(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 939 547**

(51) Int. CI.:

G16B 20/00 (2009.01)
G16B 30/00 (2009.01)
G16B 40/00 (2009.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **02.04.2014 PCT/US2014/032687**

(87) Fecha y número de publicación internacional: **09.10.2014 WO14165596**

(96) Fecha de presentación y número de la solicitud europea: **02.04.2014 E 14723612 (9)**

(97) Fecha y número de publicación de la concesión europea: **18.01.2023 EP 2981921**

(54) Título: **Métodos y procedimientos para la evaluación no invasiva de variaciones genéticas**

(30) Prioridad:

03.04.2013 US 201361808027 P
24.05.2013 US 201361827323 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
24.04.2023

(73) Titular/es:

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

(72) Inventor/es:

DZAKULA, ZELJKO;
ZHAO, CHEN y
DECIU, COSMIN

(74) Agente/Representante:

DEL VALLE VALIENTE, Sonia

DESCRIPCIÓN

Métodos y procedimientos para la evaluación no invasiva de variaciones genéticas

Campo

La presente invención se define mediante las reivindicaciones. La tecnología proporcionada en el presente documento se refiere en parte a métodos, procesos y máquinas para la evaluación no invasiva de variaciones genéticas.

Antecedentes

La información genética de organismos vivos (por ejemplo, animales, plantas y microorganismos) y otras formas de información genética de replicación (por ejemplo, virus) está codificada en ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). La información genética es una sucesión de nucleótidos o nucleótidos modificados que representan la estructura principal de los ácidos nucleicos químicos o hipotéticos. En seres humanos, el genoma completo contiene aproximadamente 30.000 genes localizados 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 a través de transcripción y traducción, cumple una función bioquímica específica dentro de una célula viva.

Muchas afecciones médicas están provocadas por una o más variaciones genéticas. Determinadas variaciones genéticas provocan 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). Tales enfermedades genéticas pueden resultar de una adición, sustitución o deleción de un solo nucleótido en el ADN de un gen particular. Determinados defectos congénitos están provocados por una anomalía cromosómica, denominada además aneuploidía, tal como trisomía 21 (síndrome de Down), trisomía 13 (síndrome de Patau), trisomía 18 (síndrome de Edward), monosomía X (síndrome de Turner) y determinadas aneuploidías en cromosomas sexuales, tales como síndrome de Klinefelter (XXY), por ejemplo. Otra variación genética es el sexo del feto, que puede determinarse a menudo basándose en los cromosomas sexuales X e Y. Algunas variaciones genéticas pueden predisponer a un individuo a, o provocar, cualquiera de varias enfermedades tales como, por ejemplo, diabetes, arteriosclerosis, obesidad, diversas enfermedades autoinmunitarias y cáncer (por ejemplo, colorrectal, de mama, de ovario, de pulmón).

La identificación de una o más varianzas o variaciones genéticas puede conducir al diagnóstico de, o determinar la predisposición a, una afección médica particular. La identificación de una varianza genética puede dar como resultado que se facilite una decisión médica y/o se emplee un procedimiento médico útil. En determinadas implementaciones, la identificación de una o más varianzas o variaciones genéticas implica el análisis del ADN libre de células. El ADN libre de células (ADNlc) se compone de fragmentos de ADN que se originan a partir de la muerte celular y circulan en la sangre periférica. Las altas concentraciones de ADNlc pueden ser indicativas de determinadas afecciones clínicas, tales como cáncer, traumatismo, quemaduras, infarto de miocardio, accidente cerebrovascular, septicemia, infección y otras enfermedades. Adicionalmente, el ADN fetal libre de células (ADNflc) puede detectarse en el torrente sanguíneo materno y usarse para diversos diagnósticos prenatales no invasivos. Por ejemplo, el documento US 2010/216153 se refiere en general a métodos para detectar ácidos nucleicos fetales y métodos para diagnosticar anomalías fetales.

Resumen

La presente invención se define mediante las reivindicaciones. En consecuencia, la presente invención se refiere a un método para determinar la presencia o ausencia de una trisomía cromosómica, que comprende: (a) obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba de una mujer embarazada; (b) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba; (c) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, que comprende determinar un coeficiente de correlación de la regresión generada en (b), generando de esta manera una evaluación, en donde la evaluación se determina según el coeficiente de correlación y un valor de corte del coeficiente de correlación, en donde: (i) el coeficiente de correlación es igual o superior al valor de corte del coeficiente de correlación y la evaluación es indicativa de una regresión lineal, o (ii) el coeficiente de correlación es inferior al valor de corte del coeficiente de correlación y la evaluación es indicativa de una regresión no lineal; d) normalizar los recuentos mediante un proceso seleccionado según la evaluación, que comprende: (i) en los casos donde la evaluación sea indicativa de una regresión lineal, restar la regresión lineal de los recuentos, o (ii) en los casos donde la evaluación es indicativa de una regresión no lineal, restar la regresión no lineal a los recuentos, generando de esta manera recuentos normalizados con sesgo de GC reducido; y (e) determinar la presencia o ausencia de una trisomía cromosómica según los recuentos normalizados.

También se proporcionan en el presente documento, en determinados aspectos, métodos para analizar el ácido nucleico de una mujer embarazada con sesgo reducido, que comprenden (a) obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba, (b) generar una regresión para (i) los recuentos

y (ii) el contenido de guanina y citosina (GC) para cada una de las porciones del genoma de referencia de la muestra de prueba, (c) determinar un coeficiente de correlación de la regresión y comparar el coeficiente de correlación con un valor de corte del coeficiente de correlación, generando de esta manera una comparación, (d) normalizar los recuentos mediante un proceso seleccionado según la comparación, generando de esta manera recuentos normalizados con sesgo reducido y (e) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados. En algunas implementaciones, una o más, o la totalidad, de las etapas (a), (b), (c), (d) y (e) son realizadas por un procesador, un microprocesador, un ordenador, junto con una memoria y/o por un aparato controlado por microprocesador.

También se proporciona en el presente documento, en determinados aspectos, un método para analizar el ácido nucleico de una mujer embarazada con sesgo reducido, que comprende (a) obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba, (b) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, (c) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, generando de esta manera una evaluación, (d) normalizar los recuentos mediante un proceso seleccionado según la evaluación, generando de esta manera recuentos normalizados con sesgo reducido y (e) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados. En determinados aspectos, la regresión de (b) es una regresión lineal y la normalización de (d) comprende, en los casos donde la evaluación es indicativa de una regresión lineal, restar la regresión lineal de los recuentos.

En determinados aspectos, la normalización de (d) comprende, en los casos donde la evaluación es indicativa de una regresión no lineal, generar una regresión no lineal para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, y restar la regresión no lineal de los recuentos. En determinados aspectos, el método comprende, antes de (a), (i) determinar un valor de incertidumbre para los recuentos mapeados de cada una de las porciones para múltiples muestras de prueba y (ii) seleccionar un subconjunto de porciones que tengan un valor de incertidumbre dentro de rango de valores de incertidumbre predeterminado, reteniendo de esta manera las porciones seleccionadas, donde (a) a (c) se realizan utilizando las porciones seleccionadas. En algunas implementaciones, una o más, o la totalidad, de las etapas (a), (b), (c), (d) y (e) son realizadas por un procesador, un microprocesador, un ordenador, junto con una memoria y/o por un aparato controlado por un microprocesador.

También se proporciona en el presente documento un sistema que comprende uno o más procesadores y una memoria, memoria que comprende instrucciones ejecutables por uno o más procesadores y memoria que comprende recuentos de lecturas de secuencia de ácidos nucleicos mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, (b) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, generando de esta manera una evaluación, (c) normalizar los recuentos mediante un proceso seleccionado según la evaluación, generando de esta manera recuentos normalizados con sesgo reducido y (d) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados.

En el presente documento, en determinados aspectos, se proporciona un método para calcular con sesgo reducido los niveles de sección genómica para una muestra de prueba, que comprende (a) obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba, (b) determinar una o más estimaciones de la curvatura para la muestra de prueba a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) una característica de mapeo para las porciones del genoma de referencia y (c) calcular un nivel de sección genómica normalizado de cada una de las porciones del genoma de referencia para la muestra de prueba según (1) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la muestra de prueba, (2) la una o más estimaciones de la curvatura determinadas en (b) para la muestra de prueba, y (3) una o más estimaciones de la curvatura específicas de la porción de cada una de las múltiples porciones del genoma de referencia de una relación ajustada entre (i) una o más estimaciones de la curvatura específicas de la muestra para una pluralidad de muestras, y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la pluralidad de muestras, proporcionando de esta manera niveles de sección genómica calculados donde se reduce el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia de los niveles de sección genómica calculados. En algunas implementaciones, una o más estimaciones de la curvatura específicas de la muestra de (c)(3) se obtienen a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) la característica de mapeo para cada una de las porciones del genoma de referencia, para cada una de la pluralidad de muestras. En determinadas implementaciones, la característica de mapeo es el contenido de guanina-citosina (GC) de cada una de las porciones del genoma de referencia. En algunas implementaciones, la relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) la característica de mapeo para cada una de las porciones del genoma de referencia, resulta del ajuste a una función seleccionada de una función polinomial; una función racional; una función trascendental; una combinación lineal de funciones exponenciales; una función exponencial de un polinomio; un producto de una función exponencialmente decreciente

y una función logarítmica; un producto de una función exponencialmente decreciente y un polinomio; una función trigonométrica; una combinación lineal de funciones trigonométricas; o combinación de las anteriores.

También se proporciona en el presente documento un sistema que comprende uno o más microprocesadores y una memoria, memoria que comprende instrucciones ejecutables por uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba y cuyas instrucciones ejecutables por uno o más microprocesadores están configuradas para (a) determinar una o más estimaciones de la curvatura para la muestra de prueba a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) una característica de mapeo para las porciones del genoma de referencia y (b) calcular un nivel de sección genómica normalizado de cada una de las porciones del genoma de referencia para la muestra de prueba según (1) los recuentos las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la muestra de prueba, (2) la una o más estimaciones de la curvatura determinadas en (b) para la muestra de prueba, y (3) una o más estimaciones de la curvatura específicas de la porción de cada una de las múltiples porciones del genoma de referencia a partir de una relación ajustada entre (i) una o más estimaciones de la curvatura específicas de la muestra para una pluralidad de muestras, y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la pluralidad de muestras, configuradas de ese modo para proporcionar niveles de sección genómica calculados, donde el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del el genoma de referencia se reduce en los niveles de sección genómica calculados.

En el presente documento, también se proporciona un método para calcular con sesgo reducido los niveles de sección genómica para una muestra de prueba, que comprende (a) obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba, (b) determinar una o más estimaciones de la linealidad para la muestra de prueba a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) una característica de mapeo para las porciones del genoma de referencia y (c) calcular un nivel de sección genómica normalizado de cada una de las porciones del genoma de referencia para la muestra de prueba según (1) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la muestra de prueba, (2) la una o más estimaciones de la curvatura determinadas en (b) para la muestra de prueba, y (3) una o más estimaciones de la linealidad específicas de la porción de cada una de las múltiples porciones del genoma de referencia de una relación ajustada entre (i) una o más estimaciones de la linealidad específicas de la muestra para una pluralidad de muestras, y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la pluralidad de muestras, proporcionando de esta manera niveles de sección genómica calculados donde se reduce el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia de los niveles de sección genómica calculados.

Determinados aspectos de la tecnología se describen adicionalmente en la siguiente descripción, los ejemplos, las reivindicaciones y figuras.

Breve descripción de los dibujos

Las figuras ilustran determinadas realizaciones de la tecnología y no son limitativas. Para mayor claridad y facilidad de ilustración, las figuras no se trazan a escala y, en algunos casos, diversos aspectos pueden mostrarse exagerados o ampliados para facilitar una comprensión de realizaciones particulares.

La FIG. 1 muestra una distribución de los valores de DAM por sección genómica, derivada de 1093 muestras euploides (por ejemplo, mujeres gestantes con un feto euploide) del estudio LDTv2CE. No se realizó escalado según una escala a los recuentos totales. Los triángulos rellenos del eje x indican los valores de corte superior e inferior a partir de los cuales se rechazaron las secciones genómicas.

La FIG. 2 muestra una correlación entre los recuentos sin procesar por sección genómica (eje x) y el contenido de GC por sección genómica (eje y) antes de aplicar la corrección LOESS aditiva. La variación de los recuentos con contenido de GC no es lineal, con un $R^2 < 0,6$.

La FIG. 3 muestra los recuentos normalizados por sección genómica (eje y) en función del contenido de GC por sección genómica (eje x) después de aplicar la corrección LOESS aditiva.

La FIG. 4 muestra una correlación entre los recuentos sin procesar por sección genómica (eje x) y el contenido de GC por sección genómica (eje y) antes de aplicar la corrección de GC basada en una regresión lineal. La variación de los recuentos con el contenido de GC es predominantemente lineal, con un $R^2 > 0,6$.

La FIG. 5 muestra los recuentos normalizados por sección genómica (eje y) en función del contenido de GC por sección genómica (eje x) después de aplicar la corrección lineal aditiva. Los valores negativos se consideran artefactos que finalmente se reemplazan por ceros.

La FIG. 6 muestra las puntuaciones Z21 según PERUN LDTv2CE (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 7 muestra las puntuaciones de Z21 corregidas para el GC híbridas aditivas de LDTv2CE (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 8 muestra las puntuaciones Z18 mediante PERUN de precisión clínica (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 9 muestra las puntuaciones de Z18 híbridas aditivas de precisión clínica (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 10 muestra las puntuaciones Z13 mediante PERUN de validación CLIA (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 11 muestra las puntuaciones de Z13 híbridas aditivas de validación CLIA (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 12 muestra las puntuaciones Z21 mediante PERUN (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 13 muestra las puntuaciones de Z21 híbridas aditivas (eje y) en función de la fracción fetal (FQA-FF) (eje x).

La FIG. 14 muestra una comparación de las puntuaciones Z normalizadas mediante PERUN (eje x) en función de las puntuaciones Z normalizadas mediante un método híbrido aditivo (eje y) para el cromosoma 21 de muestras de LDTv2CE.

La FIG. 15 muestra una comparación de puntuaciones Z normalizadas mediante PERUN (eje x) en función de las puntuaciones Z normalizadas mediante un método híbrido aditivo (eje y) para el cromosoma 21 para muestras de precisión clínica.

La FIG. 16 muestra una comparación de las puntuaciones Z normalizadas mediante PERUN (eje x) en función de las puntuaciones Z normalizadas mediante un método híbrido aditivo (eje y) para el cromosoma 21 para muestras de validación CLIA.

La FIG. 17 muestra una realización ilustrativa de un sistema en el que pueden implementarse determinadas realizaciones de la tecnología.

La FIG. 18 muestra una correlación entre los coeficientes lineales y cuadráticos observados extraídos de múltiples datos de muestra.

Descripción detallada

En el presente documento, se proporcionan métodos para determinar la presencia o ausencia de una variación genética fetal (por ejemplo, una aneuploidía cromosómica) en un feto donde se realiza una determinación, en parte y/o en su totalidad, según las secuencias de ácido nucleico. En algunas implementaciones, las secuencias de ácido nucleico se obtienen de una muestra obtenida de una mujer embarazada (por ejemplo, de la sangre de una mujer embarazada). También se proporcionan en el presente documento métodos mejorados de manipulación de datos, así como sistemas, aparatos, máquinas y módulos que, en algunas implementaciones, llevan a cabo los métodos descritos en el presente documento. En algunas implementaciones, identificar una variación genética mediante un método descrito en el presente documento puede conducir a un diagnóstico de, o a determinar una predisposición a, una afección médica particular. La identificación de una varianza genética puede dar como resultado que se facilite una decisión médica y/o se emplee un procedimiento médico útil.

Muestras

En el presente documento se proporcionan métodos y composiciones para analizar ácido nucleico. En algunas implementaciones, se analizan los fragmentos de ácido nucleico en una mezcla de fragmentos de ácido nucleico. Una mezcla de ácidos nucleicos puede comprender dos o más especies de fragmento de ácido nucleico que tienen secuencias de nucleótidos diferentes, longitudes de fragmento diferentes, orígenes diferentes (por ejemplo, orígenes genómicos, orígenes fetales frente a maternos, orígenes celulares o tisulares, orígenes de muestra, orígenes de sujeto, y similares), o combinaciones de los mismos.

El ácido nucleico o una mezcla de ácido nucleico usados en los métodos, máquinas y/o aparatos descritos en el presente documento se suele aislar de una muestra obtenida de un sujeto. Un sujeto puede ser cualquier organismo vivo o no vivo incluyendo, pero sin limitarse a, un ser humano, un animal no humano, una planta, una bacteria, un hongo o un protista. Puede seleccionarse cualquier animal humano o no humano incluyendo, pero sin limitarse a, mamífero, reptil, ave, anfibio, pez, ungulado, rumiante, bovino (por ejemplo, ganado vacuno), equino (por ejemplo, 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, pescado, delfín, ballena y tiburón. Un sujeto puede ser un hombre o una mujer (por ejemplo, una mujer, una mujer embarazada). Un sujeto puede tener cualquier edad (por ejemplo, un embrión, un feto, un bebé, un niño, un adulto).

El ácido nucleico puede aislarse de cualquier tipo de espécimen o muestra biológica adecuada (por ejemplo, una muestra de prueba). Una muestra o muestra de prueba puede ser cualquier espécimen que se aísle u obtenga de un sujeto o parte del mismo (por ejemplo, un sujeto humano, una mujer embarazada, un feto). Los ejemplos no limitativos de especímenes incluyen fluidos o tejidos 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 medular, 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 celulares fetales) o partes de las mismas (por ejemplo, mitocondriales, núcleos, extractos o similares), lavados del aparato genital femenino, orina, heces, esputo, saliva, mucosa nasal, fluido prostático, lavados, semen, fluido linfático, bilis, lágrimas, sudor, leche materna, fluido mamario, similares o combinaciones de los mismos. En algunas implementaciones, una muestra biológica es un hisopo cervicouterino de un sujeto. En algunas implementaciones, una muestra biológica puede ser sangre y, algunas 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 embarazada o una mujer a la que se le está realizando una prueba de posible embarazo. El término abarca sangre completa, hemoderivados o cualquier fracción de sangre, como suero, plasma, capa leucocitaria o similares, según se define convencionalmente. La sangre o fracciones de la misma comprenden a menudo nucleosomas (por ejemplo, nucleosomas maternos y/o fetales). Los nucleosomas comprenden ácidos nucleicos y, algunas veces, son sin células o intracelulares. La sangre comprende además capas leucocíticas. Algunas veces, las capas leucocíticas se aíslan utilizando un gradiente de Ficoll. Las capas leucocitarias pueden comprender glóbulos blancos (por ejemplo, leucocitos, linfocitos T, linfocitos B, plaquetas y similares). En determinadas implementaciones, las capas leucocitarias comprenden ácido nucleico materno y/o fetal. Plasma sanguíneo se refiere a la fracción de sangre completa que resulta de la centrifugación de sangre tratada con anticoagulantes. Suero sanguíneo se refiere a la porción acuosa del fluido que queda después de que se ha coagulado una muestra de sangre. A menudo, se recogen muestras de líquido o tejido según protocolos convencionales que siguen generalmente hospitales o clínicas. A menudo, para la sangre se recoge una cantidad adecuada de sangre periférica (por ejemplo, entre 3-40 mililitros) y puede almacenarse según procedimientos convencionales antes o después de la preparación. Una muestra de líquido o tejido de la que se extrae ácido nucleico puede ser acelular (por ejemplo, libre de células). En algunas implementaciones, una muestra de líquido o tejido puede contener elementos celulares o restos celulares. En algunas implementaciones, pueden incluirse células fetales o células cancerosas en la muestra.

Una muestra es a menudo heterogénea, lo que significa que más de un tipo de especie de ácido nucleico está presente en la muestra. Por ejemplo, el ácido nucleico heterogéneo puede incluir, pero no se limita a, (i) ácido nucleico derivado fetal y derivado materno, (ii) ácido nucleico de cáncer y distinto de cáncer, (iii) ácido nucleico patógeno y de huésped y, más generalmente, (iv) ácido nucleico mutado y de tipo natural. Una muestra puede ser heterogénea porque más de un tipo de célula está presente, tal como una célula fetal y una célula materna, una célula cancerosa y no cancerosa, o una célula patógena y de huésped. En algunas implementaciones, están presentes una especie minoritaria de ácido nucleico y una mayoría de especies de ácido nucleico.

Para las aplicaciones prenatales de la tecnología descrita en el presente documento, la muestra de líquido o tejido puede recogerse de una mujer a una edad gestacional adecuada para la prueba, o de una mujer que está sometiéndose a prueba para un posible embarazo. La edad gestacional adecuada puede variar dependiendo de la prueba prenatal que se realiza. En determinadas implementaciones, un sujeto de sexo femenino gestante algunas veces está en el primer trimestre de embarazo, a veces en el segundo trimestre de embarazo o, algunas veces, en el tercer trimestre de embarazo. En determinadas implementaciones, se recoge un líquido o tejido de una mujer embarazada de aproximadamente 1 a aproximadamente 45 semanas de gestación fetal (por ejemplo, a las 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 de aproximadamente 5 a 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 determinadas implementaciones, se recoge una muestra de líquido o tejido de una mujer embarazada durante o justo después (por ejemplo, de 0 a 72 horas después) de dar a luz (por ejemplo, parto vaginal o no vaginal (por ejemplo, parto quirúrgico)).

Adquisición de muestras de sangre y extracción de ADN

Los métodos incluidos en el presente documento suelen incluir separar, enriquecer y analizar el ADN fetal que se encuentra en la sangre materna como medio no invasivo para detectar la presencia o ausencia de una variación genética materna y/o fetal y/o controlar la salud de un feto y/o una mujer embarazada durante la gestación y, a veces, tras ella. Por lo tanto, las primeras etapas para poner en práctica determinados métodos en el presente documento suelen incluir la obtención de una muestra de sangre de una mujer embarazada y la extracción de ADN de una muestra.

Adquisición de muestras de sangre

Puede obtenerse una muestra de sangre de una mujer embarazada a una edad gestacional adecuada para la prueba utilizando un método de la presente tecnología. Una edad gestacional adecuada puede variar en función

del trastorno analizado, como se explica más adelante. La extracción de sangre de una mujer suele realizarse según el protocolo convencional que suelen seguir los hospitales o las clínicas. Se suele extraer una cantidad adecuada de sangre periférica, por ejemplo, normalmente entre 5-50 ml, que puede almacenarse según el procedimiento convencional antes de su posterior preparación. Las muestras de sangre se pueden extraer, almacenar o transportar de manera que se minimice la degradación o la calidad del ácido nucleico presente en la muestra.

Preparación de muestras de sangre

Puede realizarse un análisis del ADN fetal encontrado en la sangre materna utilizando, por ejemplo, sangre completa, suero o plasma. Se conocen métodos para preparar suero o plasma a partir de sangre materna. Por ejemplo, se puede colocar sangre de una mujer embarazada 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 luego se puede obtener plasma de la sangre completa mediante centrifugación. El suero se puede obtener con o sin centrifugación, después de la coagulación de la sangre. Si se utiliza la centrifugación, normalmente, aunque no exclusivamente, se realiza a una velocidad adecuada, por ejemplo, 1500-3000 xg. El plasma o el suero se pueden someter a etapas de centrifugación adicionales antes de transferirlos a un tubo nuevo para la extracción del ADN.

Además de la porción acelular de la sangre completa, también se puede recuperar el ADN de la fracción celular, enriquecida en la porción de la capa leucocitaria, que se puede obtener tras la centrifugación de una muestra de sangre completa de la mujer y la extracción del plasma.

Extracción de ADN

Existen numerosos métodos conocidos para extraer el ADN de una muestra biológica que incluye sangre. Se pueden seguir métodos generales de preparación de ADN (por ejemplo, descritos por Sambrook y Russell, "Molecular Cloning: A Laboratory Manual", 3.ª edición, 2001); también se pueden utilizar diferentes reactivos o kits disponibles en el mercado, tales como el kit de ácido nucleico circulante QIAamp de Qiagen, el mini kit de ADN QiaAmp o el mini kit de sangre de ADN QiaAmp (Qiagen, Hilden, Alemania), kit de aislamiento de ADN de la sangre GenomicPrep™ (Promega, Madison, Wis.) y el kit de purificación de ADN de sangre genómica GFX™ (Amersham, Piscataway, NJ) para obtener ADN de una muestra de sangre de una mujer embarazada. También se pueden utilizar combinaciones de más de uno de estos métodos.

En algunas implementaciones, la muestra puede enriquecerse o enriquecerse relativamente en ácido nucleico fetal mediante uno o más métodos. Por ejemplo, se puede realizar la diferenciación del ADN fetal y del materno mediante las composiciones y los procesos de la presente tecnología solos o junto con otros factores de diferenciación. Los ejemplos de estos factores incluyen, entre otros, diferencias de un solo nucleótido entre los cromosomas X e Y, secuencias específicas del cromosoma Y, polimorfismos ubicados 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.

Otros métodos para enriquecer una muestra para una especie particular de ácido nucleico se describen en la solicitud de patente PCT número PCT/US07/69991, presentada el 30 de mayo de 2007, solicitud de patente PCT número PCT/US2007/071232, presentada el 15 de junio de 2007, solicitudes provisionales estadounidenses 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 determinadas implementaciones, el ácido nucleico materno se elimina selectivamente (bien parcial, sustancial, casi completa o completamente) de la muestra.

