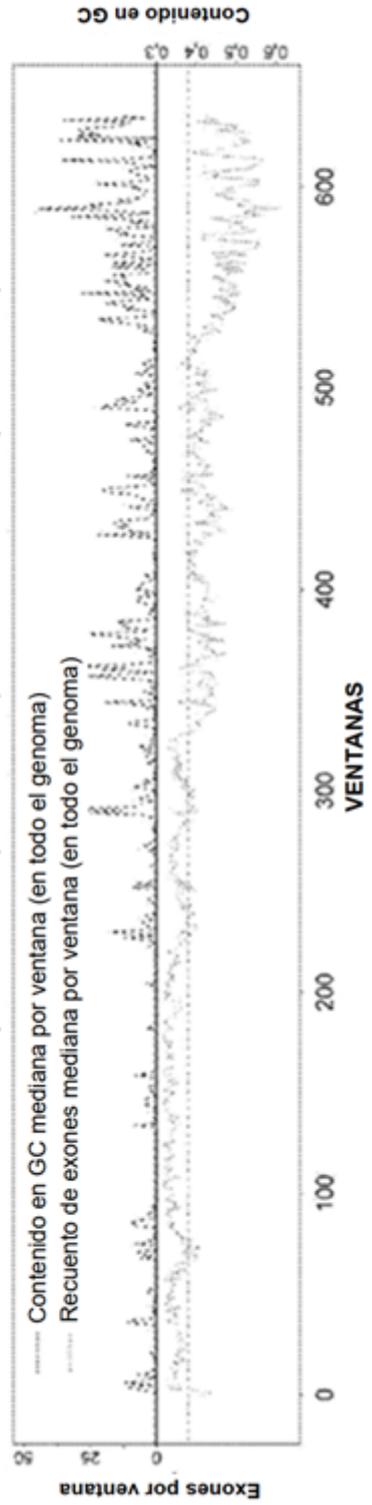


FIG. 10

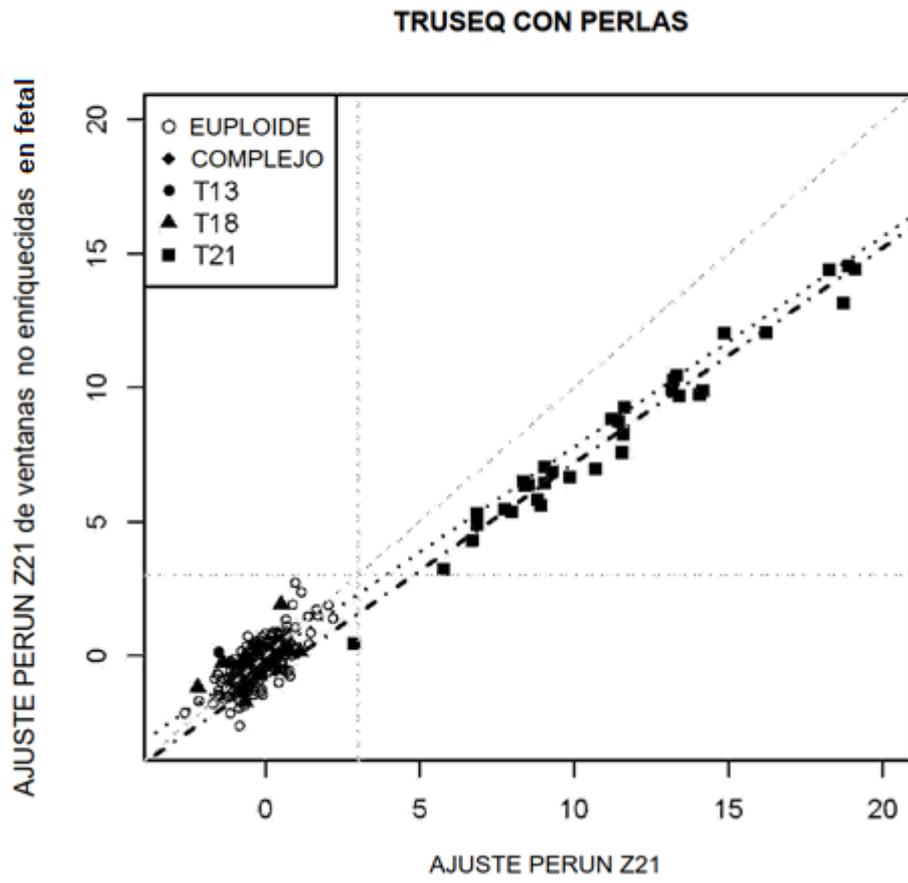


FIG. 11