Las expresiones "ácido nucleico" y "molécula de ácido nucleico" pueden utilizarse indistintamente a lo largo de la divulgación. Los términos se refieren a ácidos nucleicos de cualquier composición de, tal como ADN (por ejemplo, ADN complementario (ADNc), ADN genómico (ADNg) y similares), ARN (por ejemplo, ARN mensajero (ARNm), ARN inhibidor corto (ARNic), ARN ribosómico (ARNr), ARNt, microARN, ARN altamente expresado por el feto o la placenta, y similares), y/o análogos de ADN o ARN (por ejemplo, que contengan análogos de bases, análogos de azúcares y/o una cadena principal no nativa y similares), moléculas híbridas de ARN/ARN y ácidos nucleicos poliamídicos (ANP), la totalidad de los cuales puede estar en forma monocatenaria o bicatenaria y, a menos que se limite de otro modo, pueden abarcar análogos conocidos de nucleótidos naturales que pueden funcionar de forma similar a los nucleótidos naturales. Un ácido nucleico puede ser, o puede proceder de, un plásmido, fago, una secuencia de replicación autónoma (ARS), un centrómero, cromosoma artificial, cromosoma u otro ácido nucleico capaz de replicar o replicarse *in vitro* o en una célula huésped, una célula, un núcleo celular o un citoplasma de una célula en determinadas implementaciones. Un molde de ácido nucleico, en algunas implementaciones, puede ser de un solo cromosoma (por ejemplo, una muestra de ácido nucleico puede proceder de un cromosoma de una muestra obtenida de un organismo diploide). A menos que se limite específicamente, la expresión 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 naturales. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también incluye 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) se sustituye con restos de base mixta y/o desoxiinosina. La expresión ácido nucleico se usa indistintamente con locus, gen, ADNc y ARNm codificados por un gen. La expresión también puede incluir, como

equivalentes, derivados, variantes y análogos de ARN o ADN sintetizados a partir de análogos de nucleótidos, polinucleótidos monocatenarios (“sentido” o “antisentido”, cadena “positiva” o cadena “negativa”, marco de lectura “directo” o marco de lectura “inverso”) y bicatenarios. El término “gen” significa el segmento de ADN que interviene en la producción de una cadena polipeptídica; incluye regiones que preceden y siguen a la región codificadora (líder y remolque) que interviene 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 citosina se reemplaza por uracilo. Un ácido nucleico molde puede prepararse con el uso de un ácido nucleico obtenido de un sujeto como molde.

Aislamiento y procesamiento de ácidos nucleicos

El ácido nucleico puede derivar de una o más fuentes (por ejemplo, células, suero, plasma, capa leucocítica, líquido linfático, piel, suelo, y similares) mediante métodos conocidos en la técnica. Se puede utilizar cualquier método adecuado para aislar, extraer y/o purificar el ADN de una muestra biológica (por ejemplo, de sangre o de un producto sanguíneo), cuyos ejemplos no limitativos incluyen métodos de preparación de ADN (por ejemplo, los descritos por Sambrook y Russell, “Molecular Cloning: A Laboratory Manual” 3.ª ed., 2001), diferentes reactivos o kits disponibles en el mercado, tal como el kit de ácido nucleico circulante QIAamp de Qiagen, el mini kit de ADN QiaAmp o el mini kit de sangre de ADN QiaAmp (Qiagen, Hilden, Alemania), kit de aislamiento de ADN de la sangre GenomicPrep™ (Promega, Madison, Wis.) y el kit de purificación de ADN de sangre genómica GFX™ (Amersham, Piscataway, NJ), similares o combinaciones de los mismos.

Los procedimientos y reactivos de lisis celular se conocen en la técnica y pueden realizarse generalmente mediante métodos químicos (por ejemplo, detergentes, disoluciones hipotónicas, procedimientos enzimáticos, y similares, o una combinación de los mismos), físicos (por ejemplo, prensa francesa, sonicación, y similares) o electrolíticos de lisis. Puede utilizarse cualquier procedimiento de lisis adecuado. Por ejemplo, los métodos químicos emplean generalmente agentes de lisis para perturbar las células y extraer los ácidos nucleicos de las células, seguido por el tratamiento con sales caotrópicas. Además, son útiles los métodos físicos, tales como congelación/descongelación seguida de trituración, el uso de prensas celulares y similares. Además, se usan habitualmente procedimientos de lisis con alto contenido de sal. Por ejemplo, puede utilizarse un procedimiento de lisis alcalina. Este último procedimiento incorpora tradicionalmente el uso de disoluciones de fenol-cloroformo, y puede usarse un procedimiento alternativo libre de fenol-cloroformo que involucra tres disoluciones. En los últimos procedimientos, una solución puede contener Tris 15 mM, pH 8,0; EDTA 10 mM y 100 ug/ml de ARNasa 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 pueden encontrarse en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989).

El ácido nucleico puede aislarse en un punto de tiempo diferente en comparación con otro ácido nucleico, en donde cada una de las muestras procede de la misma fuente o de una fuente diferente. Un ácido nucleico puede ser de una biblioteca de ácido nucleico, tal como una biblioteca de ADNc o ARN, por ejemplo. Un ácido nucleico puede ser un resultado de la purificación o el aislamiento de ácido nucleico y/o la amplificación de moléculas de ácido nucleico de la muestra. El ácido nucleico proporcionado para los procedimientos 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).

Los ácidos nucleicos pueden incluir ácido nucleico extracelular en determinadas implementaciones. La expresión “ácido nucleico extracelular”, tal como se usa en el presente documento, puede referirse a ácido nucleico aislado de una fuente que no tiene sustancialmente células y también se denomina ácido nucleico “libre de células” y/o ácido nucleico “circulante libre de células”. El ácido nucleico extracelular puede estar presente en y obtenerse de sangre (por ejemplo, de la sangre de una mujer embarazada). El ácido nucleico extracelular incluye a menudo células no detectables y puede contener elementos celulares o restos celulares. Los ejemplos no limitativos de fuentes acelulares de ácido nucleico extracelular son sangre, plasma sanguíneo, suero sanguíneo y orina. Como se utiliza en el presente documento, la expresión “obtener ácido nucleico de muestra circulante libre de células” incluye obtener una muestra directamente (por ejemplo, extraer una muestra, por ejemplo, una muestra de prueba) u obtener una muestra de otra persona que haya extraído una muestra. Sin desear limitarse por la teoría, el ácido nucleico extracelular puede ser un producto de la apoptosis celular y la descomposición celular, lo que proporciona una base para el ácido nucleico extracelular que tiene a menudo una serie de longitudes a través de un espectro (por ejemplo, una “escalera”).

El ácido nucleico extracelular puede incluir diferentes especies de ácido nucleico y, por tanto, se denomina en el presente documento “heterogéneo” en determinadas implementaciones. 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 mujer embarazada puede incluir ácido nucleico materno y ácido nucleico fetal. En algunos casos, el ácido nucleico fetal algunas veces es de aproximadamente el 5 % a aproximadamente el 50 % del ácido nucleico global (por ejemplo, aproximadamente el 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 ácido nucleico total es ácido nucleico fetal). En algunas implementaciones, la mayoría del ácido nucleico fetal del ácido nucleico tiene una longitud de aproximadamente 500 pares de bases o menor, aproximadamente 250 pares de bases o menor,

aproximadamente 200 pares de bases o menor, aproximadamente 150 pares de bases o menor, aproximadamente 100 pares de bases o menor, aproximadamente 50 pares de bases o menor o aproximadamente 25 pares de bases o menor.

El ácido nucleico podría proporcionarse para realizar los métodos descritos en el presente documento sin procesar la(s) muestra(s) que contiene(n) el ácido nucleico, en determinadas implementaciones. En algunas implementaciones, se proporciona ácido nucleico para llevar a cabo 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 puede extraerse, aislarse, purificarse, purificarse parcialmente o amplificarse a partir de la(s) muestra(s). El término "aislado", tal como se usa en el presente documento, se refiere a un ácido nucleico retirado de su entorno original (por ejemplo, el entorno natural si se produce de manera natural o una célula huésped si se expresa de manera exógena) y, por tanto, se altera por intervención humana (por ejemplo, "por la mano del hombre") de su entorno original. La expresión "ácido nucleico aislado", tal como se usa en el presente documento, puede referirse a un ácido nucleico retirado de un sujeto (por ejemplo, un sujeto humano). Un ácido nucleico aislado puede proporcionarse con menos componentes distintos de ácido nucleico (por ejemplo, proteína, lípido) que la cantidad de componentes presentes en una muestra de origen. Una composición que comprende ácido nucleico aislado puede estar libre en de aproximadamente el 50 % a más del 99 % de componentes distintos de ácido nucleico. Una composición que comprende ácido nucleico aislado puede estar libre en aproximadamente el 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más del 99 % de componentes distintos de ácido nucleico. El término "purificado", tal como se usa en el presente documento, puede referirse a un ácido nucleico siempre que contenga menos componentes distintos de ácido nucleico (por ejemplo, proteína, lípido, hidrato de carbono) que la cantidad de componentes distintos de ácido nucleico presentes antes de someter el ácido nucleico a un procedimiento de purificación. Una composición que comprende ácido nucleico purificado puede estar libre en aproximadamente el 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más del 99 % de otros componentes distintos de ácido nucleico. El término "purificado", tal 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 se deriva el ácido nucleico. Una composición que comprende ácido nucleico purificado puede estar libre en aproximadamente el 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más del 99 % de otras especies de ácido nucleico. Por ejemplo, el ácido nucleico fetal puede purificarse a partir de una mezcla que comprende ácido nucleico materno y fetal. En determinados ejemplos, los nucleosomas que comprenden fragmentos pequeños 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 algunas implementaciones, 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, media o promedio de aproximadamente 5 a aproximadamente 10.000 pares de bases, de aproximadamente 100 a aproximadamente 1000 pares de bases, de 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 mediante la selección de un procedimiento de generación de fragmentos adecuado.

Los fragmentos de ácido nucleico pueden contener secuencias de nucleótidos solapantes, y esas 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, en donde x, y y z son secuencias de nucleótidos que pueden tener 5 nucleótidos de longitud o más. La secuencia solapante y puede utilizarse para facilitar la construcción de la secuencia de nucleótidos x-y-z en ácido nucleico a partir de una muestra en determinadas implementaciones. El ácido nucleico puede estar parcialmente fragmentado (por ejemplo, a partir de una reacción de escisión específica incompleta o terminada) o completamente fragmentado en determinadas implementaciones.

En algunas implementaciones, el ácido nucleico se fragmenta o escinde mediante un método adecuado, cuyos ejemplos no limitativos incluyen métodos físicos (por ejemplo, cizalla, por ejemplo, sonicación, prensa francesa, calor, radiación UV, similares), procesos enzimáticos (por ejemplo, agentes de escisión (por ejemplo, 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.° 20050112590, similares o combinaciones de los mismos.

Tal como se usa en el presente documento, "fragmentación" o "escisión" se refiere a un procedimiento o condiciones en los que una molécula de ácido nucleico, tal como una molécula de gen molde de ácido nucleico o producto amplificado de la misma, puede cortarse en dos o más moléculas de ácido nucleico más pequeñas. Tal 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.

Tal 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 molde de ácido nucleico o producto amplificado de la misma. Aunque tales fragmentos o productos escindidos pueden referirse a todas las moléculas de ácido nucleico resultantes de una reacción de escisión normalmente tales fragmentos o productos escindidos se refieren solamente a moléculas de ácido nucleico resultantes de una fragmentación o escisión de una molécula de gen 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 molde de ácido nucleico. El término "amplificado", tal como se usa en el presente documento, se refiere a someter un ácido nucleico diana en una muestra a un procedimiento que genera lineal o exponencialmente ácidos nucleicos de amplicón que tienen la misma secuencia de nucleótidos o prácticamente la misma secuencia de nucleótidos que el ácido nucleico diana o segmento del mismo. En determinadas implementaciones, 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 molde de ácido nucleico (por ejemplo, un cebador puede contener nucleótidos "extra" tales como una secuencia de iniciación de la transcripción, además de los nucleótidos complementarios a una molécula de gen molde de ácido nucleico, lo que da 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 molde de ácido nucleico). En consecuencia, los fragmentos pueden incluir fragmentos que surgen de porciones, segmentos o partes de moléculas de ácido nucleico amplificadas que contienen, al menos en parte, información de secuencia de nucleótidos de o basada en la molécula molde de ácido nucleico representativa.

Tal 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 con el uso de diferentes reactivos de escisión o mediante la alteración de la especificidad de escisión del mismo reactivo de escisión de tal manera que se generan patrones de escisión alternativos del mismo ácido nucleico o proteína diana o de referencia. En determinadas implementaciones, 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 recipientes de reacción (por ejemplo, el ácido nucleico se trata con cada agente de escisión específico en un recipiente independiente). La expresión "agente de escisión específico", tal como se usa en el presente documento, se refiere a un agente, algunas veces, una sustancia química o una enzima que puede escindir un ácido nucleico en uno o más sitios específicos. Cualquier agente de escisión enzimática inespecífico o específico adecuado puede usarse para escindir o fragmentar ácidos nucleicos. Un agente de escisión enzimático específico adecuado puede ser una enzima de restricción. Una enzima de restricción adecuada puede usarse para escindir ácidos nucleicos, en algunas implementaciones.

El ácido nucleico puede exponerse además a un procedimiento que modifica determinados nucleótidos en el ácido nucleico antes de proporcionar ácido nucleico para un método descrito en el presente documento. Un procedimiento que modifica selectivamente el ácido nucleico basándose en el estado de metilación de nucleótidos en el mismo puede aplicarse al ácido nucleico, por ejemplo. Además, condiciones tales como alta temperatura, radiación ultravioleta, radiación de rayos 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.

El ácido nucleico puede ser monocatenario o bicatenario. El ADN monocatenario, por ejemplo, puede generarse mediante la desnaturalización de ADN bicatenario mediante calentamiento o mediante tratamiento con álcali, por ejemplo. En determinadas implantaciones, el ácido nucleico se encuentra en una estructura de bucle D, formada mediante invasión de hebra de una molécula de ADN doble por un oligonucleótido o una molécula similar a ADN, tal como ácido nucleico peptídico (PNA). La formación del bucle D puede facilitarse mediante la adición de la proteína RecA de *E. coli* y/o mediante la alteración de la concentración de sal, por ejemplo, usando métodos conocidos en la técnica.

Determinación del contenido de ácido nucleico fetal

La cantidad de ácido nucleico fetal (por ejemplo, concentración, cantidad relativa, cantidad absoluta, número de copias y similares) en ácido nucleico se determina en algunas implementaciones. En determinadas implantaciones, la cantidad de ácido nucleico fetal en una muestra se denomina "fracción fetal". En algunas implantaciones, "fracción fetal" se refiere a la fracción de ácido nucleico fetal en 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 embarazada. En determinadas implementaciones, la cantidad de ácido nucleico fetal se determina según marcadores específicos de un feto de sexo masculino (por ejemplo, marcadores STR del cromosoma Y (por ejemplo, marcadores DYS 19, DYS 385, DYS 392); marcador RhD en mujeres sin expresión de RhD), proporciones alélicas de secuencias polimórficas, o según uno o más marcadores específicos del ácido nucleico fetal y no del ácido nucleico materno (por ejemplo, biomarcadores epigenéticos diferenciales (por ejemplo, metilación; se describe con más detalle a continuación) entre la madre y el feto, o marcadores de ARN fetal en el 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) se realiza, algunas veces, usando un ensayo cuantificador fetal (FQA) tal como se describe, por ejemplo, en la publicación de solicitud de patente estadounidense n.° 2010/0105049. Este tipo de ensayo permite la detección y cuantificación de ácido nucleico fetal en una muestra materna basándose en el estado de metilación del ácido nucleico en la muestra. En determinadas implantaciones, la cantidad de ácido nucleico fetal de una muestra materna puede determinarse con respecto a la cantidad total de ácido nucleico presente, lo que proporciona de ese modo el porcentaje de ácido nucleico fetal en la muestra. En determinadas implantaciones, el número de copias de ácido nucleico fetal puede determinarse en una muestra materna. En determinadas implantaciones, la cantidad de ácido nucleico fetal puede determinarse de manera

específica de secuencia (o específica de la porción) y algunas veces con suficiente sensibilidad para permitir un análisis preciso de la dosificación cromosómica (por ejemplo, para detectar la presencia o ausencia de una aneuploidía fetal).

Un ensayo cuantificador fetal (FQA) puede realizarse junto con cualquiera de los métodos descritos en el presente documento. Tal ensayo puede realizarse mediante cualquier método conocido en la técnica y/o descrito en la publicación de solicitud de patente estadounidense n.° 2010/0105049, tal como, por ejemplo, mediante un método que puede distinguir entre ADN materno y fetal basándose en el estado de metilación diferencial, y cuantificar (es decir, determinar la cantidad de) el ADN fetal. Los métodos para diferenciar ácido nucleico basados en el estado de metilación incluyen, pero no se limitan a, captura sensible a la metilación, por ejemplo, usando un fragmento Fc de MBD2 en el cual el dominio de unión a metilo de MBD2 se fusiona al fragmento Fc de un anticuerpo (MBD-FC) (Gebhard *y col*. (2006) *Cáncer Res*. 66(12):6118-28); anticuerpos específicos de la metilación; métodos de conversión de bisulfito, por ejemplo, MSP (PCR sensible a la metilación), COBRA, extensión de cebador de un solo nucleótido sensible a la metilación (Ms-SNuPE) o tecnlogía MassCLEAVE™ de Sequenom; y el uso de enzimas de restricción sensibles a la metilación (por ejemplo, digestión de ADN materno en una muestra materna usando una o más enzimas de restricción sensibles a la metilación, enriqueciendo de este modo el ADN fetal). Las enzimas sensibles a metilo pueden usarse además para diferenciar ácido nucleico basándose en el estado de metilación que, por ejemplo, puede escindir o digerir preferente o sustancialmente en su secuencia de reconocimiento de ADN si esta última no está metilada. Por tanto, se corta una muestra de ADN no metilado en fragmentos más pequeños que una muestra de ADN metilado y no se escinde una muestra de ADN hipermetilado. Excepto cuando se indique explícitamente, puede usarse cualquier método para diferenciar ácido nucleico basado en el estado de metilación con las composiciones y métodos de la tecnología en el presente documento. La cantidad de ADN fetal puede determinarse, por ejemplo, introduciendo uno o más competidores a concentraciones conocidas durante una reacción de amplificación. Además, la determinación de la cantidad de ADN fetal puede realizarse, por ejemplo, mediante RT-PCR, extensión de cebadores, secuenciación y/o recuento. En determinados casos, la cantidad de ácido nucleico puede determinarse usando la tecnología BEAMing tal como se describe en la publicación de solicitud de patente estadounidense n.° 2007/0065823. En determinadas implantaciones, puede determinarse la eficiencia de restricción y el índice de eficiencia se usa para determinar además la cantidad de ADN fetal.

En determinadas implementaciones, se puede usar un ensayo cuantificador 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 de una muestra materna utilizando una o más enzimas de restricción sensibles a la metilación, enriqueciendo de este modo 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 de este modo la concentración de ADN fetal de la muestra materna. En determinadas implantaciones, el número absoluto de copias de ácido nucleico fetal en una muestra materna puede determinarse, por ejemplo, mediante espectrometría de masas y/o un sistema que usa un enfoque competitivo de PCR para las mediciones absolutas del número de copias. Véase, por ejemplo, Ding y Cantor (2003) Proc.Natl.Acad.Sci. USA 100:3059-3064 y la publicación de solicitud de patente estadounidense n.° 2004/0081993.

En determinadas implantaciones, la fracción fetal puede determinarse basándose en las razones alélicas de secuencias polimórficas (por ejemplo, polimorfismos de un solo nucleótido (SNP)) tal como, por ejemplo, usando un método descrito en la publicación de solicitud de patente estadounidense n.° 2011/0224087. En tal método, se obtienen lecturas de secuencia de nucleótidos para una muestra materna y se determina la fracción fetal comparando el número total de lecturas de secuencia de nucleótidos que se mapean en un primer alelo y el número total de lecturas de secuencia de nucleótidos que se mapean en un segundo alelo en un sitio polimórfico informativo (por ejemplo, SNP) en un genoma de referencia. En determinadas implantaciones, los alelos fetales se identifican, por ejemplo, por su contribución minoritaria relativa a la mezcla de ácidos nucleicos fetales y maternos en la muestra cuando se compara con la contribución mayoritaria a la mezcla por los ácidos nucleicos maternos. En consecuencia, la abundancia relativa de ácido nucleico fetal en una muestra materna puede determinarse como un parámetro del número total de lecturas de secuencia 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 el presente documento. Por tanto, en determinadas implementaciones, los métodos de la tecnología descrita en el presente documento comprenden una etapa adicional para determinar la cantidad de ácido nucleico fetal. La cantidad de ácido nucleico fetal puede determinarse en una muestra de ácido nucleico de un sujeto antes o después del procesamiento para preparar el ácido nucleico de muestra. En determinadas implementaciones, la cantidad de ácido nucleico fetal se determina en una muestra después de procesar y preparar el ácido nucleico de muestra, cantidad que se usa para una evaluación adicional. En algunas implementaciones, un resultado comprende factorizar la fracción de ácido nucleico fetal en el ácido nucleico de muestra (por ejemplo, ajustar recuentos, eliminar muestras, realizar una identificación o no realizar una identificación).

La etapa de determinación puede realizarse antes, durante, en un punto cualquiera de un método descrito en el presente documento, o después de determinados métodos (por ejemplo, detección de aneuploidía, determinación del sexo del feto) descritos en el presente documento. Por ejemplo, para lograr un método de determinación de aneuploidía o sexo del feto con una sensibilidad o especificidad dadas, puede implementarse un método de cuantificación de ácido nucleico fetal antes de, durante o después de la determinación de aneuploidía o sexo del feto para identificar aquellas muestras

con más de aproximadamente el 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 algunas implementaciones, las muestras que tienen una determinada cantidad umbral de ácido nucleico fetal (por ejemplo, de aproximadamente el 15 % o más de ácido nucleico fetal; aproximadamente el 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 determinadas implementaciones, las determinaciones de, por ejemplo, sexo del feto o la presencia o ausencia de aneuploidía se seleccionan (por ejemplo, se seleccionan y comunican a un paciente) solo para muestras que tienen una determinada cantidad umbral de ácido nucleico fetal (por ejemplo, aproximadamente el 15 % o más de ácido nucleico fetal; aproximadamente el 4 % o más de ácido nucleico fetal).

En algunas implementaciones, la determinación de fracción fetal o la determinación de la cantidad de ácido nucleico fetal no es necesaria o se requiere para identificar la presencia o ausencia de una aneuploidía cromosómica. En algunas implementaciones, la identificación de la presencia o ausencia de una aneuploidía cromosómica no requiere la diferenciación de secuencias de ADN fetal en comparación con el ADN materno. En determinadas implementaciones, esto se debe a que se analiza la contribución sumada de las secuencias materna y fetal en un cromosoma particular, porción cromosómica o segmento de los mismos. En algunas implementaciones, la identificación de la presencia o ausencia de una aneuploidía cromosómica no depende de una información de secuencia a priori que distinguiría el ADN fetal del ADN materno.

Ácidos nucleicos enriquecedores

En algunas implementaciones, el ácido nucleico (por ejemplo, ácido nucleico extracelular) está enriquecido o relativamente enriquecido para una subpoblación o especie de ácido nucleico. Las subpoblaciones de ácidos nucleicos pueden incluir, por ejemplo, ácido nucleico fetal, ácido nucleico materno, ácido nucleico que comprende fragmentos de una longitud o rango de longitudes particular, o ácido nucleico de una región particular del genoma (por ejemplo, cromosoma único, conjunto de cromosomas, y/o determinadas regiones cromosómicas). Tales muestras enriquecidas pueden usarse junto con un método proporcionado en el presente documento. Por tanto, en determinadas implementaciones, los métodos de la tecnología comprenden una etapa adicional de enriquecimiento en una subpoblación de ácido nucleico en una muestra, tal como, por ejemplo, ácido nucleico fetal. En determinadas implementaciones, un método para determinar la fracción fetal descrita anteriormente puede usarse además para enriquecer el ácido nucleico fetal. En determinadas implementaciones, el ácido nucleico materno se elimina selectivamente (parcial, sustancialmente, casi completamente o completamente) de la muestra. En determinadas implementaciones, el enriquecimiento de un ácido nucleico de especie de 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 estadounidense n.° 6.927.028, la publicación de solicitud de patente internacional n.° WO2007/140417, la publicación de solicitud de patente internacional n.° WO2007/147063, la publicación de solicitud de patente internacional n.° WO2009/032779, la publicación de solicitud de patente internacional n.° WO2009/032781, la publicación de solicitud de patente internacional n.° WO2010/033639, la publicación de solicitud de patente internacional n.° WO2011/034631, la publicación de solicitud de patente internacional n.° WO2006/056480 y la publicación de solicitud de patente internacional n.° WO2011/143659.

En algunas implementaciones, el ácido nucleico se enriquece en determinadas especies de fragmento diana y/o especies de fragmento de referencia. En determinadas implementaciones, el ácido nucleico se enriquece en una longitud específica de fragmento de ácido nucleico o rango de longitudes de fragmento con el uso de uno o más métodos de separación basados en la longitud descritos a continuación. En determinadas implementaciones, el ácido nucleico se enriquece en fragmentos de una región genómica seleccionada (por ejemplo, cromosoma) con el uso de uno o más métodos de separación basados en secuencia descritos en el presente documento y/o conocidos en la técnica. Determinados métodos para enriquecer una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) en una muestra se describen con detalle más adelante.

Algunos métodos para enriquecer una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que pueden usarse con un método descrito en el presente documento incluyen métodos que aprovechan diferencias epigenéticas entre ácido nucleico materno y fetal. Por ejemplo, el ácido nucleico fetal puede diferenciarse y separarse del ácido nucleico materno basándose en diferencias de metilación. Se describen métodos de enriquecimiento de ácido nucleico fetal basados en metilación en la publicación de solicitud de patente estadounidense n.° 2010/0105049. Esos métodos implican, algunas veces, unir un ácido nucleico de muestra a un agente de unión específico de metilación (proteína de unión a metil-CpG (MBD), anticuerpos específicos de metilación y similares) y separar el ácido nucleico unido del ácido nucleico no unido basándose en el estado de metilación diferencial. Dichos métodos pueden incluir además el uso de enzimas de restricción sensibles a la metilación (tal como se ha descrito anteriormente; por ejemplo, Hhal y Hpall), que permiten el enriquecimiento de regiones de ácido nucleico fetal en una muestra materna mediante la digestión selectiva de ácido nucleico de la muestra materna con una enzima que digiere selectiva y completa o sustancialmente el ácido nucleico materno para enriquecer la muestra en al menos una región de ácido nucleico fetal.

Otro método para enriquecer una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que puede usarse con un método descrito en el presente documento es un enfoque de secuencia polimórfica potenciada por endonucleasas de restricción, tal como un método descrito en la publicación de solicitud de patente estadounidense n.° 2009/0317818. Dichos 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 de ácido nucleico no

escindido, pero no de ácido nucleico escindido, donde el ácido nucleico amplificado no escindido representa ácido nucleico diana enriquecido (por ejemplo, ácido nucleico fetal) en relación con ácido nucleico no diana (por ejemplo, ácido nucleico materno). En determinadas implementaciones, el ácido nucleico puede seleccionarse de tal manera que comprenda un alelo que tiene un sitio polimórfico susceptible a la digestión selectiva por medio de un agente de escisión, por ejemplo.

Algunos métodos para enriquecer una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que pueden usarse con un método descrito en el presente documento incluyen enfoques de degradación enzimática selectiva. Tales métodos implican proteger secuencias diana frente a la digestión con exonucleasas, facilitando de ese modo la eliminación en una muestra de secuencias no deseadas (por ejemplo, ADN materno). Por ejemplo, en un enfoque, el ácido nucleico de 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 mediante polimerización de nucleótidos que genera secuencias diana bicatenarias, y que digiere el ácido nucleico monocatenario usando una nucleasa que digiere ácido nucleico monocatenario (es decir, no diana). En determinadas implementaciones, el método puede repetirse durante al menos un ciclo adicional. En determinadas implementaciones, se usa el mismo par de cebadores específicos de diana para cebar cada uno del primer y el segundo ciclos de extensión y, en determinadas implementaciones, se usan pares de cebadores específicos de diana diferentes para el primer y el segundo ciclos.

Algunos métodos para enriquecer una subpoblación de ácido nucleico (por ejemplo, ácido nucleico fetal) que puede usarse con un método descrito en el presente documento incluyen métodos de secuenciación masiva en paralelo de firmas (MPSS). Normalmente, MPSS es un método de fase sólida que usa ligamiento de adaptador (es decir, etiqueta), seguida de decodificación de adaptador, y lectura de la secuencia de ácido nucleico en pequeños incrementos. Los productos de PCR marcados se amplifican normalmente de tal manera que cada ácido nucleico genera un producto de PCR con una etiqueta única. Las etiquetas se usan a menudo para unir los productos de PCR a microperlas. Después de varias tandas de determinación de secuencias basada en ligamiento, por ejemplo, puede identificarse una firma de secuencia a partir de cada perla. Se analiza cada secuencia de firma (etiqueta de MPSS) en un conjunto de datos MPSS, en comparación con todas las demás firmas, y se cuentan todas las firmas idénticas.

En determinadas implementaciones, determinados métodos de enriquecimiento (por ejemplo, determinados métodos de enriquecimiento basados en MPS y/o MPSS) pueden incluir enfoques basados en la amplificación (por ejemplo, PCR). En determinadas implementaciones, pueden usarse métodos de amplificación específicos de locus (por ejemplo, usando cebadores de amplificación específicos de locus). En determinadas implementaciones, puede usarse un enfoque de PCR múltiplex de alelos con SNP. En determinadas implementaciones, puede usarse un enfoque de PCR múltiplex de alelos con SNP en combinación con secuenciación uniplex. Por ejemplo, tal método puede incluir el uso de PCR múltiplex (por ejemplo, el sistema MASSARRAY) y la incorporación de secuencias de sondas de captura en los amplicones seguido por secuenciación usando, por ejemplo, el sistema Illumina de MPSS. En determinadas implementaciones, puede usarse un método de PCR múltiplex de alelos con SNP junto con un sistema de tres cebadores y secuenciación indexada. Por ejemplo, dicho enfoque puede incluir el uso de PCR múltiplex (por ejemplo, sistema MASSARRAY) con cebadores que tienen una primera sonda de captura incorporada en determinados cebadores de PCR directos específicos de locus y secuencias adaptadoras incorporadas en los cebadores de PCR inversos específicos de locus, para generar de ese modo amplicones, seguido por una PCR secundaria para incorporar las secuencias de captura inversa y los códigos de barras de índice molecular para la secuenciación usando, por ejemplo, el sistema Illumina de MPSS. En determinadas implementaciones, puede usarse un método de PCR múltiplex de alelos con SNP junto con un sistema de cuatro cebadores y secuenciación indexada. Por ejemplo, tal enfoque que puede involucrar el uso de PCR múltiplex (por ejemplo, sistema MASSARRAY) con cebadores que tienen secuencias adaptadoras incorporadas en los cebadores de PCR directos específicos de locus e inversos específicos de locus, seguido de una PCR secundaria para incorporar las secuencias de captura directa e inversa y los códigos de barra de índice molecular para la secuenciación usando, por ejemplo, el sistema Illumina de MPSS. En determinadas implementaciones, puede usarse un enfoque microfluídico. En determinadas implementaciones, puede usarse un enfoque microfluídico basado en alineamientos. Por ejemplo, tal método puede incluir el uso de un alineamiento microfluídico (por ejemplo, Fluidigm) para la amplificación a una multiplicidad baja y la incorporación de sondas de índice y captura, seguido de secuenciación. En determinadas implementaciones, puede usarse un método microfluídico en emulsión, tal como, por ejemplo, PCR digital en gotas.

En determinadas implementaciones, pueden usarse métodos de amplificación universales (por ejemplo, usando cebadores de amplificación universales o inespecíficos de locus). En determinadas implementaciones, los métodos de amplificación universales pueden usarse en combinación con enfoques de detección de interacciones (“pull-down”). En determinadas implementaciones, un método puede incluir la detección de interacciones de Ultramer biotinilado (por ejemplo, ensayos de detección de interacciones biotiniladas de Agilent o IDT) a partir de una biblioteca de secuenciación amplificada de manera universal. Por ejemplo, tal método puede implicar la preparación de una biblioteca convencional, el enriquecimiento de regiones seleccionadas mediante un ensayo de detección de interacciones y una etapa secundaria de amplificación universal. En determinadas implementaciones, los enfoques de detección de interacciones pueden usarse en combinación con métodos basados en ligamiento. En determinadas implementaciones, un método puede incluir la detección de interacciones de Ultramer biotinilado con ligamiento de adaptador específico de secuencia (por ejemplo, PCR HALOPLEX, Halo Genomics). Por ejemplo, tal método puede incluir el uso de sondas selectoras para capturar fragmentos digeridos por enzimas de restricción, seguido del ligamiento de productos capturados a un adaptador, y la amplificación universal seguida de secuenciación. En determinadas implementaciones, los enfoques de detección de interacciones pueden usarse en

combinación con métodos basados en extensión y ligamiento. En determinadas implementaciones, un método puede incluir la extensión y ligamiento de la sonda de inversión molecular (MIP). Por ejemplo, tal enfoque puede incluir el uso de sondas de inversión molecular en combinación con adaptadores de secuencia seguido de amplificación y secuenciación universal. En determinadas implementaciones, el ADN complementario puede sintetizarse y secuenciarse sin amplificación.

En determinadas implementaciones, los enfoques de extensión y ligamiento pueden realizarse sin un componente de detección de interacciones. En determinadas implementaciones, un método puede incluir hibridación de cebadores directos e inversos específicos de locus, extensión y ligamiento. Tales métodos pueden incluir además amplificación universal o síntesis de ADN complementario sin amplificación, seguida de secuenciación. Tales métodos pueden reducir o excluir secuencias de fondo durante el análisis, en determinadas implementaciones.

En determinadas implementaciones, los enfoques de detección de interacciones pueden usarse con un componente de amplificación opcional o sin componente de amplificación. En determinadas implementaciones, un método puede incluir un ensayo de detección de interacciones modificado y ligamiento con incorporación completa de sondas de captura sin amplificación universal. Por ejemplo, tal método puede incluir el uso de sondas selectoras modificadas para capturar fragmentos digeridos por enzimas de restricción, seguido por el ligamiento de productos capturados a un adaptador, amplificación opcional y secuenciación. En determinadas implementaciones, un método puede incluir un ensayo de detección de interacciones biotiniladas con extensión y ligamiento de la secuencia adaptadora en combinación con ligamiento monocatenario circular. Por ejemplo, tal enfoque puede incluir el uso de sondas selectoras para capturar regiones de interés (es decir, secuencias diana), extensión de las sondas, ligamiento de adaptador, ligamiento monocatenario circular, amplificación opcional y secuenciación. En determinadas implementaciones, el análisis del resultado de la secuenciación puede separar las secuencias diana del fondo.

En algunas implementaciones, el ácido nucleico se enriquece en fragmentos de una región genómica seleccionada (por ejemplo, cromosoma) por medio del uso de uno o más métodos de separación basados en secuencia descritos en el presente documento. La separación basada en secuencia se basa generalmente en las secuencias de nucleótidos presentes en los fragmentos de interés (por ejemplo, fragmentos diana y/o de referencia) y sustancialmente no presentes en otros fragmentos de la muestra o presentes en una cantidad insustancial de los otros fragmentos (por ejemplo, el 5 % o menos). En algunas implementaciones, la separación basada en secuencia puede generar fragmentos diana separados y/o fragmentos de referencia separados. Los fragmentos diana separados y/o fragmentos de referencia separados, se suelen aislar de los fragmentos restantes en la muestra de ácido nucleico. En determinadas implementaciones, los fragmentos diana separados y los fragmentos de referencia separados se aíslan además entre sí (por ejemplo, se aíslan en compartimentos de ensayo separados). En determinadas implementaciones, los fragmentos diana separados y los fragmentos de referencia separados se aíslan juntos (por ejemplo, se aíslan en el mismo compartimento de ensayo). En algunas implementaciones, los fragmentos no unidos pueden retirarse diferencialmente, degradarse o digerirse.

En algunas implementaciones, se usa un procedimiento de captura de ácido nucleico selectivo para separar los fragmentos diana y/o de referencia de la muestra de ácido nucleico. Los sistemas de captura de ácidos nucleicos disponibles en el mercado incluyen, por ejemplo, el sistema de captura de secuencias Nimblegen (Roche NimbleGen, Madison, WI); la plataforma BEADARRAY de Illumina (Illumina, San Diego, CA); la plataforma GENECHIP de Affymetrix (Affymetrix, Santa Clara, CA); el sistema de enriquecimiento de diaans SureSelect de Agilent (Agilent Technologies, Santa Clara, CA); y plataformas relacionadas. Tales métodos implican normalmente la hibridación de un oligonucleótido de captura a un segmento o toda la secuencia de nucleótidos de un fragmento diana o de referencia y pueden incluir el uso de una fase sólida (por ejemplo, alineamiento de fase sólida) y/o una plataforma a base de disolución. Los oligonucleótidos de captura (a veces denominados “cebo”) pueden seleccionarse o diseñarse de tal manera que se hibriden preferiblemente con fragmentos de ácido nucleico de regiones o locus genómicos seleccionados (por ejemplo, uno de los cromosomas 21, 18, 13, X o Y, o un cromosoma de referencia). En determinadas implementaciones, puede utilizarse un método basado en la hibridación (por ejemplo, utilizando matrices de oligonucleótidos) para enriquecer en las secuencias de ácidos nucleicos de determinados cromosomas (por ejemplo, un cromosoma potencialmente aneuploide, un cromosoma de referencia u otro cromosoma de interés) o segmentos de interés de los mismos.

En algunas implementaciones, el ácido nucleico se enriquece en una longitud particular de fragmento de ácido nucleico, rango de longitudes o longitudes por debajo o por encima de un umbral o punto de corte particular usando uno o más métodos de separación basados en la longitud. La longitud del fragmento de ácido nucleico se refiere normalmente al número de nucleótidos en el fragmento. La longitud del fragmento de ácido nucleico se denomina además algunas veces tamaño del fragmento de ácido nucleico. En algunas implementaciones, se realiza un método de separación basado en la longitud sin medir las longitudes de los fragmentos individuales. En algunas implementaciones, se realiza un método de separación basado en la longitud junto con un método para determinar la longitud de los fragmentos individuales. En algunas implementaciones, la separación basada en longitud se refiere a un procedimiento de fraccionamiento por tamaño, en el que la totalidad o parte de la combinación fraccionada puede aislarse (por ejemplo, retenerse) y/o analizarse. Los procedimientos de fraccionamiento por tamaños se conocen en la técnica (por ejemplo, separación en un alineamiento, separación mediante un tamiz molecular, separación por electroforesis en gel, separación por cromatografía en columna (por ejemplo, columnas de exclusión molecular) y enfoques basados en microfluídica). En determinadas implementaciones, los métodos de separación basados en la longitud pueden incluir circularización de fragmentos, tratamiento químico (por ejemplo, con formaldehído, polietilenglicol (PEG)), espectrometría de masas y/o amplificación de ácido nucleico específica de tamaño, por ejemplo.

Determinados métodos de separación basados en la longitud que pueden usarse con los métodos descritos en el presente documento usan un enfoque de marcaje con etiqueta de secuencia selectivo, por ejemplo. La expresión "marcaje con etiqueta de secuencia" se refiere a incorporar una secuencia reconocible y distinta en un ácido nucleico o una población de ácidos nucleicos. La expresión "marcaje con etiqueta de secuencia", tal como se usa en el presente documento, tiene un significado diferente a la expresión "etiqueta de secuencia" que se describe más adelante en el presente documento. En tales métodos de marcaje con etiqueta de secuencia, ácidos nucleicos de una especie de tamaño de fragmento (por ejemplo, fragmentos cortos) se someten a marcaje con etiqueta selectivo de secuencia en una muestra que incluye ácidos nucleicos largos y cortos. Tales métodos implican normalmente llevar a cabo una reacción de amplificación de ácido nucleico usando un conjunto de cebadores anidados que incluyen cebadores internos y cebadores externos. En determinadas implementaciones, uno o ambos interiores pueden marcarse con una etiqueta para introducir de ese modo una etiqueta sobre el producto de amplificación diana. Generalmente, los cebadores externos no se hibridan con los fragmentos cortos que portan la secuencia diana (interna). Los cebadores internos pueden hibridarse con los fragmentos cortos y generar un producto de amplificación que porta una etiqueta y la secuencia diana. Normalmente, el marcaje con etiqueta de los fragmentos largos se inhibe a través de una combinación de mecanismos que incluyen, por ejemplo, extensión bloqueada de los cebadores internos por la hibridación y extensión previas de los cebadores externos. El enriquecimiento en fragmentos marcados puede lograrse mediante cualquiera de una gran variedad de métodos incluyendo, por ejemplo, digestión con exonucleasa de ácido nucleico monocatenario y amplificación de los fragmentos marcados con etiqueta usando cebadores de amplificación específicos para al menos una etiqueta.

Otro método de separación basado en la longitud que puede usarse con los métodos descritos en el presente documento implica someter una muestra de ácido nucleico a precipitación con polietilenglicol (PEG). Los ejemplos de métodos incluyen los descritos en las publicaciones de solicitud de patente internacional n.ºs WO2007/140417 y WO2010/115016. Este método implica generalmente poner en contacto una muestra de ácido nucleico 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, de menos de 300 nucleótidos).

Otro método de enriquecimiento basado en tamaño que puede usarse con los métodos descritos en el presente documento implica la circularización mediante ligamiento, por ejemplo, con el uso de CircLigase. Los fragmentos de ácido nucleico cortos pueden circularizarse normalmente con una mayor eficiencia que los fragmentos largos. Las secuencias no circularizadas pueden separarse de las secuencias circularizadas, y los fragmentos cortos enriquecidos pueden usarse para un análisis adicional.

Biblioteca de ácidos nucleicos

En algunas implementaciones, 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 limitativos 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 determinadas implementaciones, se prepara una biblioteca de ácidos nucleicos antes o durante un proceso de secuenciación. Una biblioteca de ácidos nucleicos (por ejemplo, una biblioteca de secuenciación) se puede preparar mediante un método adecuado como se conoce en la técnica. Una biblioteca de ácidos nucleicos se puede preparar mediante un proceso de preparación dirigido o no dirigido.

En algunas implementaciones, 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 algunas implementaciones, 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 limitativos 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, componente del complemento C1q, proteínas de unión a ácidos nucleicos, receptores, carbohidratos, 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 epítopo antigénico y un anticuerpo o fragmento inmunológicamente reactivo del mismo; un anticuerpo y un hapteno; un resto de digoxigeno y un anticuerpo anti-digoxigeno; 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; un oligonucleótido o polinucleótido y su complemento correspondiente; similares o combinaciones de los mismos.

En algunas implementaciones, una biblioteca de ácidos nucleicos se modifica para que comprenda uno o más polinucleótidos de composición conocida, cuyos ejemplos no limitativos incluyen un identificador (por ejemplo, una etiqueta, una etiqueta de indexación), una secuencia de captura, un marcador, un adaptador, un sitio de enzima de restricción, un promotor, un potenciador, un origen de replicación, una estructura de horquilla, una secuencia complementaria (por ejemplo, un sitio de unión del cebador, un sitio de hibridación), un sitio de integración adecuado (por

ejemplo, un transposón, un sitio de integración vírica), un nucleótido modificado, similares o combinaciones de los mismos. Se pueden añadir polinucleótidos de secuencia conocida en una posición adecuada, por ejemplo, en el extremo 5', el extremo 3' o dentro de una secuencia de ácido nucleico. Los polinucleótidos de secuencia conocida pueden ser secuencias iguales o diferentes. En algunas implementaciones, un polinucleótido de secuencia conocida se configura para hibridarse 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 conocida en 5' puede hibridarse con una primera pluralidad de oligonucleótidos mientras que la secuencia conocida en 3' puede hibridarse con una segunda pluralidad de oligonucleótidos. En algunas implementaciones, una biblioteca de ácido nucleico puede comprender etiquetas específicas de cromosomas, secuencias de captura, marcadores y/o adaptadores. En algunas implementaciones, una biblioteca de ácidos nucleicos comprende uno o más marcadores detectables. En algunas implementaciones, se pueden incorporar uno o más marcadores detectables a 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 algunas implementaciones, una biblioteca de ácidos nucleicos comprende oligonucleótidos hibridados. En determinadas implementaciones, los oligonucleótidos hibridados son sondas marcadas. En algunas implementaciones, una biblioteca de ácidos nucleicos comprende sondas de oligonucleótidos hibridadas antes de la inmovilización en una fase sólida.

En algunas implementaciones, un polinucleótido de secuencia conocida comprende una secuencia universal. Una secuencia universal es una secuencia de ácido nucleico 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 se integra. Una secuencia universal se suele diseñar para hibridar y/o amplificar una pluralidad de secuencias diferentes utilizando un único cebador universal que sea complementario a una secuencia universal. En algunas implementaciones, se utilizan dos (por ejemplo, un par) o más secuencias universales y/o cebadores universales. Un cebador universal suele comprender una secuencia universal. En algunas implementaciones, los adaptadores (por ejemplo, adaptadores universales) comprenden secuencias universales. En algunas implementaciones se utilizan una o más secuencias universales para capturar, identificar y/o detectar múltiples especies o subconjuntos de ácidos nucleicos.

En determinadas implementaciones de la preparación de una biblioteca de ácidos nucleicos (por ejemplo, en determinados procedimientos de secuenciación mediante síntesis), los ácidos nucleicos se seleccionan según el tamaño y/o se fragmentan en longitudes de varios cientos de pares de bases o menores (por ejemplo, en la preparación para la generación de bibliotecas). En algunas implementaciones, la preparación de la biblioteca se realiza sin fragmentación (por ejemplo, cuando se usa ADNfcc).

En determinadas implementaciones, se usa un método de preparación de bibliotecas basado en ligamiento (por ejemplo, ILLUMINA TRUSEQ, Illumina, San Diego CA). Los métodos de preparación de bibliotecas basados en ligamiento suelen utilizar un diseño (por ejemplo, un adaptador metilado) que puede incorporar una secuencia índice en la etapa de ligamiento inicial y a menudo pueden usarse para preparar muestras para secuenciación de una sola lectura, secuenciación de extremos apareados y secuenciación multiplexada. Por ejemplo, a veces los ácidos nucleicos (por ejemplo, ácidos nucleicos fragmentados o ADNfcc) se reparan en los extremos mediante una reacción de relleno, una reacción de exonucleasa o una combinación de las mismas. En algunas implementaciones, el ácido nucleico reparado de extremo romo resultante puede extenderse luego por un solo nucleótido que sea complementario a un solo nucleótido saliente en el extremo 3' de un adaptador/cebador. Puede usarse cualquier nucleótido para los nucleótidos de extensión/saliente. En algunas implementaciones, la preparación de la biblioteca de ácidos nucleicos comprende ligar un oligonucleótido adaptador. Los oligonucleótidos adaptadores suelen ser 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 algunas implementaciones, un oligonucleótido adaptador comprende un identificador, uno o más sitios de hibridación de cebadores de secuenciación (por ejemplo, secuencias complementarias a los cebadores de secuenciación universales, cebadores de secuenciación de un solo extremo, cebadores de secuenciación de extremos apareados, cebadores de secuenciación multiplexados, y similares), o combinaciones de los mismos (por ejemplo, adaptador/secuenciación, adaptador/identificador, adaptador/identificador/secuenciación).

Un identificador puede ser un marcador detectable adecuado incorporado o unido a un ácido nucleico (por ejemplo, un polinucleótido) que permita la detección y/o identificación de ácidos nucleicos que comprendan el identificador. En algunas implementaciones, un identificador se incorpora o se une a un ácido nucleico durante un método de secuenciación (por ejemplo, mediante una polimerasa). Los ejemplos no limitativos de identificadores incluyen etiquetas de ácido nucleico, índices de ácido nucleico o códigos de barras, un marcador radioactivo (por ejemplo, un isótopo), un marcador metálico, un marcador fluorescente, un marcador quimioluminiscente, n marcador fosforescente, un inactivador de fluoróforo, un colorante, una proteína (por ejemplo, una enzima, un anticuerpo o parte del mismo, un conector, un miembro de un par de unión), similares o combinaciones de los mismos. En algunas implementaciones, un identificador (por ejemplo, un índice de ácido nucleico o un código de barras) es una secuencia única, conocida y/o identificable de nucleótidos o análogos de nucleótidos. En algunas implementaciones, los identificadores son seis o más nucleótidos contiguos. Existe una multitud de fluoróforos con una variedad de diferentes espectros de excitación y emisión. Se puede usar cualquier tipo y/o número adecuado de fluoróforos como identificador. En algunas implementaciones, se usan 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 algunas implementaciones, uno o dos

tipos de identificadores (por ejemplo, marcadores fluorescentes) están vinculados a cada ácido nucleico de una biblioteca. La detección y/o cuantificación de un identificador puede realizarse mediante un método, una máquina o un aparato adecuado, cuyos ejemplos no limitativos incluyen citometría de flujo, reacción en cadena de la polimerasa cuantitativa (qPCR), electroforesis en gel, un luminómetro, un fluorómetro, un espectrofotómetro, un análisis de genochips o micromatrices adecuado, transferencia Western, espectrometría de masas, cromatografía, análisis citofluorimétrico, microscopía de fluorescencia, un método de generación de imágenes digitales o de fluorescencia adecuado, microscopía de barrido 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 un aparato de secuenciación de ácidos nucleicos, similares y combinaciones de los mismos.

En algunas implementaciones, se usa un método de preparación de bibliotecas basado en transposones (por ejemplo, EPICENTRE NEXTERA, Epicentre, Madison WI). Los métodos basados en transposones usan normalmente la transposición *in vitro* para fragmentar y marcar con etiqueta simultáneamente ADN en una reacción de un solo tubo (que a menudo permite la incorporación de etiquetas específicas de plataforma y códigos de barras opcionales), y preparar bibliotecas listas para el secuenciador.

En algunas implementaciones, se amplifica una biblioteca de ácidos nucleicos o partes de la misma (por ejemplo, se amplifica mediante un método basado en PCR). En algunas implementaciones, un método de secuenciación comprende la amplificación de una biblioteca de ácidos nucleicos. Una biblioteca de ácidos nucleicos se puede amplificar 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 amplificación o aumento del número de un molde de ácido nucleico y/o de un complemento del mismo que están presentes (por ejemplo, en una biblioteca de ácidos nucleicos), mediante la producción de una o más copias del molde y/o su complemento. La amplificación puede llevarse a cabo mediante un método adecuado. Una biblioteca de ácidos nucleicos puede amplificarse mediante un método de termociclado o mediante un método de amplificación isotérmico. En algunas implementaciones, se utiliza un método de amplificación de círculo rodante. En algunas implementaciones, 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 determinados métodos de secuenciación, se añade una biblioteca de ácidos nucleicos a una celda de flujo y se inmoviliza mediante hibridación con anclajes en condiciones adecuadas. Este tipo de amplificación de ácidos nucleicos se suele denominar amplificación en fase sólida. En algunas implementaciones de la amplificación en fase sólida, la totalidad o una porción de los productos amplificados se sintetizan mediante una extensión que se inicia a partir de un cebador inmovilizado. Las reacciones de amplificación en fase sólida son análogas a las amplificaciones en fase de solución convencionales excepto que al menos uno de los oligonucleótidos de amplificación (por ejemplo, cebadores) se inmoviliza en un soporte sólido.