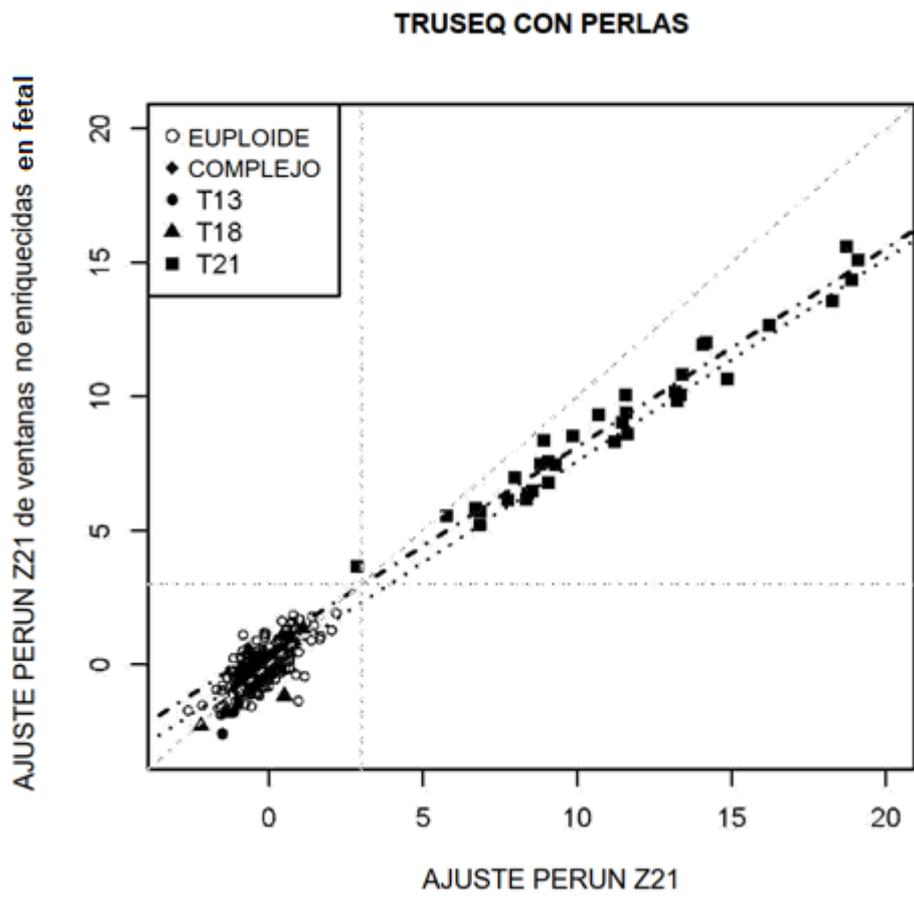


FIG. 12

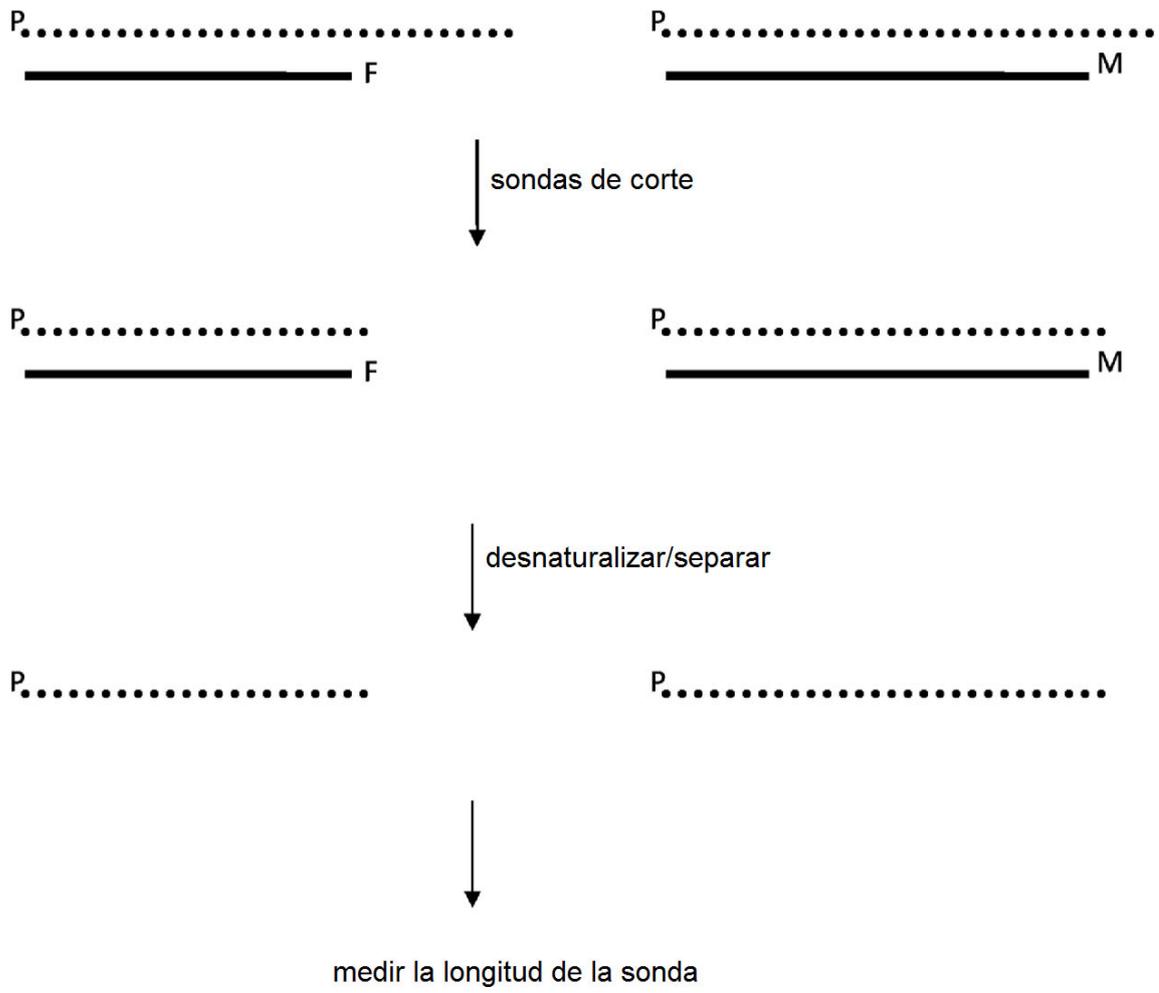


FIG. 13

Recuentos por longitud / recuentos totales

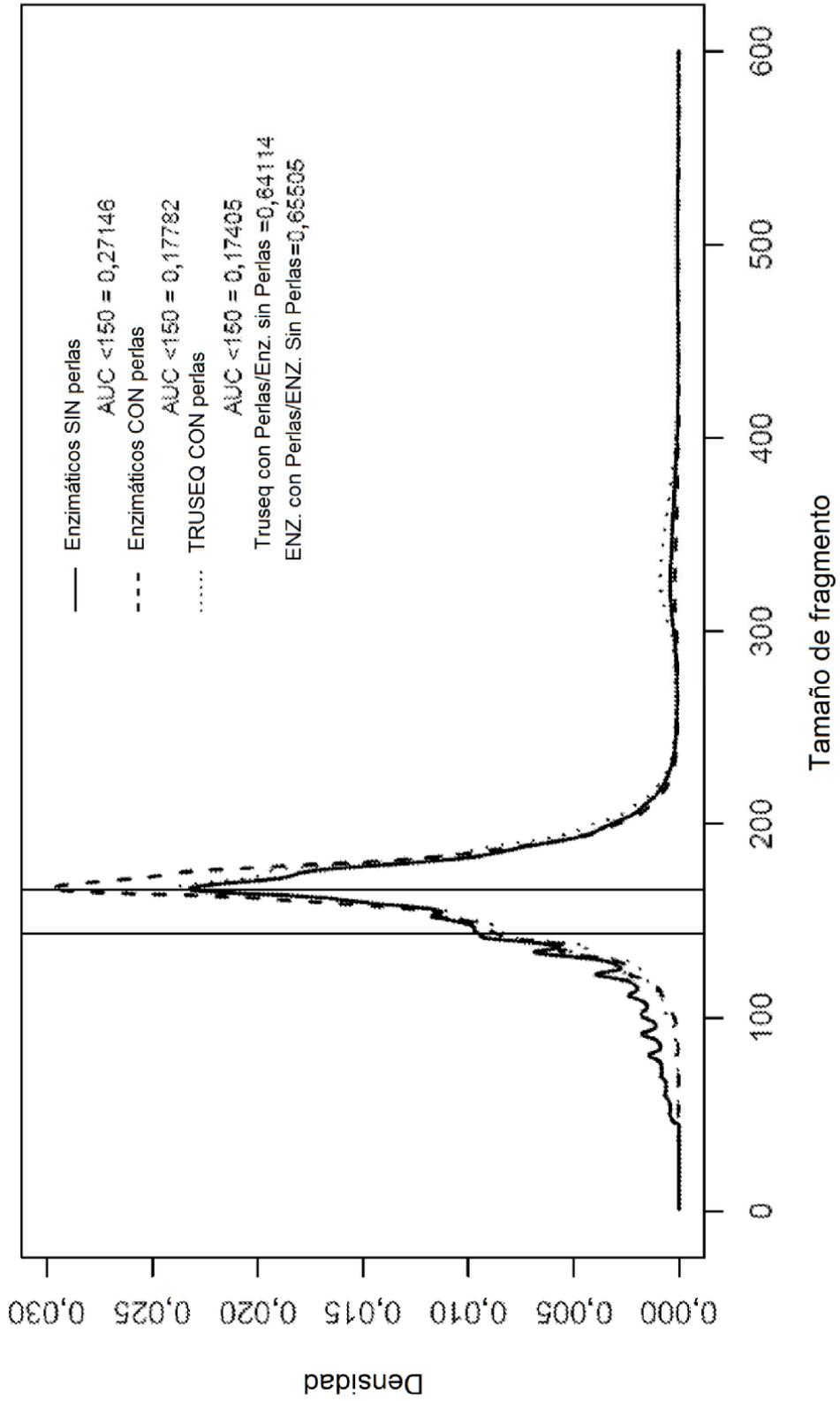


FIG. 14

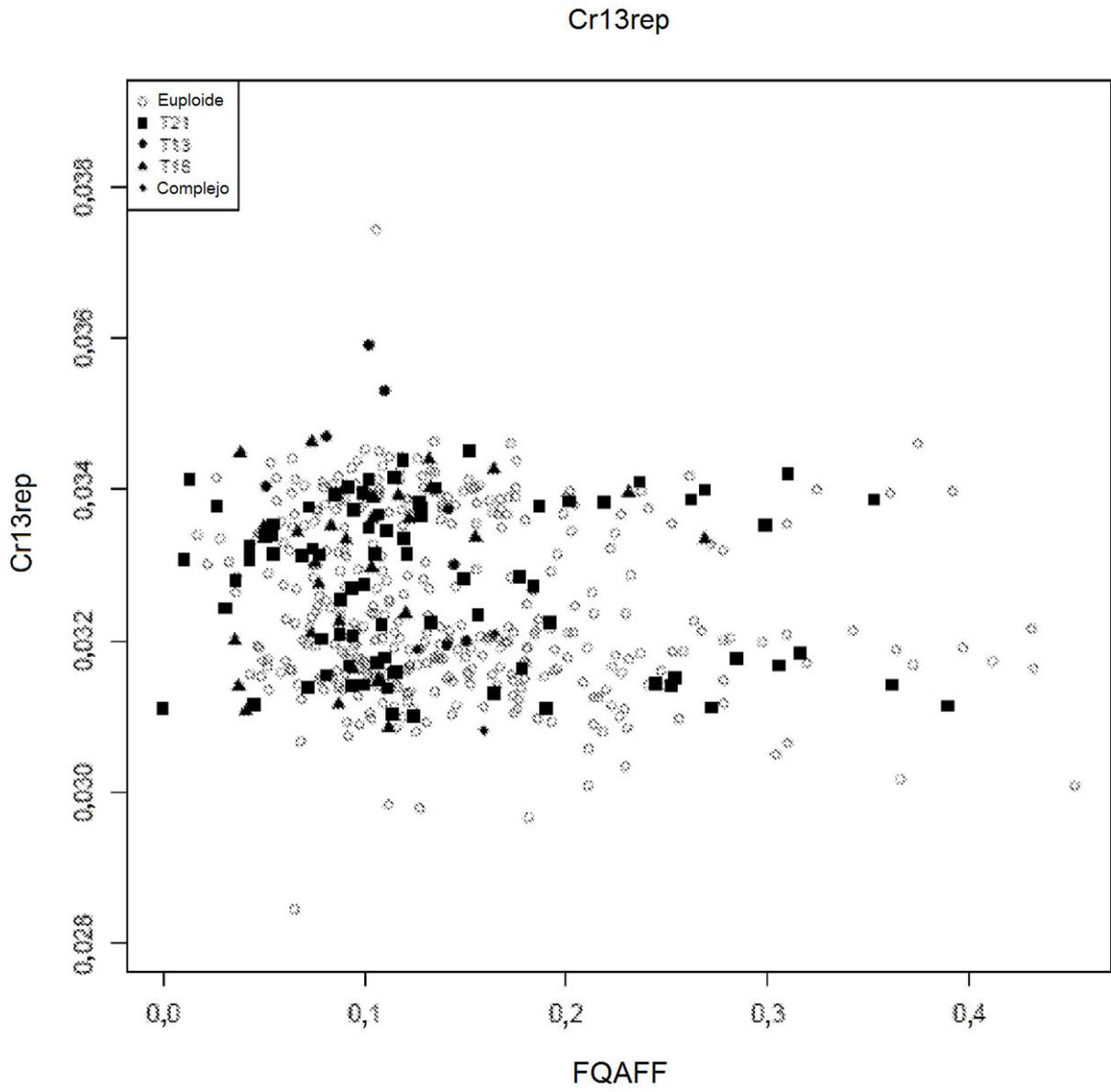