En algunas implementaciones, la amplificación en fase sólida comprende una reacción de amplificación de ácido nucleico que comprende solo una especie de cebador oligonucleotídico inmovilizado en una superficie. En determinadas implementaciones, la amplificación en fase sólida comprende una pluralidad de diferentes especies de cebadores oligonucleotídicos inmovilizados. En algunas implementaciones, la amplificación en fase sólida puede comprender una reacción de amplificación de ácido nucleico que comprende una especie de cebador oligonucleotídico inmovilizado sobre una superficie sólida y una segunda especie de cebador oligonucleotídico diferente en solución. Se pueden usar múltiples especies diferentes de cebadores inmovilizados o basados en solución. Los ejemplos no limitativos 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 estadounidense US20130012399), similares o combinaciones de las mismas.

Secuenciación

En algunas implementaciones, se secuencian ácidos nucleicos (por ejemplo, fragmentos de ácido nucleico, ácido nucleico de muestra, ácido nucleico libre de células). En determinadas implementaciones, se obtiene una secuencia completa o sustancialmente completa y, algunas veces, se obtiene una secuencia parcial.

En algunas implementaciones, algunos o todos los ácidos nucleicos de una muestra se enriquecen y/o amplifican (por ejemplo, de forma no específica, por ejemplo, mediante un método basado en PCR) antes o durante la secuenciación. En determinadas implementaciones, las porciones o subconjuntos de ácidos nucleicos específicos de una muestra se enriquecen y/o amplifican antes o durante la secuenciación. En algunas implementaciones, una porción o un subconjunto de una combinación preseleccionada de ácidos nucleicos se secuencia aleatoriamente. En algunas implementaciones, los ácidos nucleicos de una muestra no se enriquecen y/o amplifican antes o durante la secuenciación.

Tal como se usa en el presente documento, "lecturas" (es decir, "una lectura", "una lectura de secuencia") son secuencias de nucleótidos cortas producidas mediante cualquier procedimiento de secuenciación descrito en el presente documento o conocido en la técnica. Las lecturas pueden generarse a partir de un extremo de los fragmentos de ácido nucleico ("lecturas de un solo extremo") y, algunas veces, se generan a partir de ambos extremos de los ácidos nucleicos (por ejemplo, lecturas de extremos apareados, lecturas de doble extremo).

La longitud de una lectura de secuencia se asocia a menudo con la tecnología de secuenciación 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 por nanoporos, por ejemplo, puede proporcionar lecturas de secuencia que pueden variar en tamaño desde decenas hasta cientos y hasta miles de pares de bases. En algunas implementaciones, las lecturas de secuencia tienen una longitud media, promedio o absoluta, o mediana de la longitud, de aproximadamente 15 pb a aproximadamente 900 pb de longitud. En determinadas implementaciones, las lecturas de secuencia tienen una longitud media, promedio o absoluta, o mediana de la longitud, de aproximadamente 1000 pb o más.

En determinadas implementaciones, la longitud nominal, promedio, media o absoluta de las lecturas de un solo extremo es a veces de aproximadamente 15 nucleótidos contiguos a aproximadamente 50 nucleótidos contiguos o más, de aproximadamente 15 nucleótidos contiguos a aproximadamente 40 nucleótidos contiguos o más, y a veces de aproximadamente 15 nucleótidos contiguos a aproximadamente 36 nucleótidos contiguos o más. En determinadas implementaciones, la longitud nominal, promedio, media o absoluta de las lecturas de un solo extremo es de aproximadamente 20 a aproximadamente 30 bases, o de aproximadamente 24 a aproximadamente 28 bases de longitud. En determinadas implementaciones, 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 o aproximadamente 29 bases o más de longitud.

En determinadas implementaciones, la longitud nominal, promedio, media o absoluta de las lecturas de extremos apareados a veces es de aproximadamente 10 nucleótidos contiguos a aproximadamente 25 nucleótidos contiguos o más (por ejemplo, de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 nucleótidos de longitud o más), de aproximadamente 15 nucleótidos contiguos a aproximadamente 20 nucleótidos contiguos o más, y a veces es de aproximadamente 17 nucleótidos contiguos o aproximadamente 18 nucleótidos contiguos.

Las lecturas son generalmente 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 adenina, "T" representa un nucleótido timina, "G" representa un nucleótido guanina y "C" representa un nucleótido citosina, en un ácido nucleico físico. Las lecturas de secuencia obtenidas a partir de la sangre de una mujer embarazada pueden ser lecturas de una mezcla de ácido nucleico fetal y materno. Una mezcla de lecturas relativamente cortas puede transformarse mediante procedimientos descritos en el presente documento en una representación de un ácido nucleico genómico presente en la mujer embarazada y/o en el feto. Una mezcla de lecturas relativamente cortas puede transformarse en una representación de una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal), una variación genética o una aneuploidía, por ejemplo. Las lecturas de una mezcla de ácido nucleico materno y fetal pueden transformarse en una representación de un cromosoma compuesto o un segmento del mismo que comprende características de uno o ambos cromosomas maternos y fetales. En determinadas implementaciones, "obtener" lecturas de secuencia de ácido nucleico de una muestra de un sujeto y/u "obtener" lecturas de secuencia de ácido nucleico de un espécimen biológico de una o más personas de referencia pueden implicar la secuenciación directa del ácido nucleico para obtener la información de secuencia. En algunas implementaciones, "obtener" puede implicar recibir información de secuencia obtenida directamente de un ácido nucleico por otro.

En algunas implementaciones, se secuencia una fracción representativa de un genoma y, a veces, se denomina "cobertura" o "cobertura de pliegues". Por ejemplo, una cobertura de 1 vez indica que aproximadamente el 100 % de las secuencias de nucleótidos del genoma están representadas por lecturas. En algunas implementaciones, "las veces de cobertura" es una expresión relativa que se refiere a una ejecución de secuenciación anterior como referencia. Por ejemplo, una segunda ejecución de secuenciación puede tener la mitad de cobertura que una primera ejecución de secuenciación. En algunas implementaciones, se secuencia un genoma con redundancia, donde una región dada del genoma puede estar cubierta por dos o más lecturas o lecturas superpuestas (por ejemplo, "veces de cobertura" mayor de 1, por ejemplo, 2 veces de cobertura).

En algunas implementaciones, se secuencia una muestra de ácido nucleico de un individuo. En determinadas implementaciones, se secuencian los ácidos nucleicos de cada una de dos o más muestras, donde las muestras son de un individuo o de diferentes individuos. En determinadas implementaciones, las muestras de ácido nucleico de dos o más muestras biológicas se combinan, donde cada muestra biológica es de un individuo o de dos o más individuos, y se secuencia la combinación. En estas últimas implementaciones, una muestra de ácido nucleico de cada muestra biológica suele identificarse mediante uno o más identificadores únicos.

En algunas implementaciones, un método de secuenciación utiliza identificadores que permiten la multiplexación de 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 puedan multiplexarse en un proceso de secuenciación. Un proceso de secuenciación puede realizarse utilizando cualquier número adecuado de 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 la que se puede unir el ácido nucleico de una biblioteca y los reactivos pueden fluir y ponerse en contacto con el ácido nucleico anexo. Una celda de flujo a veces incluye carriles de celdas de flujo, y el uso de identificadores puede facilitar el análisis de varias muestras en cada carril. Una celda de flujo es a menudo un soporte sólido que puede configurarse para retener y/o permitir el paso ordenado de disoluciones de reactivos sobre analitos unidos. Las celdas de flujo tienen a menudo forma plana, son ópticamente transparentes, generalmente en la escala

milimétrica o submilimétrica, y tienen a menudo canales o carriles en los que se produce la interacción analito/reactivo. En algunas implementaciones, el número de muestras analizadas en un carril de celda de flujo determinado depende del número de identificadores únicos utilizados durante la preparación de la biblioteca y/o el carril de celda de flujo único del diseño de la sonda. La multiplexación usando 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 usando 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) en una celda de flujo de 8 carriles. Los ejemplos no limitativos de kits de secuenciación múltiplex disponibles comercialmente incluyen el kit de oligonucleótidos para la preparación de muestras de multiplexación de Illumina y los cebadores de secuenciación de multiplexación y el kit de control PhiX (por ejemplo, números de catálogo de Illumina PE-400-1001 y PE-400-1002, respectivamente).

Puede usarse cualquier método adecuado de secuenciación de ácidos nucleicos, cuyos ejemplos no limitativos incluyen Maxim & Gilbert, métodos de terminación de cadena, secuenciación por síntesis, secuenciación por ligación, secuenciación por espectrometría de masas, técnicas basadas en microscopía, similares o combinaciones de los mismos. En algunas implementaciones, 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 secuenciación de Sanger microfluídica, en un método proporcionado en el presente documento. En algunas implementaciones, se pueden utilizar tecnologías de secuenciación que incluyan el uso de tecnologías de obtención de imágenes de ácidos nucleicos (por ejemplo, microscopía electrónica de transmisión (TEM) y microscopía de fuerza atómica (AFM)). En algunas implementaciones, se usa un método de secuenciación de alto rendimiento. Los métodos de secuenciación de alto rendimiento generalmente incluyen moldes de ADN amplificados clonalmente o moléculas únicas de ADN que se secuencian de forma masiva en paralelo, a veces dentro de una celda de flujo. Las técnicas de secuenciación de próxima generación (por ejemplo, 2.ª y 3.ª generación) capaces de secuenciar el ADN de forma masiva en paralelo se pueden utilizar para los métodos descritos en el presente documento y se denominan colectivamente en el presente documento "secuenciación masiva en paralelo" (MPS). En algunas implementaciones, los métodos de secuenciación MPS utilizan un enfoque dirigido, donde se secuencian cromosomas, genes o regiones de interés específicos. En determinadas implementaciones, se utiliza un enfoque no dirigido donde la mayoría o todos los ácidos nucleicos de una muestra se secuencian, amplifican y/o capturan aleatoriamente.

En algunas implementaciones se utiliza un enfoque de enriquecimiento, amplificación y/o secuenciación dirigido. Un enfoque dirigido suele aislar, seleccionar y/o enriquecer un subconjunto de ácidos nucleicos de una muestra para su posterior procesamiento mediante el uso de oligonucleótidos específicos de la secuencia. En algunas implementaciones, se utiliza una biblioteca de oligonucleótidos específicos de la secuencia para dirigirse a (por ejemplo, hibridares con) uno o más conjuntos de ácidos nucleicos de una muestra. Los oligonucleótidos y/o cebadores específicos de la secuencia suelen ser selectivos para secuencias particulares (por ejemplo, secuencias de ácido nucleico únicas) presentes en uno o más cromosomas, genes, exones, intrones y/o regiones reguladoras de interés. Puede usarse cualquier método adecuado o combinación de métodos para el enriquecimiento, amplificación y/o secuenciación de uno o más subconjuntos de ácidos nucleicos diana. En algunas implementaciones, las secuencias diana se aíslan y/o enriquecen mediante la captura en una fase sólida (por ejemplo, una celda de flujo, una perla) utilizando uno o más anclajes específicos de la secuencia. En algunas implementaciones, las secuencias diana 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 conjuntos de cebadores específicos de la secuencia. Los anclajes específicos de la secuencia se suelen poder utilizar como cebadores específicos de la secuencia.

La secuenciación MPS a veces hace uso de la secuenciación por síntesis y determinados procesos de formación de imágenes. Una tecnología de secuenciación de ácidos nucleicos que puede usarse en un método descrito en el presente documento es secuenciación por síntesis y secuenciación basada en el terminador reversible (por ejemplo, el analizador genómico de Illumina; analizador genómico II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA)). Con esta tecnología, millones de fragmentos de ácido nucleico (por ejemplo, ADN) pueden secuenciarse 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 sobre cuyas superficies hay anclajes de oligonucleótidos unidos (por ejemplo, cebadores adaptadores). Una celda de flujo es a menudo un soporte sólido que puede configurarse para retener y/o permitir el paso ordenado de disoluciones de reactivos sobre analitos unidos. Las celdas de flujo tienen a menudo forma plana, son ópticamente transparentes, generalmente en la escala milimétrica o submilimétrica, y tienen a menudo canales o carriles en los que se produce la interacción analito/reactivo.

La secuenciación por síntesis, en algunas implementaciones, comprende la adición iterativa (por ejemplo, por adición covalente) de un nucleótido a un cebador o cadena de ácido nucleico preexistente de 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 cadena de ácido nucleico. La longitud de una secuencia obtenida depende, en parte, del número de etapas de adición y detección que se realicen. En algunas implementaciones de secuenciación por síntesis, se añaden y detectan uno, dos, tres o más nucleótidos del mismo tipo (por ejemplo, A, G, C o T) en una serie de adición de nucleótidos. Los nucleótidos se pueden añadir mediante cualquier método adecuado (por ejemplo, enzimática o químicamente). Por ejemplo, en algunas implementaciones, una polimerasa o una ligasa añade un nucleótido a un cebador o a una cadena de ácido nucleico preexistente de manera dirigida por el molde. En algunas implementaciones de secuenciación por síntesis, se utilizan

diferentes tipos de nucleótidos, análogos de nucleótidos y/o identificadores. En algunas implementaciones se utilizan terminadores reversibles y/o identificadores extraíbles (por ejemplo, escindibles). En algunas implementaciones se utilizan nucleótidos y/o análogos de nucleótidos marcados con fluorescencia. En determinadas implementaciones, la secuenciación por síntesis comprende una escisión (por ejemplo, escisión y extracción de un identificador) y/o una etapa de lavado. En algunas implementaciones, 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 limitativos incluyen cualquier aparato de formación de imágenes adecuado, una cámara adecuada, una cámara digital, un CCD (Dispositivo de par de carga) (por ejemplo, una cámara CCD), un aparato de formación de imágenes basado en CMOS (silicio de óxido de metal complementario) (por ejemplo, una cámara CMOS), un fotodiodo (por ejemplo, un tubo fotomultiplicador), microscopía electrónica, un transistor de efecto de campo (por ejemplo, un transistor de efecto de campo de ADN), un sensor de iones ISFET (por ejemplo, un sensor CHEMFET), similares o combinaciones de los mismos. Otros métodos de secuenciación que pueden usarse para llevar a cabo los métodos en el presente documento incluyen PCR digital y secuenciación por hibridación.

Otros métodos de secuenciación que pueden usarse para llevar a cabo los métodos en el presente documento incluyen PCR digital y secuenciación por hibridación. La reacción en cadena de la polimerasa digital (PCR digital o dPCR) puede usarse para identificar y cuantificar directamente los ácidos nucleicos en una muestra. La PCR digital puede llevarse a cabo en una emulsión, en algunas implementaciones. Por ejemplo, los ácidos nucleicos individuales se separan, por ejemplo, en un dispositivo de cámara microfluídico, y cada ácido nucleico se amplifica individualmente mediante PCR. Los ácidos nucleicos pueden separarse de tal manera que no haya más de un ácido nucleico por pocillo. En algunas implementaciones, pueden usarse sondas diferentes para distinguir diversos alelos (por ejemplo, alelos fetales y alelos maternos). Pueden enumerarse alelos para determinar el número de copias.

En determinadas implementaciones, se puede utilizar la secuenciación por hibridación. El método incluye poner en contacto una pluralidad de secuencias de polinucleótidos con una pluralidad de sondas de polinucleótidos, en donde cada una de la pluralidad de sondas de polinucleótidos puede anclarse, opcionalmente, a un sustrato. El sustrato puede ser una superficie plana con un alineamiento de secuencias de nucleótidos conocidas, en algunas implementaciones. El patrón de hibridación al alineamiento puede usarse para determinar las secuencias de polinucleótidos presentes en la muestra. En algunas implementaciones, cada sonda se conecta 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.

En algunas implementaciones, la secuenciación por nanoporos puede usarse en un método descrito en el presente documento. La secuenciación por nanoporos es una tecnología de secuenciación de una sola molécula mediante la cual una sola molécula de ácido nucleico (por ejemplo, ADN) se secuencia directamente a medida que pasa a través de un nanoporo.

Se puede utilizar un método, sistema o plataforma tecnológica de MPS adecuado para llevar a cabo los métodos descritos en el presente documento para obtener lecturas de secuenciación de ácidos nucleicos. Los ejemplos no limitativos de plataformas MPS incluyen Illumina/Solex/HiSeq (por ejemplo, el analizador genómico de Illumina; analizador genómico II; HISEQ 2000; HISEQ), SOLiD, Roche/454, PACBIO y/o SMRT, Helicos True Single Molecule Sequencing, Ion Torrent y la secuenciación basada en el semiconductor Ion (por ejemplo, desarrollado por Life Technologies), WildFire, 5500, 5500xl W y/o tecnologías basadas en el analizador genético 5500xl W (por ejemplo, desarrolladas y comercializadas por Life Technologies, publicación de patente de EE. UU. n.° US20130012399); Secuenciación de Polonia, pirosecuenciación, secuenciación masiva en paralelo (MPSS), secuenciación de ARN polimerasa (RNAP), sistemas y métodos LaserGen, plataformas basadas en nanoporos, matriz de transistores de efecto de campo sensible a productos químicos (CHEMFET), secuenciación basada en microscopía electrónica (por ejemplo, como la desarrollada por ZS Genetics, Halcyon Molecular), secuenciación de nanobolas,

En algunas implementaciones, se realiza la secuenciación específica de cromosoma. En algunas implementaciones, la secuenciación específica de cromosoma se realiza utilizando DANSR (análisis digital de regiones seleccionadas). El análisis digital de regiones seleccionadas permite la cuantificación simultánea de cientos de locus mediante catenación dependiente de ADNlc de dos oligonucleótidos específicos de locus por medio de un oligonucleótido intermedio "puente" para formar un molde para PCR. En algunas implementaciones, la secuenciación específica de cromosoma se realiza mediante la generación de una biblioteca enriquecida en secuencias específicas de cromosoma. En algunas implementaciones, las lecturas de secuencia se obtienen solamente para un conjunto seleccionado de cromosomas. En algunas implementaciones, las lecturas de secuencia se obtienen solamente para los cromosomas 21, 18 y 13.

Lecturas de mapeo

Las lecturas de secuencia pueden mapearse y la cantidad de lecturas que se mapean en una región específica de ácido nucleico (por ejemplo, un cromosoma, una porción o un segmento del mismo) se denominan recuentos. Puede usarse cualquier método de mapeo adecuado (por ejemplo, proceso, algoritmo, programa, *software*, módulo, similar o una combinación de los mismos). A continuación, se describen determinados aspectos de los procesos de mapeo.

El mapeo de las lecturas de secuencia de nucleótidos (es decir, la información de la secuencia de un fragmento cuya posición genómica física se desconoce) puede realizarse de varias maneras, y suele comprender la alineación de las lecturas de secuencia obtenidas con una secuencia coincidente en un genoma de referencia. En dichas

alineaciones, las lecturas de secuencia generalmente se alinean con una secuencia de referencia y las que se alinean se designan como "mapeadas", "una lectura de secuencia mapeada" o "una lectura mapeada". En determinadas implementaciones, una lectura de secuencia mapeada se denomina "coincidencia" o "recuento". En algunas implementaciones, las lecturas de secuencia mapeadas se agrupan según diversos parámetros y se asignan a porciones concretas, que se tratan con más detalle a continuación.

Tal como se usan en el presente documento, los términos "alineado", "alineación" o "alinear" se refieren a dos o más secuencias de ácido nucleico que pueden identificarse como apareamiento (por ejemplo, identidad del 100 %) o apareamiento parcial. Las alineaciones pueden realizarse manualmente o mediante un ordenador (por ejemplo, un *software*, programa, módulo o algoritmo), entre cuyos ejemplos no limitativos se incluye el programa informático de alineación local eficiente de datos de nucleótidos (ELAND) distribuido como parte del flujo de trabajo de análisis genómico de Illumina. La alineación de una lectura de secuencia puede tener un 100 % de apareamiento de secuencia. En algunos casos, una alineación es inferior al 100 % de apareamiento de secuencias (es decir, apareamiento no perfecto, apareamiento parcial, alineación parcial). En algunas implementaciones, una alineación es de aproximadamente el 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 apareamiento. En algunas implementaciones, una alineación comprende un apareamiento erróneo. En algunas implementaciones, una alineación comprende 1, 2, 3, 4 o 5 apareamientos erróneos. Dos o más secuencias pueden alinearse usando cualquier hebra. En determinadas implementaciones, una secuencia de ácido nucleico se alinea con el complemento inverso de otra secuencia de ácido nucleico.

Pueden usarse diversos métodos computacionales para mapear cada secuencia leída en una porción o una sección genómica. Los ejemplos no limitativos de algoritmos informáticos que pueden usarse 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 algunas implementaciones, las lecturas de secuencia pueden alinearse con las secuencias en un genoma de referencia. En algunas implementaciones, las lecturas de secuencia pueden encontrarse y/o alinearse 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). Puede usarse Blast o herramientas similares para buscar las secuencias identificadas en una base de datos de secuencias. A continuación, pueden usarse las coincidencias de búsqueda para clasificar las secuencias identificadas en porciones o secciones genómicas adecuadas (descritas más adelante en el presente documento), por ejemplo.

En algunas implementaciones, una lectura puede mapearse de manera única o no única en porciones en un genoma de referencia. Se considera que una lectura se "mapea de manera única" si se alinea con una sola secuencia en el genoma de referencia. Se considera que una lectura se "mapea de manera no única" si se alinea con dos o más secuencias en el genoma de referencia. En algunas implementaciones, las lecturas mapeadas de manera no única se eliminan del análisis posterior (por ejemplo, cuantificación). Puede permitirse que un determinado grado pequeño de apareamiento erróneo (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 mapean, en determinadas implementaciones. En algunas implementaciones, no se permite ningún grado de apareamiento erróneo para que una lectura se mapee en una secuencia de referencia.

Tal como se usa en el presente documento, la expresión "genoma de referencia" puede referirse a cualquier genoma conocido, secuenciado o caracterizado, ya sea parcial o completo, de cualquier organismo o virus que pueda usarse para referenciar secuencias identificadas de un sujeto. Por ejemplo, un genoma de referencia usado para sujetos humanos, así como muchos otros organismos, puede encontrarse 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 ácido nucleico. Tal como se usa en el presente documento, una secuencia de referencia o un genoma de referencia es a menudo una secuencia genómica ensamblada o parcialmente ensamblada de un individuo o múltiples individuos. En algunas implementaciones, un genoma de referencia es una secuencia genómica ensamblada o parcialmente ensamblada de uno o más individuos humanos. En algunas implementaciones, un genoma de referencia comprende secuencias asignadas a cromosomas.

En determinadas implementaciones, en las que un ácido nucleico de muestra proviene de una mujer embarazada, una secuencia de referencia algunas veces no es del feto, la madre del feto o el padre del feto, y se denomina en el presente documento "referencia externa". Una referencia materna puede prepararse y usarse en algunas implementaciones. Cuando se prepara una referencia de la mujer embarazada ("secuencia de referencia materna") basándose en una referencia externa, las lecturas de ADN de la mujer embarazada que no contiene sustancialmente ADN fetal se mapean en menudo en la secuencia de referencia externa y se ensamblan. En determinadas implementaciones, la referencia externa es de ADN de un individuo que tiene sustancialmente la misma etnia que la mujer embarazada. 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 aparearse perfectamente con la secuencia de ADN genómico materno (por ejemplo, la secuencia de referencia materna puede incluir múltiples apareamientos erróneos).

En determinadas implementaciones, la mapeabilidad se evalúa para una región genómica (por ejemplo, sección genómica, porción, porción genómica, porción). La capacidad de mapeo es la capacidad de alinear de manera inequívoca una secuencia de nucleótidos leída con una porción de un genoma de referencia normalmente hasta un número especificado de

apareamientos erróneos, incluyendo, por ejemplo, 0, 1, 2 o más apareamientos erróneos. Para una región genómica dada, la capacidad de mapeo esperada puede estimarse usando un enfoque de ventana deslizante de una longitud de lectura predeterminada y promediando los valores de capacidad de mapeo a nivel de lectura resultantes. Las regiones genómicas que comprenden tramos de secuencia de nucleótidos única tienen, algunas veces, un alto valor de capacidad de mapeo.

Porciones

En algunas implementaciones, las lecturas de secuencia mapeadas (es decir, recuentos o etiquetas de secuencias) se agrupan según diversos parámetros y se asignan a porciones concretas (por ejemplo, porciones de un genoma de referencia). A menudo, las lecturas de secuencia mapeadas individuales se pueden usar para identificar una porción (por ejemplo, la presencia, ausencia o cantidad de una porción) presente en una muestra. En algunas implementaciones, la cantidad de una porción es indicativa de la cantidad de una secuencia mayor (por ejemplo, un cromosoma) en la muestra. El término ‘‘porción’’ también se puede denominar en el presente documento ‘‘sección genómica’’, ‘‘bin’’, ‘‘región’’, ‘‘partición’’, ‘‘porción de un genoma de referencia’’, ‘‘porción de un cromosoma’’ o ‘‘porción genómica’’. En algunas implementaciones, una porción es un cromosoma entero, un segmento de un cromosoma, un segmento de un genoma de referencia, un segmento que abarca varios cromosomas, varios segmentos de cromosomas y/o combinaciones de los mismos. En algunas implementaciones, una porción está predefinida en función de parámetros específicos. En algunas implementaciones, una porción se define arbitrariamente basándose en la partición de un genoma (por ejemplo, partición por tamaño, contenido de GC, regiones contiguas, regiones contiguas de un tamaño definido arbitrariamente, y similares).