FIG. 15

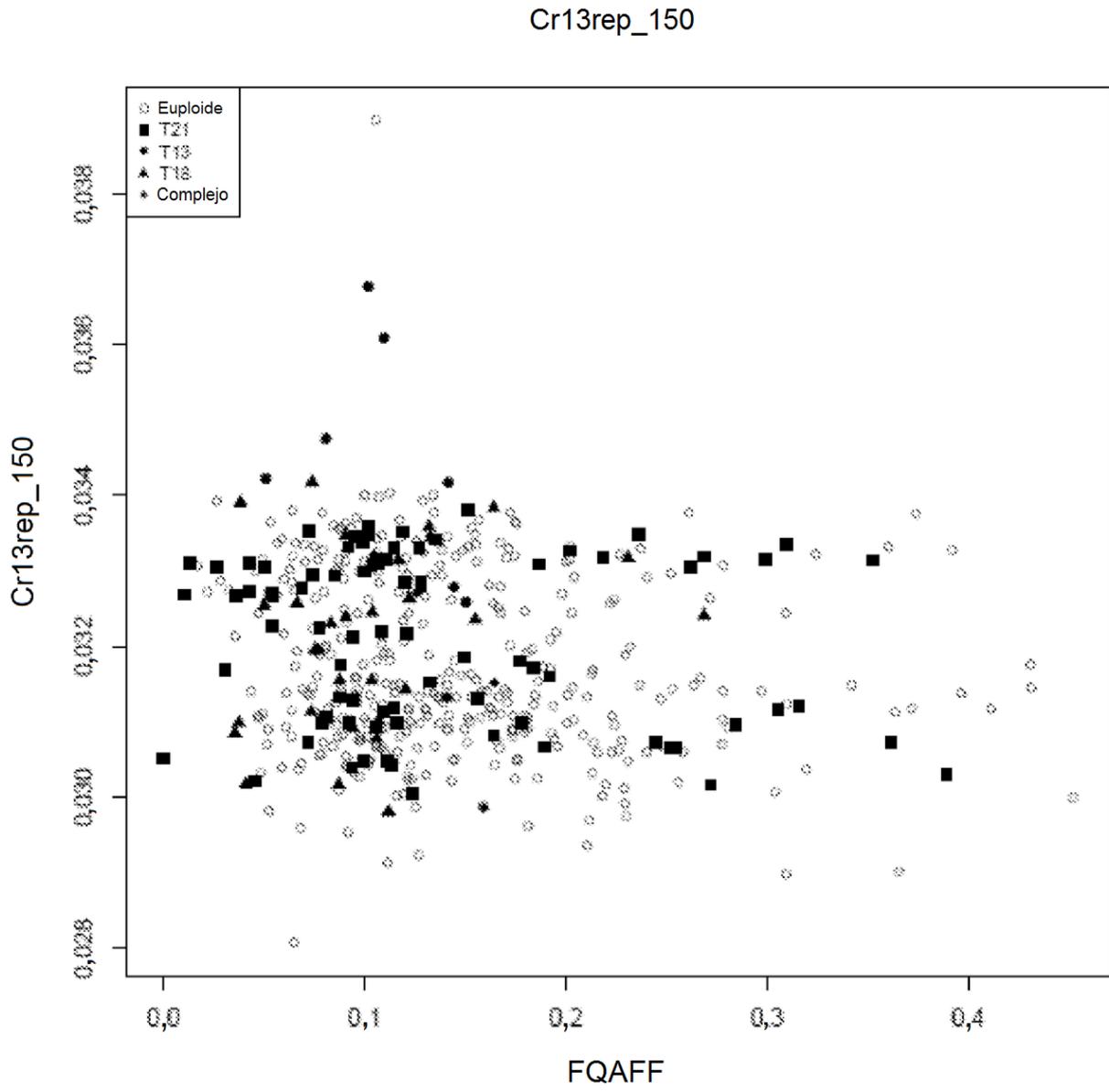


FIG. 16

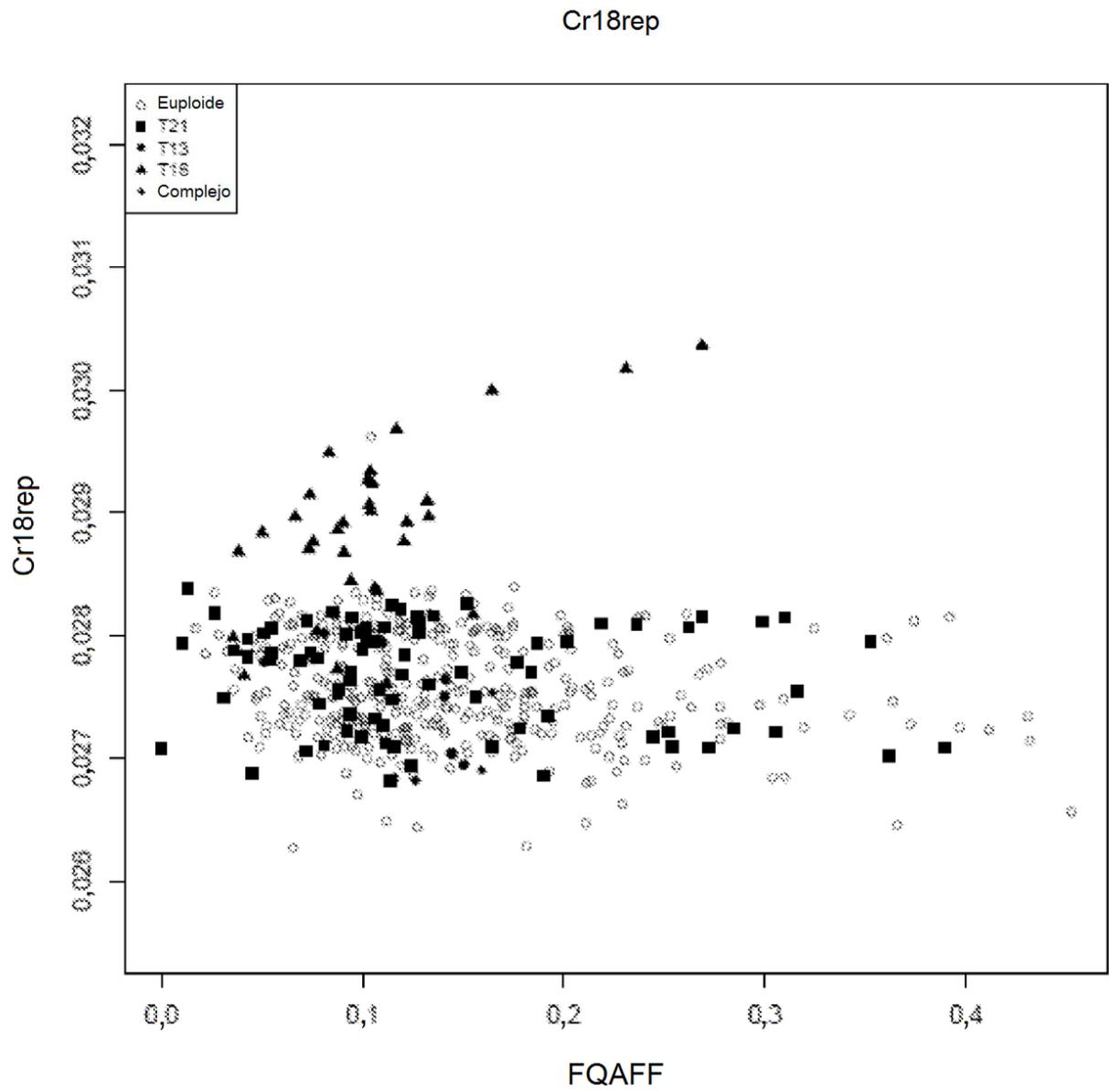


FIG. 17

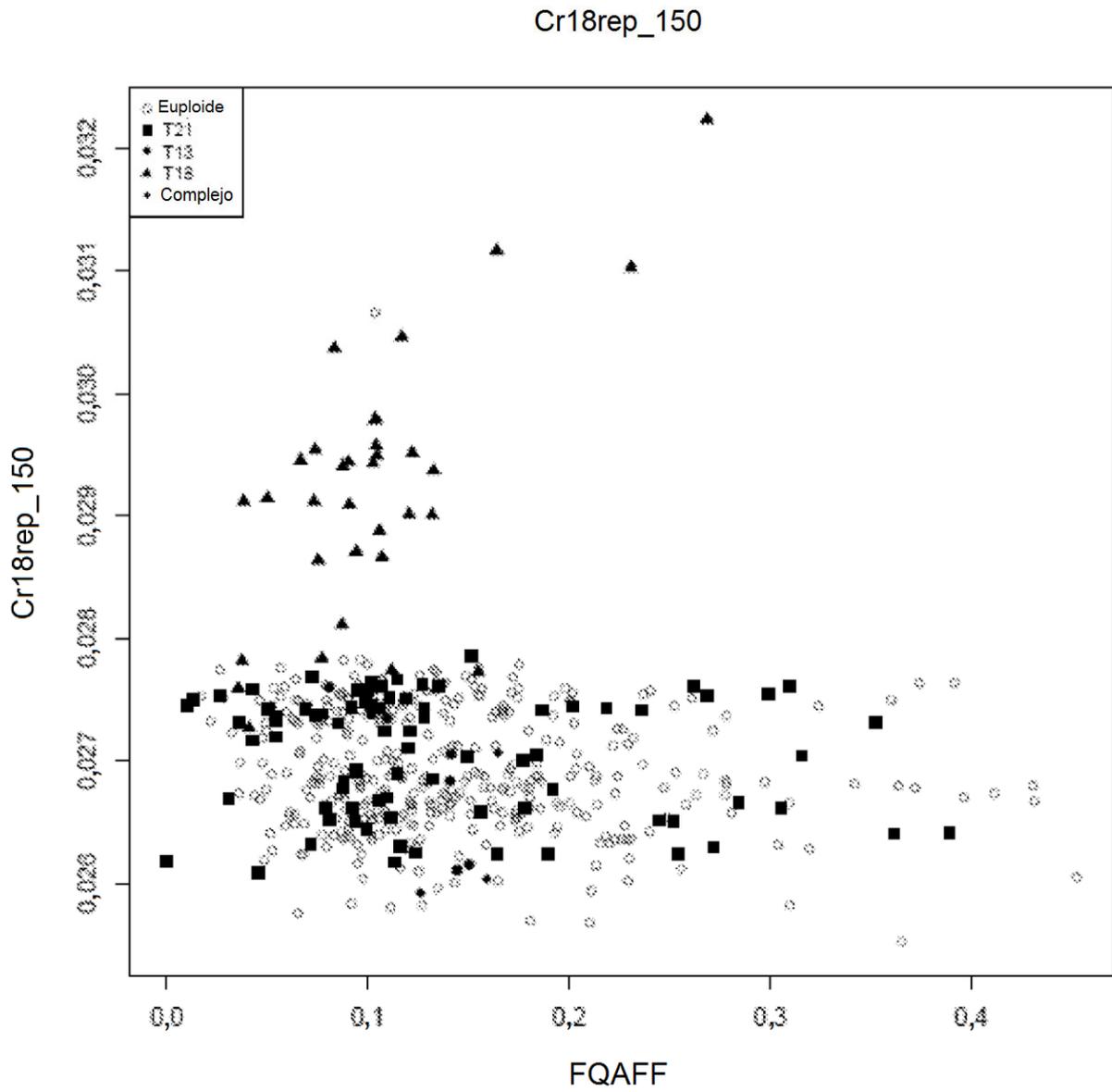


FIG. 18

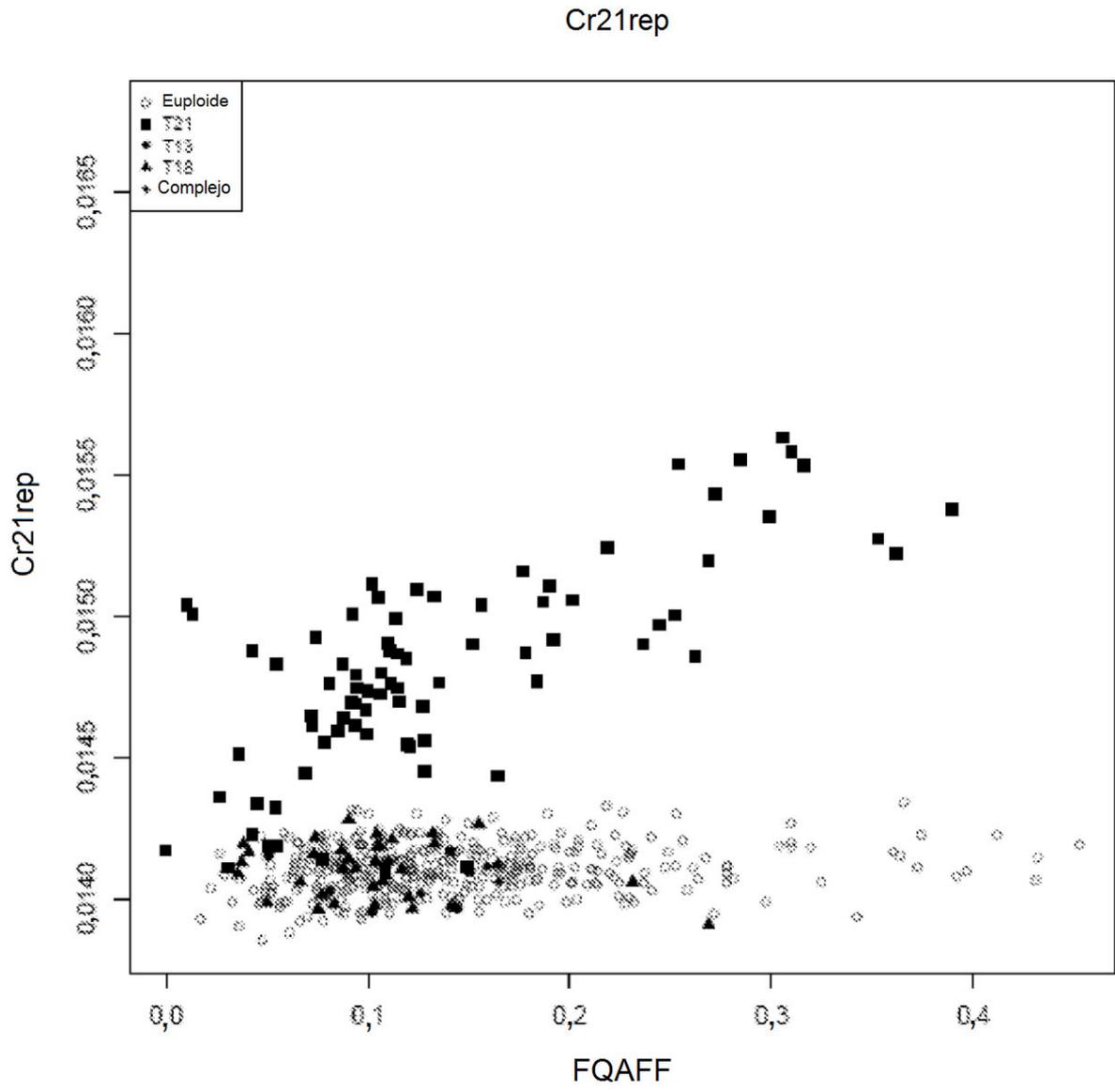


FIG. 19