En algunas implementaciones, una porción se delinea basándose en uno o más parámetros que incluyen, por ejemplo, la longitud o una o varias características particulares de la secuencia. Las porciones pueden seleccionarse, filtrarse y/o eliminarse de la consideración utilizando cualquier criterio adecuado conocido en la técnica o descrito en el presente documento. En algunas implementaciones, una porción se basa en una longitud concreta de secuencia genómica. En algunas implementaciones, un método puede incluir el análisis de múltiples lecturas de secuencia mapeadas en una pluralidad de porciones. Las porciones pueden tener aproximadamente la misma longitud o las porciones pueden tener longitudes diferentes. En algunas implementaciones, las porciones tienen aproximadamente la misma longitud. En algunas implementaciones, las porciones de longitudes diferentes se ajustan o ponderan. En algunas implementaciones, una porción es de aproximadamente 10 kilobases (kb) a aproximadamente 100 kb, de aproximadamente 20 kb a aproximadamente 80 kb, de aproximadamente 30 kb a aproximadamente 70 kb, de aproximadamente 40 kb a aproximadamente 60 kb, y a veces de aproximadamente 50 kb. En algunas implementaciones, una porción es de aproximadamente 10 kb a aproximadamente 20 kb. Una porción no se limita a tramos contiguos de secuencia. Por tanto, las porciones pueden estar formadas por secuencias contiguas y/o no contiguas. Una porción no se limita a un único cromosoma. En algunas implementaciones, una porción incluye la totalidad o parte de un cromosoma o la totalidad o parte de dos o más cromosomas. En algunas implementaciones, las porciones pueden abarcar uno, dos o más cromosomas enteros. Además, las porciones pueden abarcar regiones unidas o desunidas de múltiples cromosomas.

En algunas implementaciones, las porciones pueden ser segmentos cromosómicos particulares de un cromosoma de interés, tales como, por ejemplo, un cromosoma donde se evalúe una variación genética (por ejemplo, una aneuploidía de los cromosomas 13, 18 y/o 21 o un cromosoma sexual). Una porción puede ser además un genoma patógeno (por ejemplo, bacteriano, fúngico o vírico) o fragmento del mismo. Las porciones pueden ser genes, fragmentos génicos, secuencias reguladoras, intrones, exones y similares.

En algunas implementaciones, un genoma (por ejemplo, el genoma humano) se divide en porciones basándose en el contenido de información de regiones concretas. En algunas implementaciones, 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 único cromosoma o abarcar múltiples cromosomas. En algunas implementaciones, un genoma dividido se recorta y optimiza para una alineación más rápida, lo que suele permitir centrarse en secuencias identificables de forma única.

En algunas implementaciones, la partición puede ponderar a la baja regiones similares. A continuación, se analiza con más detalle un proceso para ponderar a la baja una porción.

En algunas implementaciones, la partición de un genoma en regiones que transcienden a los cromosomas puede basarse en la ganancia de información producida en el contexto de la clasificación. Por ejemplo, el contenido de información puede cuantificarse usando un perfil de valor de *p* que mide la significación de ubicaciones genómicas particulares para distinguir entre grupos de sujetos normales y anómalos confirmados (por ejemplo, sujetos euploides y con trisomía, respectivamente). En algunas implementaciones, la partición de un genoma en regiones que trascienden a los cromosomas puede basarse en cualquier otro criterio, tal como, por ejemplo, la velocidad/conveniencia al alinear las etiquetas, el contenido de GC (por ejemplo, alto o bajo contenido de GC), la uniformidad del contenido de GC, otras medidas del contenido de la secuencia (por ejemplo, la fracción de nucleótidos individuales, la fracción de pirimidinas o purinas, la fracción de ácidos nucleicos naturales frente a los no naturales, la fracción de nucleótidos metilados y el contenido de CpG), el estado de metilación, la temperatura de fusión de la estructura doble, la susceptibilidad a la secuenciación o la PCR, la medida de la incertidumbre asignada a porciones individuales de un genoma de referencia y/o una búsqueda dirigida para características concretas.

Un "segmento" de un cromosoma generalmente es parte de un cromosoma, y normalmente 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 que una porción, a veces no comparte un polinucleótido con una porción y, a veces, incluye un polinucleótido que se encuentra en una porción. Un segmento de un cromosoma suele contener una cantidad mayor de nucleótidos que una porción (por ejemplo, un segmento a veces incluye una porción) y, a veces, un segmento de un cromosoma contiene una cantidad menor de nucleótidos que una porción (por ejemplo, un segmento a veces está dentro de una porción).

Recuentos

Las lecturas de secuencia que se mapean o particionan en función de una característica o variable seleccionada pueden cuantificarse para determinar el número de lecturas que se mapean con una o más porciones (por ejemplo, porción de un genoma de referencia), en algunas implementaciones. En determinadas implementaciones, la cantidad de lecturas de secuencia que se mapean en una porción se denominan recuentos (por ejemplo, un recuento). A menudo, un recuento está asociado a una porción. En algunas implementaciones, los recuentos son específicos de una porción. En determinadas implementaciones, los recuentos de dos o más porciones (por ejemplo, un conjunto de porciones) se manipulan matemáticamente (por ejemplo, se promedian, se suman, se normalizan, etc. o una combinación de los mismos). En algunas implementaciones, un recuento se determina a partir de algunas o todas las lecturas de secuencia mapeadas en (es decir, asociadas con) una porción. En determinadas implementaciones, se determina un recuento a partir de un subconjunto predefinido de lecturas de secuencia mapeadas. Los subconjuntos predefinidos de lecturas de secuencia mapeadas pueden definirse o seleccionarse con el uso de cualquier característica o variable adecuada. En algunas implementaciones, los subconjuntos predefinidos de lecturas de secuencia mapeadas pueden incluir de 1 a n lecturas de secuencia, en el que n representa un número igual a la suma de todas las lecturas de secuencia generadas a partir de una muestra de sujeto de prueba o de sujeto de referencia.

En determinadas implementaciones, un recuento se 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, una operación o un procedimiento matemático adecuado. En determinadas implementaciones, un recuento se deriva de lecturas de secuencia asociadas con una porción donde algunas o todas las lecturas de secuencia se ponderan, eliminan, filtran, normalizan, ajustan, promedian, derivan como una media, se suman, o se restan o procesan mediante una combinación de los mismos. En algunas implementaciones, un recuento se deriva de lecturas de secuencia sin procesar y/o lecturas de secuencia filtradas. En determinadas implementaciones, un valor de recuento se determina mediante un proceso matemático. En determinadas implementaciones un valor de recuento es un promedio, una media o una suma de lecturas de secuencia mapeadas en una porción. En algunas implementaciones, un recuento es un número medio de recuentos. En algunas implementaciones, un recuento se asocia a una medida de incertidumbre.

En algunas implementaciones, los recuentos pueden manipularse o transformarse (por ejemplo, normalizarse, combinarse, sumarse, filtrarse, seleccionarse, promediarse, derivarse como una media, o similar o una combinación de los mismos). En algunas implementaciones, los recuentos pueden transformarse para producir recuentos normalizados. Los recuentos pueden procesarse (por ejemplo, normalizarse) mediante un método conocido en la técnica y/o tal como se describe en el presente documento (por ejemplo, normalización en porciones, normalización basada en bins, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, LOESS de GC, LOWESS, PERUN, RM, GCRM, cQn y/o combinaciones de los mismos).

En algunas implementaciones, los recuentos de una porción se proporcionan como una representación de los recuentos. En determinadas implementaciones, una representación de recuentos se determina según los recuentos de una porción divididos entre los recuentos totales de todas las porciones autosómicas (por ejemplo, todas las porciones autosómicas en un perfil o segmento de un genoma).

Los recuentos (por ejemplo, recuentos sin procesar, filtrados y/o normalizados) pueden procesarse y normalizarse a una o más elevaciones. Las elevaciones y los perfiles se describen con mayor detalle más adelante. En determinadas implementaciones, los recuentos pueden procesarse y/o normalizarse a una elevación de referencia. Las elevaciones de referencia se abordan más adelante en el presente documento. Los recuentos procesados según una elevación (por ejemplo, recuentos procesados) pueden asociarse con un valor de incertidumbre (por ejemplo, una varianza calculada, un error, desviación estándar, valor de p, desviación media absoluta, etc.). Un valor de incertidumbre define normalmente un rango por encima y por debajo de una elevación. Puede usarse un valor de desviación en lugar de un valor de incertidumbre y los ejemplos no limitativos de medidas de desviación incluyen desviación estándar, desviación absoluta promedio, mediana de desviación absoluta, puntuación estándar (por ejemplo, puntuación Z, valor de Z, puntuación normal, variable normalizada) y similares.

Los recuentos (por ejemplo, recuentos sin procesar, filtrados y/o normalizados) pueden procesarse y normalizarse a uno o más niveles. Los niveles, niveles de referencia y perfiles se describen con mayor detalle más adelante. En determinadas implementaciones, los recuentos (por ejemplo, sin procesar, procesados y/o normalizados), los recuentos de una porción, porciones, niveles y/o perfiles están asociados con una medida de incertidumbre. Los recuentos procesados según un nivel (por ejemplo, recuentos procesados) a veces se asocian con una medida de incertidumbre. Los ejemplos no limitativos de una medida de incertidumbre incluyen varianza, una varianza calculada, covarianza, una medida de error,

un error, error estándar, error absoluto, un factor R, desviación estándar, desviación absoluta, puntuación Z, valor de *p*, puntuación estándar, error absoluto medio (EAM), desviación absoluta promedio, desviación absoluta media (DAM), mediana de la desviación absoluta, similares o combinaciones de los mismos. En algunas implementaciones, una medida de incertidumbre define un rango por encima y por debajo de un nivel (por ejemplo, un nivel de sección genómica).

Los recuentos se obtienen a menudo a partir de una muestra de ácido nucleico de una mujer embarazada que porta un feto. Los recuentos de lecturas de secuencia de ácido nucleico mapeadas en una o más porciones son a menudo recuentos representativos tanto del feto como de la madre del feto (por ejemplo, sujeto de sexo femenino gestante). En determinadas implementaciones, algunos de los recuentos mapeados en una porción proceden de un genoma fetal y algunos de los recuentos mapeados en la misma porción proceden de un genoma materno.

En algunas implementaciones, una proporción de todas las lecturas de secuencia son de un cromosoma que interviene una aneuploidía (por ejemplo, el cromosoma 13, el cromosoma 18, el cromosoma 21), y otras lecturas de secuencia son de otros cromosomas. Teniendo en cuenta el tamaño relativo del cromosoma involucrado en la aneuploidía (por ejemplo, "cromosoma diana": cromosoma 21) en comparación con otros cromosomas, podría obtenerse una frecuencia normalizada, dentro de un rango de referencia, de secuencias específicas del cromosoma diana, en algunas implementaciones. Si el feto tiene una aneuploidía 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 del cromosoma no diana, lo que permite así la detección de la aneuploidía. El grado de cambio en la frecuencia normalizada dependerá de la concentración fraccional de ácidos nucleicos fetales en la muestra analizada, en algunas implementaciones.

Niveles

En algunas implementaciones, se atribuye un valor (por ejemplo, un número, un valor cuantitativo) a un nivel. Un nivel puede determinarse mediante un método, una operación o un procedimiento matemático adecuado (por ejemplo, un nivel procesado). A menudo, un nivel es, o se deriva de, recuentos (por ejemplo, recuentos normalizados) para un conjunto de porciones. En algunas implementaciones, el nivel de una porción es sustancialmente igual al número total de recuentos mapeados en una porción (por ejemplo, recuentos, recuentos normalizados). Con frecuencia, un nivel se determina a partir de recuentos que se procesan o manipulan mediante un método, una operación o un procedimiento matemático adecuado conocido en la técnica. En algunas implementaciones, un nivel se deriva de recuentos que se procesan y los ejemplos no limitativos de recuentos procesados incluyen recuentos ponderados, eliminados, filtrados, normalizados, ajustados, promediados, derivados como una media (por ejemplo, nivel medio), sumados, restados, transformados o combinaciones de los mismos. En algunas implementaciones, un nivel comprende recuentos que están normalizados (por ejemplo, recuentos normalizados de porciones). Un nivel puede ser para recuentos normalizados mediante un procedimiento adecuado, los ejemplos no limitativos de los cuales incluyen normalización basada en porciones, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, 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 algunas implementaciones un nivel es para recuentos o recuentos normalizados de dos o más porciones que se promedian, y el nivel se denomina nivel medio. En algunas implementaciones un nivel es para un conjunto de porciones que tienen un recuento medio o media de recuentos normalizados que se denomina nivel medio. En algunas implementaciones se deriva un nivel para porciones que comprenden recuentos sin procesar y/o filtrados. En algunas implementaciones, un nivel se basa en recuentos que están sin procesar. En algunas implementaciones, un nivel está asociado con una medida de incertidumbre (por ejemplo, una desviación estándar, una DAM). En algunas implementaciones, un nivel se representa mediante una puntuación Z o un valor de *p*. En el presente documento, un nivel para una o más porciones es sinónimo de un "nivel de sección genómica".

Los recuentos normalizados o no normalizados para dos o más niveles (por ejemplo, dos o más niveles en un perfil) pueden, algunas veces, manipularse matemáticamente (por ejemplo, sumarse, multiplicarse, promediarse, normalizarse, similares o una combinación de los mismos) según los niveles. Por ejemplo, los recuentos normalizados o no normalizados para dos o más niveles pueden normalizarse según una, algunas o todos los niveles de un perfil. En algunas implementaciones, los recuentos normalizados o no normalizados de todos los niveles en un perfil se normalizan según un nivel del perfil. En algunas implementaciones, los recuentos normalizados o no normalizados de un primer nivel en un perfil se normalizan según los recuentos normalizados o no normalizados de un segundo nivel en el perfil.

Los ejemplos no limitativos de un nivel (por ejemplo, un primer nivel, un segundo nivel) son un nivel para un conjunto de porciones que comprenda recuentos procesados, un nivel para un conjunto de porciones que comprenda una media, mediana o promedio de recuentos, un nivel para un conjunto de porciones que comprenda recuentos normalizados, similares o cualquier combinación de los mismos. En algunas implementaciones, un primer nivel y un segundo nivel en un perfil se derivan de recuentos de porciones mapeadas en el mismo cromosoma. En algunas implementaciones, un primer nivel y un segundo nivel en un perfil se derivan de recuentos de porciones mapeadas en cromosomas diferentes.

En algunas implementaciones, un nivel se determina a partir de recuentos normalizados o no normalizados mapeados en una o más porciones. En algunas implementaciones, 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 suelen ser aproximadamente los mismos. 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 que en otras porciones del conjunto (por ejemplo, picos y/o caídas). Cualquier número adecuado de recuentos normalizados o no normalizados asociados a cualquier número adecuado de porciones puede definir un nivel.

En algunas implementaciones, uno o más niveles pueden determinarse 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 de un segmento del mismo. En algunas implementaciones, dos o más recuentos derivados de dos o más porciones (por ejemplo, un conjunto de porciones) determinan un nivel. En algunas implementaciones, dos o más recuentos (por ejemplo, recuentos de dos o más porciones) determinan un nivel. En algunas implementaciones, los recuentos de 2 a aproximadamente 100.000 porciones determinan un nivel. En algunas implementaciones, recuentos de 2 a aproximadamente 50.000, de 2 a aproximadamente 40.000, de 2 a aproximadamente 30.000, de 2 a aproximadamente 20.000, de 2 a aproximadamente 10.000, de 2 a aproximadamente 5000, de 2 a aproximadamente 2500, de 2 a aproximadamente 1250, de 2 a aproximadamente 1000, de 2 a aproximadamente 500, de 2 a aproximadamente 250, de 2 a aproximadamente 100 o de 2 a aproximadamente 60 porciones determinan un nivel. En algunas implementaciones, los recuentos de aproximadamente 10 a aproximadamente 50 porciones determinan un nivel. En algunas implementaciones, los recuentos de aproximadamente 20 a aproximadamente 40 o más porciones determinan un nivel. En algunas implementaciones, 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 algunas implementaciones, 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 algunas implementaciones, se determina un nivel para recuentos normalizados o no normalizados de porciones que son contiguas. En algunas implementaciones, 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 uniendo las porciones extremo con extremo, pueden representar un conjunto 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 del mismo. En algunas implementaciones, un nivel se determina a partir de una colección (por ejemplo, un conjunto) de porciones contiguas y/o porciones no contiguas.

Elevaciones

En algunas implementaciones, se atribuye un valor a una elevación (por ejemplo, un número). Una elevación puede determinarse mediante un método, una operación o un procedimiento matemático adecuado (por ejemplo, una elevación procesada). El término "nivel", tal como se usa en el presente documento, es sinónimo del término "elevación" tal como se usa en el presente documento. Una elevación es, o se deriva de, recuentos (por ejemplo, recuentos normalizados) para un conjunto de porciones. En algunas implementaciones, una elevación de una porción es sustancialmente igual al número total de recuentos mapeados en una porción (por ejemplo, recuentos normalizados). Con frecuencia, una elevación se determina a partir de recuentos que se procesan o manipulan mediante un método, una operación o un procedimiento matemático adecuado conocido en la técnica. En algunas implementaciones, una elevación se deriva de recuentos que se procesan y los ejemplos no limitativos de recuentos procesados incluyen recuentos ponderados, eliminados, filtrados, normalizados, ajustados, promediados, derivados como una media (por ejemplo, elevación media), sumados, restados, transformados o combinaciones de los mismos. En algunas implementaciones, una elevación comprende recuentos que están normalizados (por ejemplo, recuentos normalizados de porciones). Una elevación puede ser para recuentos normalizados mediante un procedimiento adecuado, los ejemplos no limitativos de los cuales incluyen normalización basada en bins, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, LOESS de GC, LOWESS, PERUN, RM, GCRM, cQn, similares y/o combinaciones de los mismos. Una elevación puede comprender recuentos normalizados o cantidades relativas de recuentos. En algunas implementaciones, una elevación es para recuentos o recuentos normalizados de dos o más porciones que se promedian, y la elevación se denomina elevación promedio. En algunas implementaciones, una elevación es para un conjunto de porciones que tienen un recuento medio o media de recuentos normalizados que se denomina elevación media. En algunas implementaciones, una elevación se deriva para porciones que comprenden recuentos sin procesar y/o filtrados. En algunas implementaciones, una elevación se basa en recuentos que son sin procesar. En algunas implementaciones, una elevación se asocia a un valor de incertidumbre. Una elevación para una porción, o una "elevación de sección genómica", es sinónimo de un "nivel de sección genómica" en el presente documento.

Los recuentos normalizados o no normalizados para dos o más elevaciones (por ejemplo, dos o más elevaciones en un perfil) pueden, algunas veces, manipularse matemáticamente (por ejemplo, sumarse, multiplicarse, promediarse, normalizarse, similares o una combinación de los mismos) según las elevaciones. Por ejemplo, los recuentos normalizados o no normalizados para dos o más elevaciones pueden normalizarse según una, algunas o todas las elevaciones de un perfil. En algunas implementaciones, los recuentos normalizados o no normalizados de todas las elevaciones de un perfil se normalizan según una elevación del perfil. En algunas implementaciones, los recuentos normalizados o no normalizados de una primera elevación en un perfil se normalizan según los recuentos normalizados o no normalizados de una segunda elevación en el perfil.

Los ejemplos no limitativos de una elevación (por ejemplo, una primera elevación, una segunda elevación) son una elevación para un conjunto de porciones que comprenden recuentos procesados, una elevación para un conjunto de porciones que comprenden una media, mediana o promedio de recuentos, una elevación para un conjunto de porciones que comprenden recuentos normalizados, similares o cualquier combinación de los mismos. En algunas implementaciones, una primera elevación y una segunda elevación en un perfil se derivan de recuentos de porciones mapeadas en el mismo cromosoma. En algunas implementaciones, una primera elevación y una segunda elevación en un perfil se derivan de recuentos de porciones mapeadas en cromosomas diferentes.

En algunas implementaciones, una elevación se determina a partir de recuentos normalizados o no normalizados mapeados en una o más porciones. En algunas implementaciones, una elevación se determina a partir de recuentos normalizados o no normalizados mapeados en dos o más porciones, donde los recuentos normalizados para cada porción suelen ser aproximadamente los mismos. Puede haber variación en los recuentos (por ejemplo, recuentos normalizados) en un conjunto de porciones para una elevación. En un conjunto de porciones para una elevación puede haber una o más porciones que tengan recuentos que sean significativamente diferentes que en otras porciones del conjunto (por ejemplo, picos y/o caídas). Cualquier número adecuado de recuentos normalizados o no normalizados asociados a cualquier número adecuado de porciones puede definir una elevación.

En algunas implementaciones, una o más elevaciones pueden determinarse a partir de recuentos normalizados o no normalizados de todas o algunas de las porciones de un genoma. A menudo, puede determinarse una elevación a partir de todos o algunos de los recuentos normalizados o no normalizados de un cromosoma o segmento del mismo. En algunas implementaciones, dos o más recuentos derivados de dos o más porciones (por ejemplo, un conjunto de secciones genómicas) determinan una elevación. En algunas implementaciones, dos o más recuentos (por ejemplo, recuentos de dos o más porciones) determinan una elevación. En algunas implementaciones, los recuentos de 2 a aproximadamente 100.000 porciones determinan una elevación. En algunas implementaciones, recuentos de 2 a aproximadamente 50.000, de 2 a aproximadamente 40.000, de 2 a aproximadamente 30.000, de 2 a aproximadamente 20.000, de 2 a aproximadamente 10.000, de 2 a aproximadamente 5000, de 2 a aproximadamente 2500, de 2 a aproximadamente 1250, de 2 a aproximadamente 1000, de 2 a aproximadamente 500, de 2 a aproximadamente 250, de 2 a aproximadamente 100 o de 2 a aproximadamente 60 porciones determinan una elevación. En algunas implementaciones, los recuentos de aproximadamente 10 a aproximadamente 50 porciones determinan una elevación. En algunas implementaciones, los recuentos de aproximadamente 20 a aproximadamente 40 o más porciones determinan una elevación. En algunas implementaciones, una elevación 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 algunas implementaciones, una elevación 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 algunas implementaciones, se determina una elevación para recuentos normalizados o no normalizados de porciones que son contiguas. En algunas implementaciones, 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 uniendo las porciones extremo con extremo, pueden representar un conjunto 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 del mismo. En algunas implementaciones, una elevación partir de una colección (por ejemplo, un conjunto) de porciones contiguas y/o porciones no contiguas.

Procesamiento y normalización de datos

Las lecturas de secuencia mapeadas que se han contado se denominan en el presente documento datos sin procesar, ya que los datos representan recuentos sin manipular (por ejemplo, recuentos sin procesar). Los datos de lecturas de secuencia (por ejemplo, recuentos) de un conjunto de datos pueden procesarse posteriormente (por ejemplo, manipularse matemática y/o estadísticamente) y/o visualizarse para facilitar la obtención de un resultado. En determinadas implementaciones, los conjuntos de datos, incluyendo conjuntos de datos más grandes, pueden beneficiarse del preprocesamiento para facilitar un análisis adicional. El preprocesamiento de conjuntos de datos a veces implica la eliminación de porciones, bins o porciones redundantes y/o poco informativas de un genoma de referencia (por ejemplo, bins o porciones de un genoma de referencia con datos poco informativos, lecturas mapeadas redundantes, porciones o bins con recuentos de mediana nula, secuencias representadas en exceso o en defecto). Sin desear limitarse por la teoría, el procesamiento y/o preprocesamiento de datos puede (i) eliminar datos con ruido, (ii) eliminar datos no informativos, (iii) eliminar datos redundantes, (iv) reducir la complejidad de conjuntos de datos más grandes, (v) reducir el sesgo experimental y/o sistemático y/o (vi) facilitar la transformación de los datos de una forma a otra u otras formas. Los términos “preprocesamiento” y “procesamiento”, cuando se usan con respecto a los datos o conjuntos de datos, se denominan colectivamente en el presente documento “procesamiento”. El procesamiento puede hacer que los datos sean más susceptibles de análisis adicional, y puede generar un resultado en algunas implementaciones. En algunas implementaciones, uno o más o todos los métodos de procesamiento (por ejemplo, métodos de normalización, filtrado de bins o porciones, mapeo, validación, similares o combinaciones de los mismos) son realizados por un procesador, un microprocesador, un ordenador, junto con la memoria y/o por un aparato controlado por microprocesador.

La expresión “datos con ruido”, tal 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 desviación estándar significativa (por ejemplo, mayor de 3 desviaciones estándar), (c) datos que tienen un error estándar de la media significativo, similares y combinaciones de los anteriores. Los datos con ruido suceden, algunas veces, debido a la cantidad y/o calidad del material de partida (por ejemplo, muestra de ácido nucleico) y, algunas veces, suceden como parte de los procedimientos para preparar o replicar el ADN usado para generar lecturas de secuencia. En determinadas implementaciones, el ruido resulta de determinadas secuencias sobrerrepresentadas cuando se preparan usando métodos basados en PCR. Los métodos descritos en el presente documento pueden reducir o eliminar la contribución de los datos con ruido y, por tanto, reducir el efecto de los datos con ruido sobre el resultado proporcionado.