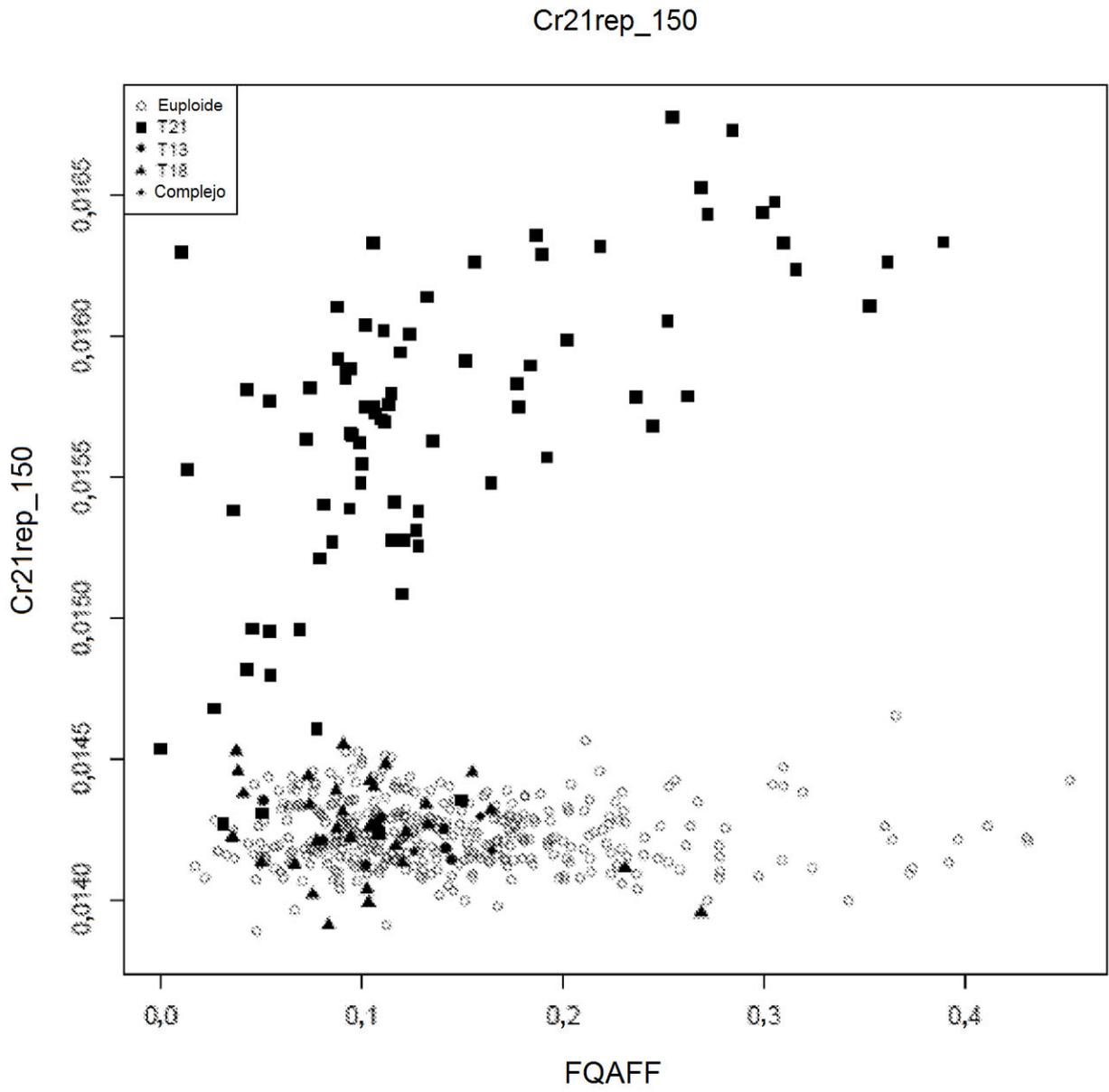


FIG. 20

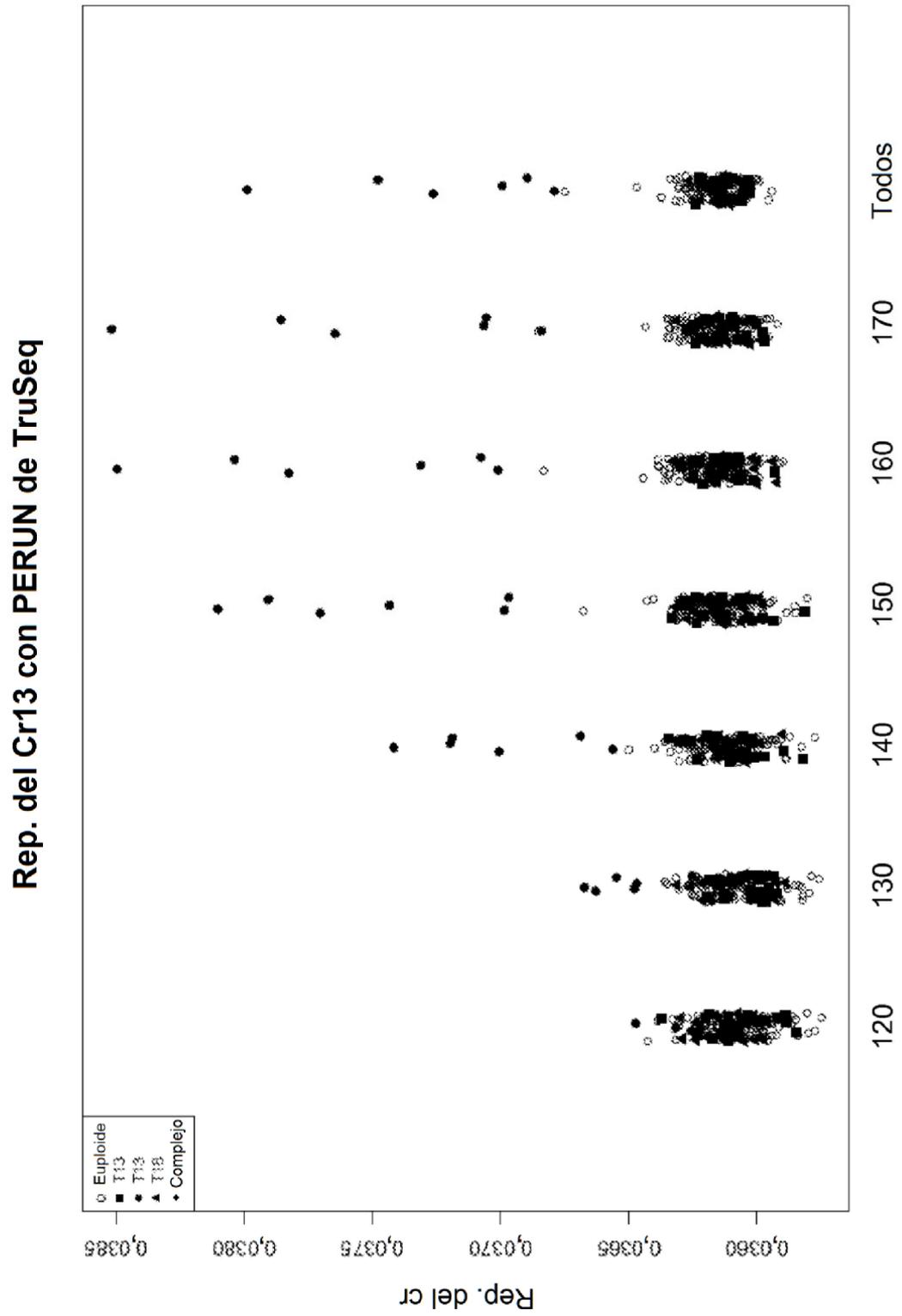


FIG. 21

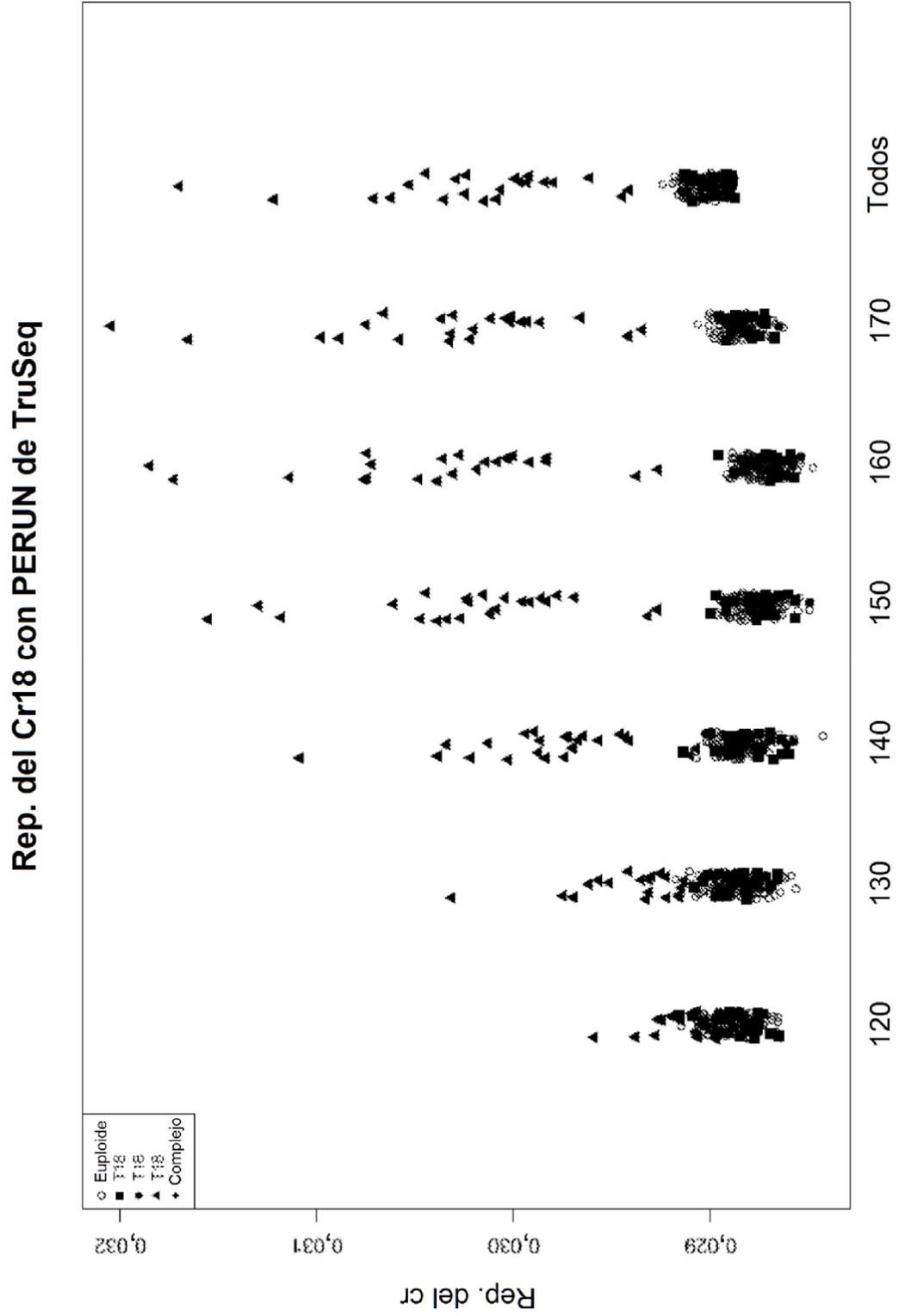


FIG. 22

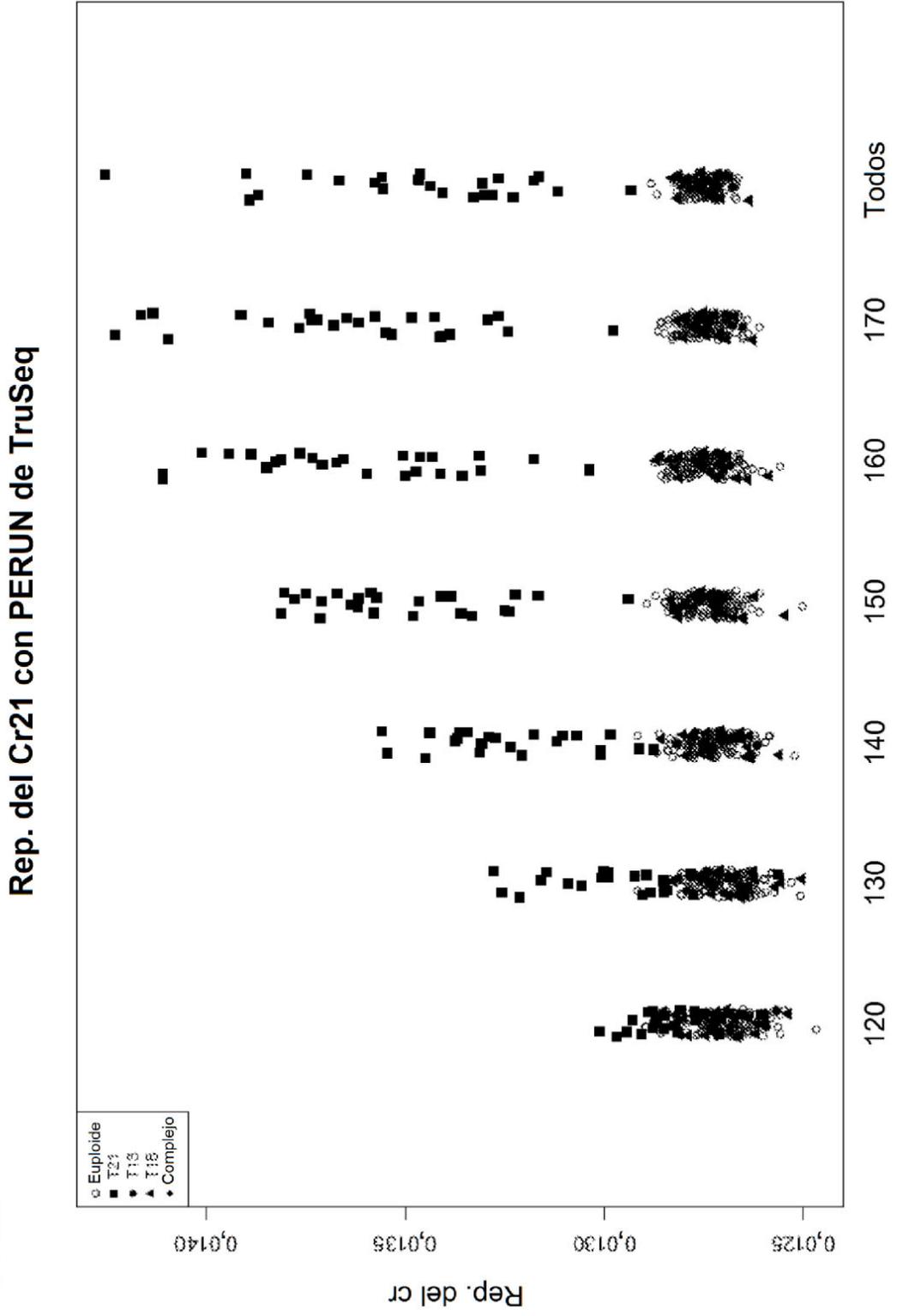
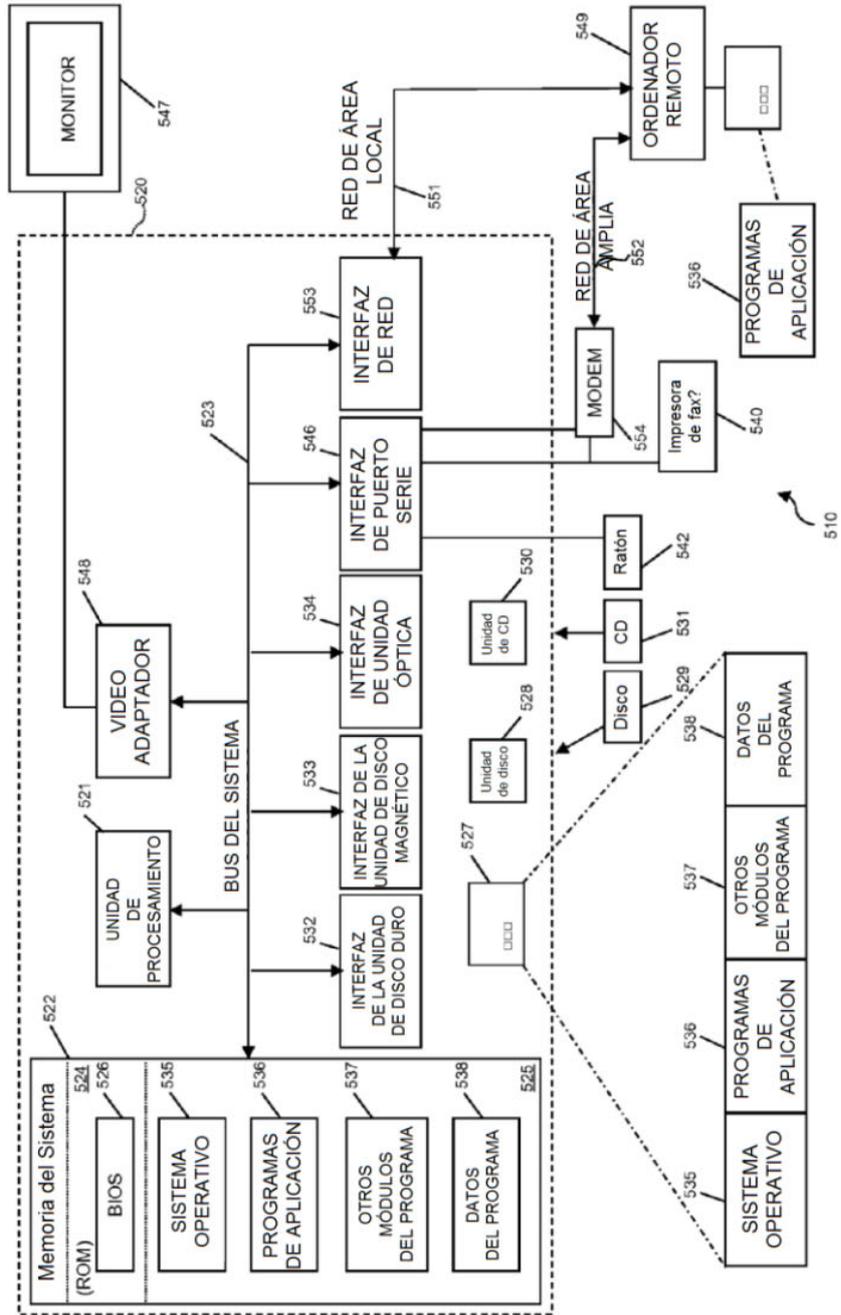