Las expresiones “datos no informativos”, “bins no informativos”, “porciones no informativas de un genoma de referencia” y “porciones no informativas”, tal y como se utilizan en el presente documento, se refieren a porciones, o datos derivados de las mismas, que tienen un valor numérico que es significativamente diferente de un valor umbral predeterminado o que cae fuera de un rango de valores de corte predeterminado. Las expresiones “umbral” y “valor umbral” en el presente documento se refieren a cualquier número que se calcula usando un conjunto de datos calificados y sirve como límite de diagnóstico de una variación genética (por ejemplo, una variación en el número de copias, una aneuploidía, una aberración cromosómica y similares). En determinadas implementaciones, se supera un umbral por los resultados obtenidos mediante los métodos descritos en el presente documento y a un sujeto se le diagnostica una variación genética (por ejemplo, trisomía 21). A menudo, se calcula un valor umbral o rango de valores al manipular matemática y/o estadísticamente los datos leídos de secuencias (por ejemplo, a partir de una referencia y/o sujeto), en algunas implementaciones, y en determinadas implementaciones, los datos leídos de secuencias manipulados para generar un valor umbral o rango de valores son datos leídos de secuencias (por ejemplo, a partir de una referencia y/o sujeto). En algunas implementaciones, se determina una medida de incertidumbre. En algunas implementaciones, se determina un valor de incertidumbre. Se puede determinar un valor de incertidumbre mediante un método adecuado. Un valor de incertidumbre es generalmente una medida de varianza o error y puede ser cualquier medida de varianza o error adecuada. En algunas implementaciones, un valor de incertidumbre es una desviación estándar, un error estándar, una varianza calculada, un valor de *p* o una desviación absoluta media (DAM).

Puede usarse cualquier procedimiento adecuado para procesar conjuntos de datos descritos en el presente documento. Los ejemplos no limitativos de procedimientos adecuados para su uso para procesar conjuntos de datos incluyen filtrado, normalización, ponderación, monitorización de alturas de pico, monitorización de áreas de pico, monitorización de bordes de pico, determinación de razones 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 el procesamiento adicional, similares y combinaciones de los anteriores. En algunas implementaciones, los conjuntos de datos se procesan basándose en diversas características (por ejemplo, contenido de GC, lecturas mapeadas redundantes, regiones de centrómero, regiones de telómero, similares y combinaciones de los mismos) y/o variables (por ejemplo, sexo del feto, edad materna, ploidía materna, contribución porcentual de ácido nucleico fetal, similares o combinaciones de los mismos). En determinadas implementaciones, el procesamiento de conjuntos de datos tal 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 limitativo de un conjunto de datos complejos incluye datos de lectura de secuencia generados a partir de uno o más sujetos de prueba y una pluralidad de sujetos de referencia de diferentes edades y orígenes étnicos. En algunas implementaciones, los conjuntos de datos pueden incluir de miles a millones de lecturas de secuencia para cada sujeto de prueba y/o referencia.

El procesamiento de datos puede realizarse en cualquier número de etapas, en determinadas implementaciones. Por ejemplo, los datos pueden procesarse usando solamente un único procedimiento de procesamiento en algunas implementaciones, y en determinadas implementaciones los datos pueden procesarse usando 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 o 20 o más etapas de procesamiento). En algunas implementaciones, las etapas de procesamiento pueden ser la misma etapa repetida dos o más veces (por ejemplo, filtrar dos o más veces, normalizar dos o más veces), y en determinadas implementaciones, las etapas de procesamiento pueden ser dos o más etapas de procesamiento diferentes (por ejemplo, filtrado, normalización; normalización, monitorización de alturas y bordes de pico; filtrado, normalización, normalización con respecto a una referencia, manipulación estadística para determinar valores de p, y similares), realizadas de forma simultánea o secuencial. En algunas implementaciones, puede usarse cualquier número y/o combinación adecuada de las mismas o diferentes etapas de procesamiento para procesar datos de lectura de secuencia para facilitar la provisión de un resultado. En determinadas implementaciones, el procesamiento de conjuntos de datos por los criterios descritos en el presente documento puede reducir la complejidad y/o dimensionalidad de un conjunto de datos.

En algunas implementaciones, una o más etapas de procesamiento pueden comprender una o más etapas de filtrado. El término “filtrado”, tal y como se utiliza en el presente documento, se refiere a la eliminación de porciones o bins o porciones de un genoma de referencia de la consideración. Los bins o porciones de un genoma de referencia pueden

seleccionarse para su eliminación (por ejemplo, filtrarse) en función de cualquier criterio adecuado, incluidos, entre otros, los datos redundantes (por ejemplo, lecturas mapeadas redundantes o solapadas), los datos no informativos (por ejemplo, bins o porciones de un genoma de referencia con recuentos medios nulos), bins o porciones de un genoma de referencia con secuencias representadas en exceso o en defecto, contenido de GC, datos con ruido, mapeabilidad, recuentos, variabilidad de recuentos, una medida de incertidumbre, una medida de repetibilidad, similares, o combinaciones de los anteriores. Un proceso de filtrado a menudo implica eliminar una o más porciones de un genoma de referencia de la consideración y/o restar los recuentos en la una o más porciones de un genoma de referencia seleccionadas para su eliminación de los recuentos contados o sumados para las porciones de un genoma de referencia, cromosoma o cromosomas, o genoma en consideración. En determinadas implementaciones, un proceso de filtrado implica eliminar uno o más bins de la consideración y restar los recuentos en el uno o más bins seleccionados para su eliminación de los recuentos contados o sumados para los bins, cromosoma o cromosomas, o genoma en consideración. Las porciones o bins a veces se filtran antes y/o después del procesamiento (por ejemplo, promediando, normalizando, ajustando y/o similares). En algunas implementaciones, las porciones se filtran una vez determinados los niveles de sección genómica normalizados para una porción. Por ejemplo, a veces se filtran porciones tras determinar los niveles de sección genómica normalizados según una normalización PERUN (por ejemplo, una normalización lineal, no lineal, cuadrática, semicuadrática, cuasicuadrática, canónica o similar). A menudo, las porciones se filtran según una medida de incertidumbre (por ejemplo, una medida de error) determinada para y/o asociada con recuentos de una porción y/o un nivel de sección genómica (por ejemplo, un nivel de sección genómica normalizado) determinado para una porción. En algunas implementaciones, las porciones se filtran según un umbral o rango predeterminado, donde el umbral o rango se determina según una medida de incertidumbre (por ejemplo, un rango de +/-DE, un rango de DAM). Por ejemplo, a veces los niveles de sección genómica normalizados se determinan según un método PERUN, se determina una medida de incertidumbre para todos los niveles de sección genómica en un perfil y las porciones se filtran según un umbral predeterminado de la medida de incertidumbre (por ejemplo, todas las porciones con una DAM superior a 3 se eliminan de la consideración).

En algunas implementaciones, las porciones se pueden filtrar según una medida de error o incertidumbre (por ejemplo, desviación estándar, error estándar, varianza calculada, valor de *p*, error absoluto medio (EAM), desviación absoluta promedio y/o desviación absoluta media (DAM)). En determinadas implementaciones, una medida de incertidumbre se refiere a la variabilidad del recuento. En algunas implementaciones, las porciones se filtran según la variabilidad del recuento. En determinadas implementaciones, la variabilidad del recuento es una medida de la incertidumbre del error para los recuentos mapeados en una porción (es decir, una porción) de un genoma de referencia para múltiples muestras (por ejemplo, múltiples muestras obtenidas 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 algunas implementaciones, las porciones con una variabilidad de recuento por encima de un rango superior predeterminado se filtran (por ejemplo, se excluyen de la consideración). En algunas implementaciones, un rango superior predeterminado es un valor de la DAM igual o superior a 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 de 74 o igual o superior a aproximadamente 76. En algunas implementaciones, las porciones con una variabilidad de recuento por debajo de un rango inferior predeterminado se filtran (por ejemplo, se excluyen de la consideración). En algunas implementaciones, un rango inferior predeterminado es un valor de la DAM igual o inferior a aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25, aproximadamente 20, aproximadamente 15, aproximadamente 10, aproximadamente 5, aproximadamente 1, o igual o inferior a aproximadamente 0. En algunas implementaciones, las porciones con una variabilidad de recuento fuera de un rango predeterminado se filtran (por ejemplo, se excluyen de la consideración). En algunas implementaciones, un rango predeterminado es un valor de la DAM superior a cero e inferior a aproximadamente 76, inferior a aproximadamente 74, inferior a aproximadamente 73, inferior a aproximadamente 72, inferior a aproximadamente 71, inferior a aproximadamente 70, inferior a aproximadamente 69, inferior a aproximadamente 68, inferior a aproximadamente 67, inferior a aproximadamente 66, inferior a aproximadamente 65, inferior a aproximadamente 64, inferior a aproximadamente 62, inferior a aproximadamente 60, inferior a aproximadamente 58, inferior a aproximadamente 56, inferior a aproximadamente 54, inferior a aproximadamente 52 o inferior a aproximadamente 50. En algunas implementaciones, un rango predeterminado es un valor de la DAM superior a cero e inferior a aproximadamente 67,7. En algunas implementaciones, se seleccionan porciones con una variabilidad de recuento dentro de un rango predeterminado (por ejemplo, se usan para determinar la presencia o ausencia de una variación genética).

En algunas implementaciones, la variabilidad del recuento de porciones representa una distribución (por ejemplo, una distribución normal). En algunas implementaciones, las porciones se seleccionan dentro de un cuantil de la distribución. En algunas implementaciones, se seleccionan las 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 algunas implementaciones, se seleccionan porciones dentro de un cuantil del 99 % de la distribución de la variabilidad del recuento. En algunas implementaciones, las porciones con DAM > 0 y DAM < 67,725 están dentro del cuantil del 99 % y se seleccionan, lo que resulta en la identificación de un conjunto de porciones estables de un genoma de referencia. En algunas implementaciones, se seleccionan porciones con una DAM > 0 y una DAM < 67,725 dentro del cuantil del 99 %, lo que da como resultado la identificación de un conjunto de bins estables.

En el presente documento y en la solicitud de patente internacional n.° PCT/US12/59123 (WO2013/0522913), se proporcionan ejemplos no limitativos de filtrado de porciones con respecto a PERUN.

Se puede utilizar una medida de error que comprenda valores absolutos de desviación, tales como un factor R, para la eliminación, el filtrado y/o la ponderación de porciones en determinadas implementaciones. En algunas implementaciones, un factor R representa un error residual en un modelo o una variación no explicada. Un factor R puede determinarse mediante un método adecuado. Un factor R, en algunas implementaciones, se define como la suma de las desviaciones absolutas de los valores de recuento predichos con respecto a las mediciones reales dividida por los valores de recuento predichos con respecto a las mediciones reales (por ejemplo, la ecuación B del presente documento). En algunas implementaciones, las porciones se filtran con un factor R (por ejemplo, para recuentos de lecturas de secuencia para una porción) de aproximadamente el 1 % a aproximadamente el 20 %, de aproximadamente el 5 % a aproximadamente el 15 %, de aproximadamente el 7 % a aproximadamente el 10 %, o con un Factor R de aproximadamente el 6 %, 7 %, 8 %, 9 %, 10 %, 11 % o 12 %. Aunque puede usarse una medida de error que comprende valores absolutos de desviación, puede usarse alternativamente una medida de incertidumbre adecuada. En determinadas implementaciones puede usarse una medida de incertidumbre que no comprende valores absolutos de desviación, tal como una dispersión basada en cuadrados. En algunas implementaciones, las porciones se filtran o se ponderan según una medida de mapeabilidad (por ejemplo, una puntuación de mapeabilidad). A veces, una porción se filtra o se pondera según 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). En algunas implementaciones, una porción con cero lecturas mapeadas en ella (por ejemplo, mapeabilidad nula), se filtra y se elimina de la consideración. Las porciones pueden filtrarse o ponderarse según el tipo de análisis que se realice. Por ejemplo, para el análisis de aneuploidía del cromosoma 13, 18 y/o 21, los cromosomas sexuales pueden filtrarse, y pueden analizarse solo autosomas, o un subconjunto de autosomas.

En implementaciones particulares, puede emplearse el siguiente procedimiento de filtrado. Se selecciona el mismo conjunto de porciones (por ejemplo, bins o porciones de un genoma de referencia) dentro de un cromosoma determinado (por ejemplo, el cromosoma 21) y se compara el número de lecturas en las muestras afectadas y no afectadas. La distancia (por ejemplo, la diferencia de niveles) relaciona la trisomía 21 y las muestras euploides y afecta a un conjunto de porciones que abarcan la mayor parte del cromosoma 21. El conjunto de porciones es el mismo entre las muestras euploides y las T21. En algunas implementaciones, la distinción entre un conjunto de porciones y una única sección no es crucial, ya que se puede definir una porción. Se compara la misma región genómica en diferentes pacientes (por ejemplo, diferentes muestras de prueba). Este procedimiento puede utilizarse para un análisis de trisomía, tal como para T13 o T18 además de, o en lugar de, T21.

Después de que los conjuntos de datos se han contado, opcionalmente filtrado y normalizado, los conjuntos de datos procesados pueden manipularse mediante ponderación, en algunas implementaciones. Una o más porciones pueden seleccionarse para la ponderación para reducir la influencia de los datos (por ejemplo, datos con ruido, datos no informativos) contenidos en las porciones seleccionadas, en determinadas implementaciones, y en algunas implementaciones, una o más porciones pueden seleccionarse para la ponderación para mejorar o aumentar la influencia de los datos (por ejemplo, datos con pequeña varianza medida) contenidos en las porciones seleccionadas. En algunas implementaciones, se pondera un conjunto de datos usando una única función de ponderación 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 algunas implementaciones, se genera un gráfico de perfiles de datos procesados manipulados adicionalmente mediante ponderación para facilitar la clasificación y/o proporcionar un resultado. Puede proporcionarse un resultado basado en una representación gráfica de perfiles de datos ponderados.

En algunas implementaciones, una etapa de procesamiento comprende una ponderación. Los términos "ponderado" y "ponderación" o la expresión "función de ponderación" o sus derivados gramaticales o equivalentes, tal y como se utilizan en el presente documento, se refieren a una manipulación matemática de los recuentos, de una parte o de la totalidad de un conjunto de datos utilizada en ocasiones para alterar la influencia de determinadas características o variables del conjunto de datos con respecto a otras características o variables del conjunto de datos (por ejemplo, aumentar o disminuir la importancia y/o la contribución de los datos contenidos en una o varias porciones o porciones de un genoma de referencia, en función de la calidad o la utilidad de los datos de la porción o porciones seleccionadas de un genoma de referencia). Las porciones de un genoma de referencia pueden ponderarse en función de cualquier criterio adecuado, incluidos, aunque no de forma limitativa, los datos redundantes (por ejemplo, lecturas mapeadas redundantes o solapadas), los datos no informativos (por ejemplo, porciones de un genoma de referencia con recuentos de mediana nula), las porciones de un genoma de referencia con secuencias representadas en exceso o en defecto, el contenido de GC, el ruido (por ejemplo, datos con ruido), la mapeabilidad, los recuentos, la variabilidad de los recuentos, una medida de incertidumbre, una medida de repetibilidad, similares o combinaciones de las anteriores. Una función de ponderación puede usarse 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, en algunas implementaciones. Por ejemplo, las porciones de un genoma de referencia con datos de secuencia poco representados o de baja calidad (por ejemplo, mapeabilidad nula, es decir, ninguna lectura mapeada en la porción) pueden "ponderarse a la baja" para minimizar la influencia en un conjunto de datos, mientras que las porciones seleccionadas de un genoma de referencia pueden "ponderarse al alza" para aumentar la influencia en un conjunto de datos. Un ejemplo no limitativo de una función de ponderación es $[1/(\text{desviación estándar})^2]$. A veces, una etapa de ponderación se realiza de una manera sustancialmente similar a una etapa de normalización. En

algunas implementaciones, un conjunto de datos se divide entre una variable predeterminada (por ejemplo, variable de ponderación). A menudo, se selecciona una variable predeterminada (por ejemplo, función objetivo minimizada, Phi) para ponderar distintas partes de un conjunto de datos de manera diferente (por ejemplo, aumentar la influencia de determinados tipos de datos mientras se reduce la influencia de otros tipos de datos). A veces, las porciones se ponderan antes y/o después del procesamiento (por ejemplo, promediando, normalizando, ajustando y/o similares). En algunas implementaciones, las porciones se ponderan después de determinar los niveles de sección genómica normalizados para una porción. Por ejemplo, a veces las porciones se ponderan después de determinar los niveles de sección genómica normalizados según una normalización PERUN (por ejemplo, una normalización lineal, no lineal, cuadrática, semicuadrática, cuasicuadrática, canónica o similar). A menudo, las porciones se ponderan según una medida de incertidumbre (por ejemplo, una medida de error) determinada para y/o asociada con recuentos de una porción y/o un nivel de sección genómica (por ejemplo, un nivel de sección genómica normalizado) determinado para una porción. En algunas implementaciones, las porciones se ponderan según un umbral o rango predeterminado, donde el umbral o rango se determina según una medida de incertidumbre (por ejemplo, un rango de +/-DE, un rango de DAM). Por ejemplo, a veces los niveles de sección genómica normalizados se determinan según un método PERUN, se determina una medida de incertidumbre para todos los niveles de sección genómica en un perfil y las porciones se ponderan según un umbral predeterminado de la medida de incertidumbre.

El filtrado o la ponderación de porciones puede realizarse en uno o más puntos adecuados en un análisis. Por ejemplo, las porciones pueden filtrarse o ponderarse antes o después de mapear las lecturas de secuencia en porciones de un genoma de referencia. Las porciones pueden filtrarse o ponderarse antes o después de determinar un sesgo experimental para las porciones individuales del genoma en algunas implementaciones. En determinadas implementaciones, las porciones pueden filtrarse o ponderarse antes o después de que se calculen las elevaciones de las porciones o los niveles de las secciones genómicas.

En algunas implementaciones, las porciones de un genoma de referencia pueden eliminarse sucesivamente (por ejemplo, de una en una para permitir la evaluación del efecto de la eliminación de cada bien o porción individual) y, en determinadas implementaciones, todos los bins o porciones de un genoma de referencia marcados para su eliminación pueden eliminarse al mismo tiempo. En algunas implementaciones, se eliminan porciones de un genoma de referencia caracterizadas por una varianza por encima o por debajo de un determinado nivel, lo que a veces se denomina en el presente documento filtrado de porciones "con ruido" de un genoma de referencia. En determinadas implementaciones, un proceso de filtrado comprende la obtención de puntos de datos de un conjunto de datos que se desvían de la elevación media del perfil o del nivel medio del perfil de una porción, un cromosoma o un segmento de un cromosoma en un múltiplo predeterminado de la varianza del perfil, y en determinadas implementaciones, un proceso de filtrado comprende la eliminación de puntos de datos de un conjunto de datos que no se desvían de la elevación media del perfil o del nivel medio del perfil de una porción, un cromosoma o segmento de un cromosoma en un múltiplo predeterminado de la varianza del perfil. En algunas implementaciones, se utiliza un proceso de filtrado para reducir el número de porciones candidatas de un genoma de referencia analizado para detectar la presencia o ausencia de una variación genética. Reducir el número de porciones candidatas de un genoma de referencia analizado para determinar la presencia o ausencia de una variación genética (por ejemplo, microdeleción, microduplicación) reduce a menudo 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 algunas implementaciones, una o más etapas de procesamiento pueden comprender una o más etapas de normalización. La normalización puede realizarse mediante un método descrito en el presente documento adecuado o conocido en la técnica. En determinadas implementaciones, la normalización comprende ajustar los valores medidos en diferentes escalas a una escala teóricamente común. En determinadas implementaciones, la normalización comprende un ajuste matemático sofisticado para llevar a alineación distribuciones de probabilidad de valores ajustados. En algunas implementaciones, la normalización comprende alinear distribuciones a una distribución normal. En determinadas implementaciones, la normalización comprende ajustes matemáticos que permiten la comparación de los valores normalizados correspondientes para diferentes conjuntos de datos, de manera que se eliminen los efectos de determinadas influencias macroscópicas (por ejemplo, error y anomalías). En determinadas implementaciones, la normalización comprende el escalado. La normalización comprende a veces la división de uno o más conjuntos de datos por una variable o fórmula predeterminada. Los ejemplos no limitativos de métodos de normalización incluyen normalización basada en bins o normalización basada en porciones, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, LOESS, LOESS de GC, LOWESS (suavizado de diagrama de dispersión ponderado localmente), PERUN, enmascaramiento de repetición (RM), normalización de GC y enmascaramiento de repetición (GCRM), cQn y/o combinaciones de los mismos. En algunas implementaciones, la determinación de una presencia o ausencia de una variación genética (por ejemplo, una aneuploidía) utiliza un método de normalización (por ejemplo, normalización basada en bins o basada en porciones, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, LOESS, LOESS de GC, LOWESS (suavizado de diagrama de dispersión ponderado localmente), PERUN, enmascaramiento de repetición (RM), normalización de GC y enmascaramiento de repetición (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 por regresión conocido en la técnica que combina modelos de regresión múltiple en un metamodelo basado en k vecinos más cercanos. LOESS se denomina, algunas veces, regresión polinómica ponderada localmente. LOESS de GC, en algunas implementaciones, aplica un modelo de LOESS a la relación entre el recuento de fragmentos (por ejemplo, lecturas de secuencia, recuentos) y composición de GC para porciones de un genoma

de referencia. Representar gráficamente una curva suave a través de un conjunto de puntos de datos usando LOESS se denomina, algunas veces, curva de LOESS, particularmente cuando cada valor suavizado viene dado por una regresión cuadrática de mínimos cuadrados ponderados sobre el tramo de valores de la variable de criterio de diagrama de dispersión del eje y. Para cada punto en un conjunto de datos, el método de LOESS ajusta un polinomio de bajo grado a un subconjunto de los datos, con valores de variables explicativas cerca del punto cuya respuesta se estima. El polinomio se ajusta usando mínimos cuadrados ponderados, lo que da más peso a puntos cerca del punto cuya respuesta se estima y menos peso a puntos más lejos. Después, se obtiene el valor de la función de regresión para un punto mediante la evaluación del polinomio local usando los valores de variables explicativas para ese punto de datos. Algunas veces, el ajuste de LOESS 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, tales como el grado del modelo polinómico y los pesos, son flexibles.

Puede usarse cualquier número adecuado de normalizaciones. En algunas implementaciones, 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). Los ejemplos no limitativos de tipos de normalizaciones de datos que se pueden usar incluyen normalizar los datos de recuento sin procesar para una o más pruebas o porciones de referencia seleccionadas con respecto al número total de recuentos mapeados en el cromosoma o lo totalidad del genoma en el que se mapean la porción o secciones genómicas seleccionadas; normalizar los datos de recuento sin procesar de una o varias porciones seleccionadas con respecto a la mediana de un recuento de referencia de una o varias porciones o del cromosoma en el que se mapea una o varias porciones seleccionadas; normalizar datos de recuento sin procesar con respecto a datos previamente normalizados o derivados de los mismos; y normalizar datos previamente normalizados con respecto a una o más variables de normalización predeterminadas diferentes. Normalizar 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. A veces, normalizar un conjunto de datos permite además comparar las características de datos de datos que tienen escalas diferentes, al llevar los datos a una escala común (por ejemplo, variable de normalización predeterminada). En algunas implementaciones, pueden usarse una o más normalizaciones con respecto a un valor derivado estadísticamente para minimizar las diferencias de los datos y disminuir la importancia de los datos atípicos. La normalización de porciones, o bins o porciones de un genoma de referencia, con respecto a un valor de normalización, a veces se denomina ‘‘normalización basada en bins’’ o ‘‘normalización basada en porciones’’.

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

En algunas implementaciones, la normalización con respecto a una ventana deslizante o móvil puede generar uno o más valores, donde cada valor representa la normalización con respecto a un conjunto diferente de porciones de referencia seleccionadas de regiones diferentes de un genoma (por ejemplo, cromosoma). En determinadas implementaciones, el uno o más valores generados son sumas acumulativas (por ejemplo, una estimación numérica de la integral del perfil de recuento normalizado en la porción seleccionada, dominio (por ejemplo, parte del cromosoma) o cromosoma). Los valores generados por el procedimiento de ventana deslizante o móvil pueden usarse para generar un perfil y facilitar que se llegue un resultado. En algunas implementaciones, las sumas acumulativas de una o más porciones pueden mostrarse en función de la parte. El análisis de ventana móvil o deslizante a veces se usa para analizar un genoma para determinar la presencia o ausencia de microdeleciones y/o microinserciones. En determinadas implementaciones, 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 algunas implementaciones, el análisis de ventana móvil o deslizante se usa para identificar regiones genómicas que contienen microdeleciones y, en determinadas implementaciones, 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 indicadores de ácido nucleico se denomina, en el presente documento, eliminación del error parametrizado y normalización no sesgada (PERUN) descrito en el presente documento y en la solicitud internacional de patente n.° PCT/US12/59123 (WO2013/0522913). La metodología PERUN puede aplicarse a una gran variedad de indicadores de ácido nucleico (por ejemplo, lecturas de secuencia de ácido nucleico) con el propósito de reducir los efectos de error que confunden predicciones basadas en tales indicadores.

Por ejemplo, la metodología PERUN puede aplicarse a lecturas de secuencia de ácido nucleico de una muestra y reducir los efectos de error que pueden afectar a las determinaciones de elevación de ácido nucleico (por ejemplo, determinaciones de elevación de porciones) y/o determinaciones a nivel de sección genómica. Dicha aplicación es útil para utilizar lecturas de secuencia de ácido nucleico para determinar la presencia o ausencia de una variación genética en un sujeto manifestada como una elevación o nivel variable de una secuencia de nucleótidos (por ejemplo, una porción, un nivel de sección genómica). Los ejemplos no limitativos 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 distinto de un cromosoma sexual) puede denominarse autosoma afectado. Otros ejemplos no limitativos de variaciones en las elevaciones de porciones o los niveles de sección genómica incluyen microdeleciones, microinserciones, duplicaciones y mosaicismo.

En determinadas aplicaciones, la metodología PERUN puede reducir el sesgo experimental mediante la normalización de indicadores de ácido nucleico para grupos genómicos particulares, denominándose estos últimos bins. Los bins incluyen una colección adecuada de indicadores de ácido nucleico, un ejemplo no limitativo de los mismos incluye una longitud de nucleótidos contiguos, a lo que se hace referencia en el presente documento como sección genómica o porción de un genoma de referencia. Los bins pueden incluir otros indicadores de ácido nucleico tal como se describe en el presente documento. En tales aplicaciones, la metodología PERUN normaliza generalmente los indicadores de ácido nucleico en bins particulares a través de varias muestras en tres dimensiones. En determinadas aplicaciones, la metodología PERUN puede reducir el sesgo experimental al normalizar las lecturas de ácido nucleico mapeadas en porciones particulares de un genoma de referencia, las últimas de las cuales se denominan porciones y, a veces, porciones de un genoma de referencia. En dichas aplicaciones, la metodología PERUN generalmente normaliza los recuentos de lecturas de ácidos nucleicos en porciones particulares de un genoma de referencia a través de una serie de muestras en tres dimensiones. En el apartado de ejemplos del presente documento, en la solicitud de patente internacional n.° PCT/US12/59123 (WO2013/0522913) y en la publicación de solicitud de patente US-20130085681, se proporciona una descripción detallada de PERUN y sus aplicaciones.

En determinadas implementaciones, la metodología PERUN incluye calcular una elevación de porción para cada bin a partir de una relación ajustada entre (i) sesgo experimental para un bin de un genoma de referencia en el que se mapean las lecturas de secuencia y (ii) recuentos de lecturas de secuencia mapeadas en el bin. El sesgo experimental para cada uno de los bins puede determinarse a través de múltiples muestras según una relación ajustada para cada muestra entre (i) los recuentos de lecturas de secuencia asignadas a cada uno de los bins, y (ii) una característica de mapeo para cada uno de los bins. Esta relación ajustada para cada muestra puede ensamblarse para múltiples muestras en tres dimensiones. El conjunto puede ordenarse según el sesgo experimental en determinadas implementaciones, aunque la metodología PERUN puede ponerse en práctica sin ordenar el conjunto según el sesgo experimental.

En determinadas implementaciones, 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 mapeados en una porción de un genoma de referencia para una muestra de prueba, (b) sesgo experimental (por ejemplo, sesgo de GC) para la muestra de prueba, y (c) uno o más parámetros de ajuste (por ejemplo, estimaciones de ajuste) para una relación ajustada entre (i) el sesgo experimental para una porción de un genoma de referencia en el que se mapean las lecturas de secuencia y (ii) los recuentos de secuencia lecturas mapeados en la porción. El sesgo experimental para cada una de las porciones de un genoma de referencia puede determinarse a través de múltiples muestras según una relación ajustada para cada muestra entre (i) los recuentos de lecturas de secuencia 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. Una característica de mapeo puede ser cualquier parámetro adecuado, variable y/o fuente de sesgo, cuyos ejemplos no limitativos incluyen cualquier parámetro de contenido de nucleótidos (por ejemplo, contenido de A, T, G y/o C), contenido de GC, adenina/timidina (A/T), relación de intrones/exones, contenido de intrones, contenido de exones, contenido de regiones codificantes, contenido de regiones no codificantes, contenido de secuencias repetitivas, Tf (por ejemplo, punto de fusión asociado con segmentos de un genoma), contenido de mutaciones (por ejemplo, contenido de SNP), fracción fetal, similares o combinaciones de los mismos. En algunas implementaciones, una característica de mapeo es el contenido de GC (por ejemplo, contenido de GC para una porción). Esta relación ajustada para cada muestra puede ensamblarse para múltiples muestras en tres dimensiones. El conjunto puede ordenarse según el sesgo experimental en determinadas implementaciones, aunque la metodología PERUN puede ponerse en práctica sin ordenar el conjunto según 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 función no lineal mediante un proceso de ajuste adecuado conocido en la técnica.

Una relación, como se menciona en el presente documento, es una correspondencia matemática, geométrica y/o gráfica entre dos o más parámetros (por ejemplo, valores medidos o conocidos) o variables. Una relación a veces se denomina correspondencia. Una relación puede describirse matemáticamente (por ejemplo, mediante una ecuación o fórmula) y/o

gráficamente (por ejemplo, representada como una gráfica o representada gráficamente). Una relación, o partes de ella, pueden estar asociadas con una medida de incertidumbre. En algunas implementaciones, una relación (por ejemplo, una fórmula matemática que describe una relación) comprende una o más constantes, variables y/o coeficientes. Una relación puede generarse mediante un método descrito en el presente documento o mediante un método adecuado conocido en la técnica. Puede generarse una relación en dos dimensiones para cada muestra en determinadas implementaciones, y puede seleccionarse una variable probatoria de error, o posiblemente probatoria de error, para una o más de las dimensiones. Una relación puede generarse, por ejemplo, utilizando un *software* de representación gráfica conocido en la técnica que traza una representación gráfica utilizando los valores de dos o más parámetros y/o variables proporcionados por un usuario. En determinadas implementaciones, una relación es una regresión (por ejemplo, una línea de regresión). Una relación o regresión puede ser lineal o no lineal. Una relación puede ajustarse utilizando un método adecuado descrito en el presente documento o conocido en la técnica (por ejemplo, mediante el uso de *software* de representación gráfica). Por ejemplo, una relación puede ajustarse mediante una regresión lineal y/o una regresión no lineal (por ejemplo, una función parabólica, hiperbólica o exponencial (por ejemplo, una función cuadrática)). En algunas implementaciones, una relación se ajusta según una expresión donde una o más constantes y/o coeficientes (por ejemplo, un valor de pendiente, un valor de intersección y similares) son fijos y/o predeterminados (por ejemplo, determinados a partir de otra relación). En determinadas implementaciones, una relación puede ajustarse o corregirse, por ejemplo, mediante la normalización. Por ejemplo, una relación que comprende puntos de datos puede normalizarse restando, sumando, multiplicando o dividiendo una línea de regresión. En algunas implementaciones, una relación puede comprender dos, tres, cuatro o más dimensiones. Se puede generar una relación en dos dimensiones para una o más muestras.

En la metodología PERUN, una o más de las relaciones ajustadas pueden ser lineales. Para un análisis de ácido nucleico circulante libre de células de mujeres embarazadas, donde el sesgo experimental es el 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 secuencia mapeadas en caca bin o porción, y (ii) el contenido de GC para cada bin o porción de un genoma de referencia, puede ser lineal. Para esta última relación ajustada, la pendiente pertenece a el sesgo de GC, y puede determinarse un coeficiente de sesgo de GC para cada muestra cuando las relaciones ajustadas se ensamblan a través de múltiples muestras. En dichas implementaciones, la relación ajustada para múltiples muestras y un bin o porción entre (i) el coeficiente de sesgo de GC para el bin o la porción, y (ii) los recuentos de lecturas de secuencia mapeadas en el bin o la porción, también puede ser lineal. Puede obtenerse una ordenada en el origen y pendiente a partir de esta última relación ajustada. En dichas aplicaciones, la pendiente puede abordar el sesgo específico de la muestra basado en el contenido de GC y la intersección puede abordar un patrón de atenuación específico del bin o de la porción común a todas las muestras. La metodología PERUN puede reducir significativamente dicho sesgo específico de la muestra y la atenuación específica del bin o de la porción al calcular las elevaciones de la porción o los niveles de la sección genómica para proporcionar un resultado (por ejemplo, presencia o ausencia de variación genética; determinación del sexo del feto).

En algunas implementaciones, 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 las mismas.

Ecuación A:

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

Ecuación B:

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

En algunas implementaciones, L es un nivel o perfil normalizado mediante PERUN (por ejemplo, un nivel de sección genómica normalizado, un nivel de sección genómica calculado). En algunas implementaciones, L es el resultado deseado del procedimiento de normalización PERUN. En determinadas implementaciones, L es una porción específica. En algunas implementaciones, L se determina según múltiples porciones de un genoma de referencia y representa un nivel normalizado mediante PERUN de un genoma, cromosoma, porciones o segmento del mismo. El nivel L suele utilizarse para análisis posteriores (por ejemplo, para determinar los valores de Z, las deleciones/duplicaciones maternas, las microdeleciones/microduplicaciones fetales, el sexo fetal, las aneuploidías sexuales, etc.). El método de normalización según la ecuación B se denomina eliminación de error parametrizado y normalización no sesgada (PERUN).

En algunas implementaciones de PERUN, G es un coeficiente de sesgo de GC medido utilizando un modelo lineal, LOESS, o cualquier enfoque equivalente. En algunas implementaciones, G es una pendiente. En algunas implementaciones, el coeficiente de sesgo G de GC se evalúa como la pendiente de una regresión para recuentos (por ejemplo, M, recuentos sin procesar, C_i) para la porción i y el contenido de GC de la porción i determinado a partir de un genoma de referencia. En algunas implementaciones, G representa información secundaria, extraída de M y determinada según una relación. En algunas implementaciones, G representa una relación para un conjunto de recuentos específicos de la porción y un conjunto de valores de contenido de GC específicos de la porción para una muestra (por ejemplo, una muestra de prueba). En algunas implementaciones, el contenido de GC específico de la porción se deriva de un genoma de referencia. En algunas implementaciones, el contenido de GC específico de la porción se deriva del contenido de GC observado o medido (por ejemplo, medido a partir de la muestra). A menudo se determina un coeficiente de sesgo de GC para cada muestra en un grupo de muestras y generalmente se determina para una muestra de prueba. Un coeficiente de sesgo de GC a menudo es específico de la muestra. En algunas implementaciones, un coeficiente de sesgo de GC es una constante (por ejemplo, una

vez obtenido para una muestra, no cambia) para una muestra en particular. Un "coeficiente de sesgo de GC" como se denomina en el presente documento es una estimación de la linealidad.

En algunas implementaciones, I es una intersección y S es una pendiente derivada de una relación lineal. I y S a menudo se determinan a partir de una relación para una pluralidad de muestras. En algunas implementaciones, la relación de la que se derivan I y S es diferente de la relación de la que se deriva G. En algunas implementaciones, la relación de la que se derivan I y S es fija para una configuración experimental dada. En algunas implementaciones, I y S se derivan de una relación lineal según los recuentos (por ejemplo, recuentos sin procesar, recuentos para la porción i de una muestra) y un coeficiente de sesgo de GC (por ejemplo, G determinado para una muestra) según múltiples muestras. En algunas implementaciones, I y S se derivan independientemente de la muestra de prueba. I y S a menudo son porciones específicas. En algunas implementaciones, I y S se determinan asumiendo que L = 1 para todas las porciones de un genoma de referencia en muestras euploides. En algunas implementaciones, se determina una relación lineal para muestras euploides y se determinan los valores de I y S específicos para una porción seleccionada (suponiendo que L = 1). En determinadas implementaciones, 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. Los coeficientes I y S, a los que se hace referencia en el presente documento, y como se describe para las ecuaciones (A) y (B), son estimaciones de linealidad específicas de la porción.

En algunas implementaciones, se aplica un enfoque de validación cruzada. La validación cruzada, a veces se denomina estimación de rotación. En algunas implementaciones, se aplica un enfoque de validación cruzada para evaluar con qué precisión funcionará un modelo predictivo (por ejemplo, tal como PERUN) en la práctica utilizando una muestra de prueba. En algunas implementaciones, una serie de validación cruzada comprende particionar 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 usando otro subconjunto (por ejemplo, a veces llamado conjunto de validación o conjunto de prueba). En determinadas implementaciones, se realizan múltiples series de validación cruzada usando diferentes particiones y/o diferentes subconjuntos). Los ejemplos no limitativos de enfoques de validación cruzada incluyen dejar uno fuera, bordes deslizantes, de K veces, del doble, submuestreo aleatorio repetido, similares o combinaciones de los mismos. En algunas implementaciones, 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 determinadas implementaciones, la selección aleatoria se repite 100 veces, lo que produce un conjunto de 100 pendientes y 100 intersecciones para cada porción.

En algunas implementaciones, el valor de M es un valor medido derivado de una muestra de prueba. En algunas implementaciones, M es los recuentos medidos (por ejemplo, recuentos sin procesar) para una porción. En algunas implementaciones, donde los valores I y S están disponibles para una porción, la medición de M se determina a partir de una muestra de prueba y se usa para determinar el nivel L normalizado mediante PERUN para un genoma, cromosoma, segmento o porción del mismo según la Ecuación (B).

Por tanto, la aplicación de la metodología PERUN a las lecturas de secuencia en múltiples muestras paralelas puede reducir significativamente el error provocado por (i) el sesgo experimental específico de muestra (por ejemplo, el sesgo de GC) y (ii) la atenuación específica del bin o específica de la porción común a las muestras. Otros métodos en los cuales cada una de estas dos fuentes de error se abordan a menudo por separado o en serie no pueden reducirlas tan eficazmente como la metodología PERUN. Sin desear limitarse por la teoría, se espera que la metodología PERUN reduzca el error más eficazmente en parte porque sus procedimientos generalmente aditivos no aumentan la dispersión tanto como los procedimientos generalmente multiplicativos usados en otros enfoques de normalización (por ejemplo, LOESS de GC).

Pueden usarse técnicas de normalización y estadísticas adicionales en combinación con la metodología PERUN. Puede aplicarse un procedimiento adicional antes, después y/o durante el empleo de la metodología PERUN. Se describen más adelante en el presente documento ejemplos no limitativos de procedimientos que pueden usarse en combinación con la metodología PERUN.

En algunas implementaciones, se puede utilizar una normalización secundaria o un ajuste del nivel de elevación de una porción o sección genómica para el contenido de GC junto con la metodología PERUN. Puede usarse un procedimiento de normalización o ajuste del contenido de GC adecuado (por ejemplo, LOESS de GC, GCRM). En determinadas implementaciones, puede identificarse una muestra particular para la aplicación de un procedimiento de normalización de GC adicional. Por ejemplo, la aplicación de la metodología PERUN puede determinar el sesgo de GC para cada muestra, y una muestra asociada con un sesgo de GC por encima de determinado umbral puede seleccionarse para un procedimiento adicional de normalización de GC. En tales implementaciones, puede usarse una elevación o un nivel umbral predeterminada para seleccionar tales muestras para la normalización de GC adicional.

En determinadas implementaciones, puede usarse un procedimiento de ponderación o filtrado de bins o porciones junto con la metodología PERUN. Se puede utilizar un proceso adecuado de filtrado o ponderación de bins o porciones, y los ejemplos no limitativos se describen en el presente documento, en la solicitud de patente internacional n.° PCT/US12/59123 (WO2013/0522913) y la publicación de solicitud de patente US-20130085681. En algunas implementaciones, se usa una técnica de normalización que reduce el error asociado con inserciones, duplicaciones y/o deleciones maternas (por ejemplo, variaciones del número de copias materno y/o fetal) junto con la metodología PERUN.

Las elevaciones de porción calculadas mediante la metodología PERUN pueden utilizarse directamente para proporcionar un resultado. En algunas implementaciones, las elevaciones de porción pueden usarse directamente para proporcionar un resultado para las muestras en las cuales la fracción fetal es de aproximadamente el 2 % a aproximadamente el 6 % o más (por ejemplo, fracción fetal de aproximadamente el 4 % o más). Las elevaciones de porción calculadas mediante la metodología PERUN, algunas veces, se procesan adicionalmente para proporcionar un resultado. En algunas implementaciones, las elevaciones de porción calculadas están normalizadas. En determinadas implementaciones, la suma, media o mediana de las elevaciones de porción calculadas para una porción de prueba (por ejemplo, el cromosoma 21) puede dividirse entre la suma, media o mediana de las elevaciones de porción calculadas para porciones distintas de la porción de prueba (por ejemplo, autosomas distintos del cromosoma 21), para generar una elevación de porción experimental. Una elevación de porción experimental o una elevación de porción sin procesar puede utilizarse como parte de un análisis de normalización, como el cálculo de una puntuación Z o un valor de Z. Puede generarse una puntuación Z para una muestra restando una elevación de porción esperada de una elevación de porción experimental o de porción sin procesar, y el valor resultante puede dividirse por una desviación estándar para las muestras. Las puntuaciones Z resultantes pueden distribuirse para diferentes muestras y analizarse, o pueden relacionarse con otras variables, tales como fracción fetal y otras, y analizarse, para proporcionar un resultado, en determinadas implementaciones.

Los niveles de sección genómica calculadas mediante la metodología PERUN pueden utilizarse directamente para proporcionar un resultado. En algunas implementaciones, los niveles de sección genómica pueden usarse directamente para proporcionar un resultado para las muestras en las cuales la fracción fetal es de aproximadamente el 2 % a aproximadamente el 6 % o más (por ejemplo, fracción fetal de aproximadamente el 4 % o más). Los niveles de sección genómica calculadas mediante la metodología PERUN, algunas veces, se procesan adicionalmente para proporcionar un resultado. En algunas implementaciones, los niveles de sección genómica calculadas están normalizadas. En determinadas implementaciones, la suma, media o mediana de los niveles de sección genómica calculadas para una porción de prueba (por ejemplo, el cromosoma 21) puede dividirse entre la suma, media o mediana de los niveles de sección genómica calculados para porciones distintas de la porción de prueba (por ejemplo, autosomas distintos del cromosoma 21), para generar un nivel de sección genómica experimental. Un nivel de sección genómica experimental o un nivel de sección genómica sin procesar pueden utilizarse como parte de un análisis de normalización, tal como el cálculo de una puntuación Z o la puntuación Z. Puede generarse una puntuación Z para una muestra restando un nivel de sección genómica esperado de un nivel de sección genómica experimental o de un nivel de sección genómica sin procesar, y el valor resultante puede dividirse entre una desviación estándar para las muestras. Las puntuaciones Z resultantes pueden distribuirse para diferentes muestras y analizarse, o pueden relacionarse con otras variables, tales como fracción fetal y otras, y analizarse, para proporcionar un resultado, en determinadas implementaciones.

Tal como se indica en el presente documento, la metodología PERUN no se limita a la normalización según 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 limitativo de una fuente de sesgo distinto del contenido de GC es la capacidad de mapeo. Cuando se abordan parámetros de normalización distintos del sesgo y el contenido de GC, una o más de las relaciones ajustadas pueden ser no lineales (por ejemplo, hiperbólica, exponencial). Cuando el sesgo experimental se determina a partir de una relación no lineal, por ejemplo, una estimación de la curvatura del sesgo experimental puede analizarse en algunas implementaciones.

La metodología PERUN puede aplicarse a una gran variedad de indicadores de ácido nucleico. Los ejemplos no limitativos de indicadores de ácido nucleico son lecturas de secuencia de ácido nucleico y elevaciones o niveles de ácido nucleico en una ubicación particular en un microalineamiento. Los ejemplos no limitativos 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 puede aplicarse a lecturas de secuencia mapeadas en secuencias de referencia adecuadas, tales como ADN genómico de referencia, ARN celular de referencia (por ejemplo, transcriptoma) y porciones de los mismos (por ejemplo, parte(s) de un complemento genómico de transcriptoma de ADN o ARN, parte(s) de un cromosoma).

Por tanto, en determinadas implementaciones, un ácido nucleico celular (por ejemplo, ADN o ARN) puede servir como indicador de ácido nucleico. Las lecturas de ácidos nucleicos celulares mapeadas en porciones del genoma de referencia pueden normalizarse usando la metodología PERUN.

El ácido nucleico celular es, en algunas implementaciones, una asociación con una o más proteínas, y un agente que captura ácido nucleico asociado a proteínas puede usarse para enriquecer este último, en algunas implementaciones. Un agente en determinados casos es un anticuerpo o fragmento de anticuerpo que se une de manera específica a una proteína en asociación con ácido nucleico celular (por ejemplo, un anticuerpo que se une de manera específica a una proteína cromatina (por ejemplo, proteína histona)). Los procedimientos en los cuales se usa un anticuerpo o fragmento de anticuerpo para enriquecer el ácido nucleico celular unido a una proteína particular a veces se denominan procedimientos de inmunoprecipitación de cromatina (ChIP). El ácido nucleico enriquecido mediante ChIP es un ácido nucleico en asociación con una proteína celular, tal como ADN o ARN, por ejemplo. Las lecturas de ácido nucleico enriquecido mediante ChIP pueden obtenerse con el uso de tecnología conocida en la técnica. Las lecturas de ácido nucleico enriquecido mediante ChIP pueden mapearse en una o más porciones de un genoma de referencia, y los resultados pueden normalizarse usando la metodología PERUN para proporcionar un resultado.

Por tanto, en determinadas implementaciones, también se proporcionan métodos para calcular con sesgo reducido elevaciones de porción para una muestra de prueba, que comprenden: (a) obtener recuentos de lecturas de secuencia mapeadas en bins de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico celular de una muestra de prueba obtenida por aislamiento de una proteína a la que se asoció el ácido nucleico; (b) determinar el sesgo experimental para cada uno de los bins a través de múltiples muestras a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins, y (ii) una característica de mapeo para cada uno de los bins; y (c) calcular una elevación de porción para cada uno de los bins a partir de una relación ajustada entre el sesgo experimental y los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins, proporcionando de este modo elevaciones de porción calculadas, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins se reduce en las elevaciones de porción calculadas.

En determinadas implementaciones, un ARN celular puede servir como indicadores de ácido nucleico. Las lecturas de ARN celular pueden mapearse en porciones de ARN de referencia y normalizar con el uso de la metodología PERUN para proporcionar un resultado. Las secuencias conocidas para ARN celular, denominado transcriptoma, o un segmento del mismo, pueden usarse como una referencia en la que pueden mapearse lecturas de ARN a partir 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 con el uso de la metodología PERUN para proporcionar un resultado.

Por tanto, en algunas implementaciones, también se proporcionan métodos para calcular con sesgo reducido elevaciones de porciones para una muestra de prueba, que comprenden: (a) obtener recuentos de lecturas de secuencia mapeadas en bins de ARN de referencia (por ejemplo, transcriptoma de referencia o segmento(s) del mismo), lecturas de secuencia que son lecturas de ARN celular de una muestra de prueba; (b) determinar el sesgo experimental para cada uno de los bins a través de múltiples muestras a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins, y (ii) una característica de mapeo para cada uno de los bins; y (c) calcular una elevación de porción para cada uno de los bins a partir de una relación ajustada entre el sesgo experimental y los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins, proporcionando de este modo elevaciones de porción calculadas, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada uno de los bins se reduce en las elevaciones de porción calculadas.

En algunas implementaciones, los niveles de ácido nucleico de microalineamiento pueden servir como indicadores de ácido nucleico. Los niveles de ácido nucleico a través de las muestras para una dirección particular o ácido nucleico de hibridación en un alineamiento pueden analizarse con el uso de la metodología PERUN, normalizando de ese modo los indicadores de ácido nucleico proporcionados por el análisis de microalineamiento. De esta manera, una dirección particular o ácido nucleico de hibridación en una microalineamiento es análogo a un bin o una porción para lecturas de secuencia de ácido nucleico mapeadas, y puede usarse la metodología PERUN para normalizar los datos de microalineamiento para proporcionar un resultado mejorado.

Por tanto, en determinadas implementaciones se proporcionan métodos para reducir el error de nivel de ácido nucleico de microalineamiento para una muestra de prueba, que comprenden: (a) obtener niveles de ácido nucleico en una micromatriz a la que se ha asociado el ácido nucleico de la muestra de prueba, micromatriz que incluye una matriz de ácidos nucleicos de captura; (b) determinar el sesgo experimental para cada uno de los ácidos nucleicos de captura en múltiples muestras a partir de una relación ajustada entre (i) los niveles de ácido nucleico de la muestra de prueba asociados con cada uno de los ácidos nucleicos de captura, y (ii) una característica de asociación para cada uno de los ácidos nucleicos de captura; y (c) calcular un nivel de ácido nucleico de la muestra de prueba para cada uno de los ácidos nucleicos de captura a partir de una relación ajustada entre el sesgo experimental y los niveles del ácido nucleico de la muestra de prueba asociado con cada uno de los ácidos nucleicos de captura, proporcionando de este modo niveles calculados, mediante los cuales el sesgo en los niveles de ácido nucleico de muestra de prueba asociado con cada uno de los ácidos nucleicos de captura se reduce en los niveles calculados. La característica de asociación mencionada anteriormente puede ser cualquier característica correlacionada con la hibridación de un ácido nucleico de muestra de prueba con un ácido nucleico de captura que da lugar a, o puede dar lugar a, error en la determinación del nivel de ácido nucleico de muestra de prueba asociado con un ácido nucleico de captura.

En algunas implementaciones, una etapa de procesamiento comprende una ponderación, como se ha descrito anteriormente. Los términos “ponderado” y “ponderación” y la expresión “función de ponderación” o sus derivados gramaticales o equivalentes, tal y como se utilizan en el presente documento, se refieren a una manipulación matemática de una parte o de la totalidad de un conjunto de datos que a veces se utiliza para alterar la influencia de determinadas 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 bins, en función de la calidad o la utilidad de los datos del bin o bins seleccionados). Por ejemplo, los bins con datos de secuencias poco representados o de baja calidad pueden “ponderarse a la baja” para minimizar la influencia en un conjunto de datos, mientras que los bins seleccionados pueden “ponderarse al alza” para aumentar la influencia en un conjunto de datos.