FIG. 23

Celda de Flujo	Muestras	Longitud de lectura	Bioquímica	Limpieza automatizada de perlas	Estudio	Placa	FQA	NTC	Agrupados
RDFC0050	32	36	ENZ	SI	480v2		32/32	0	0
RDFC0051	32	36	ENZ	SI	480v2		32/32	0	0
RDFC010044	96	45	ENZ	NO	WI	19	88/96	1	1
RDFC010045	96	45	TRUSEQ	SI	WI	EPT	0/96	0	0
RDFC010049	96	36	ENZ	NO	WI	17	84/96	1	1
RDFC010050	96	36	ENZ	NO	WI	18	90/96	1	1
RDFC010232	96	36	TRUSEQ	SI	CEWI	PE_P2	88/96	0	8
RDFC010239	96	36	TRUSEQ	SI	CEWI	PE_P1	88/96	0	8

FIG. 24



REFERENCIAS CITADAS EN LA DESCRIPCIÓN

La lista de referencias citadas por el solicitante es, únicamente, para conveniencia del lector. No forma parte del documento de patente europea. Si bien se ha tenido gran cuidado al compilar las referencias, no pueden excluirse errores u omisiones y la OEP declina toda responsabilidad a este respecto.

5

Documentos de patente citados en la descripción

- WO 2013055817 A [0004]
- WO 2011091063 A [0004]
- WO 2010065470 A [0004]
- WO 2012118745 A [0004]
- US 0769991 W [0025]
- US 2007071232 W [0025]
- US 968876 [0025]
- US 60968878 B [0025]
- EP 05012707 W [0025]
- US 20050112590 [0035]
- US 20100105049 [0047] [0048] [0056]
- US 20070065823 [0048]
- US 20040081993 [0049]
- US 20110224087 [0050]
- US 6927028 B [0054]
- WO 2007140417 A [0054] [0079]
- WO 2007147063 A [0054]
- WO 2009032779 A [0054]
- WO 2009032781 A [0054]
- WO 2010033639 A [0054]
- WO 2011034631 A [0054]
- WO 2006056480 A [0054]
- WO 2011143659 A [0054]
- US 20090317818 [0057]
- WO 2010115016 A [0079]
- US 5720928 A [0106]
- US 20130012399 A [0131] [0155]
- US 20090029377 A [0132]
- US 669136 [0204]
- US 1259123 W [0204] [0207] [0220] [0234] [0366]
- WO 2013052913 A [0204] [0207] [0220] [0234] [0366]
- US 20130085681 A [0207] [0220]

10

Literatura no patente citada en la descripción

- **T. STRACHAN.** The Human Genome. BIOS Scientific Publishers, 1992 [0002]
- Duchenne Muscular Dystrophy (DMD), Huntington's Disease (HD), Alzheimer's Disease and Cystic Fibrosis (CF). **D. N. COOPER ; M. KRAWCZAK.** Human Genome Mutations. BIOS Publishers, 1993 [0003]
- **KITZMAN JO et al.** *Sci Transl Med.*, 06 June 2012, vol. 4 (137), 137ra76 [0004]
- **TABOR HK et al.** *Am J Med Genet A.*, October 2012, vol. 158A (10), 2382-4 [0004]
- **SAMBROOK ; RUSSELL.** Molecular Cloning: A Laboratory Manual 3d ed. 2001 [0023]
- **SAMBROOK ; RUSSELL.** *Molecular Cloning: A Laboratory Manual 3d ed.*, 2001 [0027] [0100]
- Current Protocols in Molecular Biology. John Wiley & Sons, 1989, 6.3.1-6.3.6 [0028] [0071]
- **CHAN et al.** *Clin. Chem.*, 2004, vol. 50, 88-92 [0045]
- **LO et al.** *Sci. Transl. Med.*, 2010, vol. 2, 61-ra91 [0045]
- **LO.** *Journal of Histochemistry and Cytochemistry*, 2005, vol. 53 (3), 293-296 [0046]
- **GEBHARD et al.** *Cancer Res.*, 2006, vol. 66 (12), 6118-28 [0048]
- **DING ; CANTOR.** *PNAS USA*, 2003, vol. 100, 3059-3064 [0049]
- **BEAUCAGE ; CARUTHERS.** *Tetrahedron Letts*, 1981, vol. 22, 1859-1862 [0068]
- **LIN ; BROWN.** *Nucleic Acids Res.*, 1989, vol. 17, 10383 [0116]
- **BROWN ; LIN.** *Carbohydrate Research*, 1991, vol. 216, 129-139 [0116]
- **NEEDHAM-VANDEVANTER et al.** *Nucleic Acids Res.*, 1984, vol. 12, 6159-6168 [0068]
- **PEARSON ; REGNIER.** *J. Chrom.*, 1983, vol. 255, 137-149 [0068]
- **JURINKE et al.** *Mol. Biotechnol.*, 2004, vol. 26, 147-164 [0082]
- **BURLINGAME et al.** *Anal. Chem.*, 1998, vol. 70, 647R-716R [0084]
- **VERBECK et al.** *Journal of Biomolecular Techniques*, vol. 13 (2), 56-61 [0088]
- **DING ; CANTOR.** *PNAS USA.*, 18 March 2003, vol. 100 (6), 3059-64 [0098]
- **LAI et al.** *Nat Genet*, 1999, vol. 23 (3), 309-13 [0106]
- **ASTON et al.** *Trends Biotechnol.*, 1999, vol. 17 (7), 297-302 [0106]
- **ASTON et al.** *Methods Enzymol.*, 1999, vol. 303, 55-73 [0106]
- **JING et al.** *Proc Natl Acad Sci USA.*, 1998, vol. 95 (14), 8046-51 [0106]
- **NICHOLS et al.** *Nature*, 1994, vol. 369, 492-493 [0116]
- **BERGSTROM et al.** *J. Am. Chem. Soc.*, 1995, vol. 117, 1201-1209 [0116]
- **LOAKES ; BROWN.** *Nucleic Acids Res.*, 1994, vol. 22, 4039-4043 [0116]
- **LIN ; BROWN.** *Nucleic Acids Res.*, 1992, vol. 20, 5149-5152 [0116]
- **BERGER et al.** *Nucleic Acids Res.*, 2000, vol. 28 (15), 2911-2914 [0116]
- **PALOMAKI et al.** *Genet. Med.*, 2011, vol. 13 (11), 913-20 [0502]