En determinadas implementaciones, 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. Puede usarse

cualquier número adecuado de manipulaciones matemáticas y/o estadísticas. En algunas implementaciones, un conjunto de datos puede manipularse matemática y/o estadísticamente 1 o más, 5 o más, 10 o más o incluso 20 o más veces. Los ejemplos no limitativos de manipulaciones matemáticas y estadísticas que pueden usarse incluyen 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 elevaciones o niveles de pico, determinación de ubicaciones de bordes de pico, cálculo de razones de áreas de pico, análisis de la mediana de elevación o nivel cromosómico, cálculo de la desviación media absoluta, suma de residuos al cuadrado, media, desviación estándar, error estándar, similares o combinaciones de los mismos. Puede realizarse una manipulación matemática y/o estadística en todos o en una parte de los datos de lectura de secuencia o productos procesados de los mismos. Los ejemplos no limitativos de variables o características del conjunto de datos que pueden manipularse estadísticamente incluyen recuentos sin procesar, recuentos filtrados, recuentos normalizados, alturas de pico, anchuras de pico, áreas de pico, bordes de pico, tolerancias laterales, valores de *p*, mediana de elevaciones o niveles, elevaciones o niveles medios, distribución de recuentos dentro de una región genómica, representación relativa de especies de ácido nucleico, similares o combinaciones de los mismos.

En algunas implementaciones, 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. Puede usarse cualquier número adecuado de algoritmos estadísticos. En algunas implementaciones, un conjunto de datos puede analizarse usando 1 o más, 5 o más, 10 o más o incluso 20 o más algoritmos estadísticos. Los ejemplos no limitativos de algoritmos estadísticos adecuados para su uso con los métodos descritos en el presente documento incluyen árboles de decisión, valores contranulos, comparaciones múltiples, prueba ómnibus, problema de Behrens-Fisher, remuestreo de tipo *bootstrapping*, método de Fisher para combinar pruebas independientes de significación, hipótesis nula, error tipo I, error tipo II, prueba exacta, prueba Z de una muestra, prueba Z de dos muestras, prueba de la t de una muestra, prueba de la t para datos emparejados, prueba de la *t* agrupada de dos muestras que tienen varianzas iguales, prueba de la t no agrupada de dos muestras que tienen varianzas desiguales, prueba z de una proporción, prueba z de dos proporciones agrupadas, prueba z de dos proporciones no agrupadas, prueba de chi cuadrado de una muestra, prueba F de dos muestras para determinar la igualdad de varianzas, intervalo de confianza, intervalo creíble, significación, metaanálisis, regresión lineal simple, regresión lineal robusta, similares o combinaciones de los anteriores. Los ejemplos no limitativos de variables o características del conjunto de datos que pueden analizarse usando algoritmos estadísticos incluyen recuentos sin procesar, recuentos filtrados, recuentos normalizados, alturas de pico, anchuras de pico, bordes de pico, tolerancias laterales, valores de *p*, mediana de elevaciones o niveles, elevaciones o niveles medios, distribución de recuentos dentro de una región genómica, representación relativa de especies de ácido nucleico, similares o combinaciones de los mismos.

En determinadas implementaciones, un conjunto de datos puede analizarse utilizando algoritmos estadísticos múltiples (por ejemplo, 2 o más) (por ejemplo, regresión por mínimos cuadrados, análisis de componentes principales, análisis discriminante lineal, análisis discriminante cuadrático, agregación de tipo *bootstrap*, redes neurales, modelos de máquinas de vectores de soporte, bosques aleatorios, modelos de árboles de clasificación, K vecinos más cercanos, regresión logística y/o suavizado de pérdida) y/o manipulaciones matemáticas y/o estadísticas (por ejemplo, a las que se hace referencia en el presente documento como manipulaciones). El uso de múltiples manipulaciones puede generar un espacio N-dimensional que puede usarse para proporcionar un resultado, en algunas implementaciones. En determinadas implementaciones, 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 múltiples manipulaciones en un conjunto de datos de referencia puede generar un espacio N-dimensional (por ejemplo, gráfico 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 las muestras de prueba usando un conjunto sustancialmente similar de manipulaciones puede usarse para generar un punto N-dimensional para cada una de las muestras de prueba. A veces, la complejidad y/o dimensionalidad de un conjunto de datos de un sujeto de prueba se reduce a un solo valor o punto N-dimensional que puede compararse fácilmente con el espacio N-dimensional generado a partir de los datos de referencia. Los datos de muestra de prueba que se encuentran dentro del espacio N-dimensional poblado por los datos del sujeto de referencia son indicativos de un estado genético prácticamente similar al de los sujetos de referencia. Los datos de muestra de prueba 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 algunas implementaciones, las referencias son euploides o no tienen de cualquier otra manera una variación genética o afección médica.

Después de que los conjuntos de datos se han contado, opcionalmente filtrado y normalizado, los conjuntos de datos procesados pueden manipularse adicionalmente mediante uno o más procedimientos de filtrado y/o normalización, en algunas implementaciones. Un conjunto de datos que se ha manipulado adicionalmente mediante uno o más procedimientos de filtrado y/o normalización puede usarse para generar un perfil, en determinadas implementaciones. El uno o más procedimientos de filtrado y/o normalización a veces pueden reducir la complejidad y/o dimensionalidad del conjunto de datos, en algunas implementaciones. Puede proporcionarse un resultado basado en un conjunto de datos de complejidad y/o dimensionalidad reducidas.

Después de que los conjuntos de datos se han contado, opcionalmente filtrado, normalizado y, opcionalmente, ponderado, los conjuntos de datos procesados pueden manipularse mediante una o más manipulaciones matemáticas y/o estadísticas (por ejemplo, funciones estadísticas o algoritmo estadístico), en algunas implementaciones. En

determinadas implementaciones, los conjuntos de datos procesados pueden manipularse adicionalmente mediante el cálculo de puntuaciones Z para una o más porciones, cromosomas o porciones de cromosomas seleccionados. En algunas implementaciones, los conjuntos de datos procesados pueden manipularse adicionalmente mediante el cálculo de valores de p. En la Ecuación 1 (Ejemplo 2) se presenta una implementación de una ecuación para calcular una puntuación Z y un valor de *p*. En determinadas implementaciones, las manipulaciones matemáticas y/o estadísticas incluyen una o más suposiciones que pertenecen a ploidía y/o fracción fetal. En algunas implementaciones, se genera un gráfico de perfiles de datos procesados manipulados adicionalmente mediante una o más manipulaciones estadísticas y/o matemáticas para facilitar la clasificación y/o proporcionar un resultado. Puede proporcionarse un resultado basado en un gráfico de perfiles de datos manipulados estadística y/o matemáticamente. Un resultado proporcionado basado en un gráfico de perfiles de datos manipulados estadística y/o matemáticamente incluye a menudo una o más suposiciones que pertenecen a ploidía y/o fracción fetal.

En determinadas implementaciones, se realizan múltiples manipulaciones en conjuntos de datos procesados para generar un espacio N-dimensional y/o un punto N-dimensional, después de que los conjuntos de datos se han contado, opcionalmente, filtrado y normalizado. Puede proporcionarse un resultado basado en un gráfico de perfiles de conjuntos de datos analizados en N dimensiones.

En algunas implementaciones, los conjuntos de datos se procesan usando uno o más análisis de elevación o nivel de pico, análisis de anchura de pico, análisis de ubicación de borde de pico, tolerancias laterales de pico, similares, derivaciones de los mismos o combinaciones de los anteriores, como parte de o después de procesar y/o manipular los conjuntos de datos. En algunas implementaciones, se genera una representación gráfica de perfiles de datos procesados usando uno o más análisis de elevación o nivel de pico, análisis de anchura de pico, análisis de ubicación de borde de pico, tolerancias laterales de pico, similares, derivaciones de los mismos o combinaciones de los anteriores para facilitar la clasificación y/o proporcionar un resultado. Puede proporcionarse un resultado basado en una representación gráfica de perfiles de datos que se procesaron usando uno o más análisis de elevación o nivel de pico, análisis de anchura de pico, análisis de ubicación de borde de pico, tolerancias laterales de pico, similares, derivaciones de los mismos o combinaciones de los anteriores.

En algunas implementaciones, el uso de una o más muestras de referencia que están prácticamente exentas de una variación genética en cuestión puede usarse para generar una mediana de perfil de recuento de referencia, 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 las áreas correspondientes a la ubicación genómica en la cual la variación genética está ubicada en el sujeto de prueba, si el sujeto de prueba presentase la variación genética. En los sujetos de prueba en riesgo de, o que padecen, una afección médica asociada con una variación genética, se espera que el valor numérico para la porción o secciones seleccionadas varíe significativamente con respecto al valor predeterminado para ubicaciones genómicas no afectadas. En determinadas implementaciones, el uso de una o más muestras de referencia que se sabe que portan la variación genética en cuestión puede usarse para generar una mediana de perfil de recuento 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 las áreas correspondientes a la ubicación genómica en la que un sujeto de prueba no porta la variación genética. En sujetos de prueba que no están en riesgo de o que padecen una afección médica asociada con una variación genética, se espera que el valor numérico para la porción o secciones seleccionadas varíe significativamente con respecto al valor predeterminado para las ubicaciones genómicas afectadas.

En algunas implementaciones, el análisis y procesamiento de datos pueden incluir el uso de una o más suposiciones. Puede usarse un número o tipo adecuado de suposiciones para analizar o procesar un conjunto de datos. Los ejemplos no limitativos de suposiciones que pueden usarse para el procesamiento y/o análisis de datos incluyen ploidía materna, contribución fetal, prevalencia de determinadas secuencias en una población de referencia, origen étnico, prevalencia de una afección médica seleccionada en miembros de la familia emparentados, paralelismo entre perfiles de recuento sin procesar de diferentes pacientes y/o ejecuciones después de la normalización de GC y enmascaramiento de repetición (por ejemplo, GCRM), las coincidencias idénticas representan artefactos de PCR (por ejemplo, posición de base idéntica), las suposiciones inherentes en un ensayo cuantificador fetal (por ejemplo, FQA), las suposiciones con respecto a gemelos (por ejemplo, si hay 2 gemelos y solo 1 se ve afectado, la fracción fetal efectiva es solo el 50 % de la fracción fetal total medida (de manera similar para trillizos, cuatrillizos y similares)), el ADN fetal libre de células (por ejemplo, ADNlc) cubre uniformemente todo el genoma, similares y combinaciones de los mismos.

En los casos en donde la calidad y/o profundidad de las lecturas de secuencia mapeadas no permite una predicción de resultados de la presencia o ausencia de una variación genética a un nivel de confianza deseado (por ejemplo, del 95 % o mayor nivel de confianza), basándose en los perfiles de recuento normalizados, pueden usarse 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 proporcionar un resultado. La expresión "perfil de recuento normalizado", tal como se usa en el presente documento, se refiere a un perfil generado usando recuentos normalizados. Los ejemplos de métodos que pueden usarse para generar recuentos normalizados y perfiles de recuento normalizados se describen en el presente documento. Tal como se indicó, las lecturas de secuencia mapeadas que se han contado pueden normalizarse con respecto a los recuentos de muestras de prueba o recuentos de muestras de referencia. En algunas implementaciones, un perfil de recuento normalizado puede presentarse como una representación gráfica.

Métodos adicionales de PERUN

En la metodología PERUN, una o más relaciones ajustadas pueden ser lineales y/o no lineales. En determinadas implementaciones, se puede ajustar una relación lineal y/o una relación no lineal a una función no lineal. En algunas implementaciones, una relación lineal y/o una relación no lineal se define y/o describe mediante una función no lineal. Una normalización PERUN puede comprender una relación que se puede ajustar a una función no lineal adecuada. Los ejemplos no limitativos de una función no lineal que se puede utilizar para un enfoque de normalización PERUN incluyen una función polinomial; una función racional; una función trascendental; una combinación lineal de funciones exponenciales; una función exponencial de un polinomio (por ejemplo, una función cuadrática); un producto de una función exponencialmente decreciente y una función logarítmica (por ejemplo, exp(-x)log(1+x)); un producto de una función exponencialmente decreciente y un polinomio; una función trigonométrica; una combinación lineal de funciones trigonométricas; o combinación de las anteriores.

En la metodología PERUN, una relación puede ajustarse a una función según una o más estimaciones de la curvatura. Una estimación de la curvatura, como se menciona en el presente documento, puede ser un coeficiente (por ejemplo, un coeficiente de regresión) y/o una constante. Por ejemplo, una estimación de la curvatura puede ser uno o más coeficientes que definen, en parte, una función (por ejemplo, una expresión, una expresión cuadrática). Una estimación de la curvatura puede determinarse mediante un método adecuado descrito en el presente documento o conocido en la técnica. En algunas implementaciones, una estimación de la curvatura se determina según y/o se define mediante una relación. Por ejemplo, se pueden determinar una o más estimaciones de la curvatura ajustando una regresión no lineal (por ejemplo, una expresión cuadrática) a un conjunto de datos que da como resultado una expresión cuadrática que define una línea de regresión ajustada. En algunas implementaciones, una o más de las estimaciones de la curvatura determinadas para la nueva expresión cuadrática pueden mantenerse constantes, y se pueden utilizar la nueva expresión cuadrática y las estimaciones de la curvatura para 1) normalizar otro conjunto de datos o 2) ajustar un nuevo conjunto de datos a otra función. Una relación se puede ajustar a una función que comprende 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 o 10 o más estimaciones de la curvatura. En algunas implementaciones, el número de estimaciones de la curvatura para una función se determina según el grado, la dimensión y/o el orden de una función. Una estimación de la curvatura puede ser específica de la muestra o independiente de la muestra (por ejemplo, se aplica a una o más muestras). Una estimación de la curvatura puede ser específica de la porción o independiente de la porción.

En algunas implementaciones, una estimación de la curvatura es específica de la muestra e independiente de la porción. Se puede determinar una estimación de la curvatura según cualquier muestra adecuada. En algunas implementaciones, se determina una estimación de la curvatura según una muestra de prueba. En algunas implementaciones, se determina una estimación de la curvatura según múltiples muestras (por ejemplo, muestras de referencia o muestras de prueba y muestras de referencia).

En determinadas implementaciones de la normalización PERUN, se determina un nivel de sección genómica normalizado, en parte, según una o más estimaciones de la curvatura específicas de la muestra. En algunas implementaciones, se determinan una o más estimaciones de la curvatura específicas de la muestra para una relación ajustada entre (i) los recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, y (ii) una función de mapeo para cada una de las porciones de un genoma de referencia. En determinadas implementaciones, se determinan una o más estimaciones de la curvatura específicas de la muestra para una relación ajustada entre (i) los recuentos de porciones de un genoma de referencia y (ii) una característica de mapeo para cada una de las porciones donde la característica de mapeo es el contenido de GC. En determinadas implementaciones, se determinan una o más estimaciones de la curvatura específicas de la muestra para una relación ajustada que resulta del ajuste a una función no lineal adecuada. Los ejemplos no limitativos de una función no lineal que se puede usar para determinar estimaciones de la curvatura específicas de la muestra incluyen una función polinomial; una función racional; una función trascendental; una combinación lineal de funciones exponenciales; una función exponencial de un polinomio (por ejemplo, una función cuadrática); un producto de una función exponencialmente decreciente y una función logarítmica (por ejemplo, exp(-x)log(1+x)); un producto de una función exponencialmente decreciente y un polinomio; una función trigonométrica; una combinación lineal de funciones trigonométricas; o combinación de las anteriores. Una función no lineal puede ser una función exponencial de un polinomio, por ejemplo, una función cuadrática o una función semicuadrática. Una función semicuadrática a veces es una función cuasicuadrática o una versión canónica de una función cuadrática de PERUN. En algunas implementaciones, se determinan una o más estimaciones de la curvatura específicas de la muestra según una función cuadrática representada por la siguiente ecuación (30):

$$c_i = G_0 + G_1 g_i + G_2 g_i^2 \qquad (30),$$

donde c_i es el recuento para la porción i (por ejemplo, el recuento sin procesar, por ejemplo, una representación del recuento), g_i es el contenido de GC de la porción i y G_0, G_1 y G_2 son las estimaciones de la curvatura (por ejemplo, de orden cero, primero y segundo respectivamente). En algunas implementaciones G_0, G_1 y G_2 son estimaciones de la curvatura específicas de la muestra. Puede utilizarse cualquier método de regresión adecuado para determinar una estimación de la curvatura (por ejemplo, un valor de una estimación de la curvatura). En algunas

implementaciones, las estimaciones de la curvatura (por ejemplo, estimaciones de la curvatura específicas de la muestra y/o específicas de la porción) se determinan, en parte, mediante un proceso de optimización adecuado.

La implementación de la ecuación (30) comprende un polinomio de segundo orden. La función cuadrática representada por la ecuación (30) puede extenderse a otras formas funcionales adecuadas y/o a funciones polinómicas de orden superior. En algunas implementaciones, la ecuación (30) se puede generalizar como la siguiente ecuación 31:

$$c_i = \sum_{m=0}^{N} G_n g_i^n \qquad (31),$$

donde c_i es el recuento para la porción i (por ejemplo, el recuento sin procesar, por ejemplo, una representación del recuento), g_i es el contenido de GC de la porción i, N representa el nivel de truncamiento y G representa una estimación específica de la muestra de la curvatura de orden n.

En determinadas implementaciones de la normalización PERUN, se determina (por ejemplo, se calcula) un nivel de sección genómica normalizado, en parte, según una o más estimaciones de la curvatura específicas de la porción (por ejemplo, m_0, m_1 y m_2 en la siguiente ecuación 32). Las estimaciones de la curvatura específicas de la porción a veces son independientes de la muestra. En algunas implementaciones, las estimaciones de la curvatura específicas de la porción se derivan de una pluralidad de muestras para una porción de un genoma de referencia. Una pluralidad de muestras puede ser aproximadamente 50 o más, 100 o más, 500 o más, 1000 o más o 10.000 o más muestras. Las estimaciones de la curvatura específicas de la porción se pueden determinar según las estimaciones de la curvatura específicas de la muestra (por ejemplo, G_0, G_1 y G_2). En determinadas implementaciones, las estimaciones de la curvatura específicas de la porción se determinan según una o más funciones que describen una relación entre una o más estimaciones de la curvatura específicas de la muestra (por ejemplo, G_0, G_1 y G_2) y recuentos específicos de porciones (por ejemplo, c_i) para una porción seleccionada i para una pluralidad de muestras. Las estimaciones de la curvatura específicas de la porción pueden determinarse a partir de una o más relaciones ajustadas entre las estimaciones de la curvatura específicas de la muestra (por ejemplo, G_0, G_1 y G_2) y recuentos específicos de la porción (por ejemplo, recuentos de una porción de un genoma de referencia) determinados para una pluralidad de muestras. En algunas implementaciones, se determinan una o más estimaciones de la curvatura específicas de la porción para una o más relaciones ajustadas que resultan del ajuste a una función no lineal adecuada. Los ejemplos no limitativos de una función no lineal que se puede usar para determinar estimaciones de la curvatura específicas de la porción incluyen una función polinomial; una función racional; una función trascendental; una combinación lineal de funciones exponenciales; una función exponencial de un polinomio (por ejemplo, una función cuadrática); un producto de una función exponencialmente decreciente y una función logarítmica (por ejemplo, exp(-x)log(1+x)); un producto de una función exponencialmente decreciente y un polinomio; una función trigonométrica; una combinación lineal de funciones trigonométricas; o combinación

de los anteriores. En algunas implementaciones, las estimaciones de la curvatura específicas de la porción se determinan según la siguiente ecuación (32):

$$c_i = G_0 + m_0 + G_1 m_1 + G_2 m_2 \qquad (32),$$

donde c_i es el recuento para la porción i (por ejemplo, el recuento sin procesar, por ejemplo, una representación del recuento), G_0, G_1 y G_2 son estimaciones de la curvatura específicas de la muestra (por ejemplo, de cero, primer y segundo orden, respectivamente) y m_0, m_1 y m_2 son estimaciones de la curvatura específicas de la porción (por ejemplo, de cero, primer y segundo orden, respectivamente). En algunas implementaciones m_0, m_1 y m_2 son estimaciones de la curvatura independientes de la muestra. En algunas implementaciones m_0, m_1 y m_2 se derivan de la ecuación (32) para múltiples valores de c_i, G_0, G_1 y G_2 determinados a partir de una pluralidad de muestras. Algunas veces, en el presente documento la ecuación (32) se denomina PERUN cuadrática.

En algunas implementaciones, una función lineal puede describir una relación entre G_1 y G_2 obtenida a partir de una pluralidad de muestras. En determinadas implementaciones, una relación entre G_1 y G_2 se determina según la siguiente ecuación (33):

$$G_2 = K_0 + G_1 K_2 \qquad (33),$$

donde K_0 y K_2 son estimaciones de la curvatura que definen una relación lineal entre G_1 y G_2. En algunas implementaciones, K_0 y K_2 son coeficientes de regresión lineal que definen una relación lineal entre G_1 y G_2. En algunas implementaciones, K_0 y K_2 se obtienen de una regresión lineal de G_1 y G_2 para una pluralidad de muestras.

En algunas implementaciones, las estimaciones de la curvatura específicas de la porción se determinan según la siguiente ecuación (34):

$$c_i = G_0 + m_0 + G_1 m_1 + (K_0 + G_1 K_2) m_2 \qquad (34).$$

A la ecuación (34) a veces se la denomina PERUN cuasicuadrática. En algunas implementaciones, la ecuación (34) se puede expresar como la siguiente ecuación (35):

$$c_i = G_0 + a_0 + G_1 a_1 \qquad (35),$$

donde,

$$a_0 = m_0 + K_0 m_2 \qquad (36),$$

y,

$$a_1 = m_1 + K_2 m_2 \qquad (37).$$

En algunas implementaciones, las estimaciones de la curvatura específicas de la porción se determinan según la siguiente ecuación (38):

$$c_i = G_0 + m_0 + G_1 m_1 + (K_0 + G_1 K_2) m_2 = G_0 + a_0 + G_1 a_1 \qquad (38)$$

En algunas implementaciones, las estimaciones de la curvatura G_1 y G_2 se transforman en un conjunto de coordenadas generalizadas X_1 y X_2 utilizando una transformación canónica (por ejemplo, véase el Ejemplo 4). En algunas implementaciones, las coordenadas canónicas X_1 y X_2 se obtienen como elementos de vectores propios de una matriz de covarianza y se dividen entre las raíces cuadradas de los valores propios correspondientes. En determinadas implementaciones, las coordenadas canónicas X_1 y X_2 se utilizan a continuación para definir una versión canónica de PERUN cuadrático que se ilustra con la siguiente ecuación (39):

$$c_i = G_0 + \mu_0 + X_1 \mu_1 + X_2 \mu_2 \qquad (39),$$

en donde c_i es los recuentos de una porción i.° de un genoma de referencia; G_0 es una estimación de la curvatura específica de la muestra; X_1 y X_2 son coordenadas canónicas; μ_0, μ_1 y μ_2 están relacionados con el conjunto cuadrático de parámetros m_0, M_1 y m_2 por una inversa de una transformación de coordenadas lineales utilizada para generar coordenadas canónicas X_1 y X_2; y m_0, m_1 y m_2 son estimaciones de la curvatura específicas de la porción. En algunas implementaciones μ_0, μ_1 y μ_2 se determinan aplicando una regresión lineal a un gran conjunto de muestras de referencia.

En algunas implementaciones, estimaciones de la curvatura específicas de la porción (por ejemplo, m_0, m_1 y m_2), coeficientes de regresión lineal X_1 y X_2, y/o parámetros μ_0, μ_1 y μ_2 están determinados, en parte, por un proceso de optimización adecuado. Los ejemplos no limitativos de un proceso de optimización que se puede utilizar incluyen un proceso simple cuesta abajo; proceso de búsqueda o bisección de extremos y relación áurea; un proceso de interpolación parabólica; un proceso de gradientes conjugados; un proceso del mayor descenso newtoniano; un proceso Broyden-Fletcher-Goldfarb-Shanno (BFGS); una versión de base limitada de un proceso BFGS; un proceso del mayor descenso cuasinewtoniano; un proceso de hibridación simulada; un proceso de metrópolis de MonteCarlo; un proceso de toma de muestras de Gibbs; un proceso de algoritmo E-M; o combinación de las anteriores. Un proceso simple cuesta abajo a veces se denomina proceso de Nelder/Mead o proceso ameba. En algunas implementaciones, las estimaciones de la curvatura específicas de la porción (por ejemplo, m_0, m_1 y m_2), los coeficientes de regresión lineal X_1 y X_2, y/o los parámetros μ_0, μ_1 y μ_2 se determinan según una relación ajustada entre (i) recuentos de lecturas de secuencia mapeadas en las porciones del genoma de referencia y (ii) una característica de mapeo para cada una de las porciones del genoma de referencia obtenidas de múltiples muestras y un proceso de optimización. En algunas implementaciones, las estimaciones de la curvatura específicas de la porción (por ejemplo, m_0, m_1 y m_2), los coeficientes de regresión lineal X_1 y X_2, y/o los parámetros μ_0, μ_1 y μ_2 se determinan según una relación ajustada entre (i) recuentos específicos de la porción de lecturas de secuencia mapeadas en porciones de un genoma de referencia, y (ii) estimaciones de la curvatura específicas de la muestra obtenidas de múltiples muestras y un proceso de optimización. En algunas implementaciones, las estimaciones de la curvatura específicas de la porción (por ejemplo, m_0, m_1 y n_2), los coeficientes de regresión lineal X_1 y X_2, y/o los parámetros μ_0, μ_1 y μ_2 se determinan directa o indirectamente según la ecuación (32), una pluralidad de muestras y un proceso de optimización.

En algunas implementaciones de una normalización PERUN, se determinan los niveles de sección genómica normalizados para una muestra de prueba, según los recuentos medidos para una porción de un genoma de referencia, las estimaciones de la curvatura específicas de la muestra y las estimaciones de la curvatura específicas de la porción. En algunas implementaciones, los niveles de sección genómica normalizados se determinan para una muestra de prueba, según la siguiente ecuación (40):

$$l_i = \frac{c_i}{G_0 + m_0 + G_1 m_1 + G_2 m_2} \quad (40),$$

donde l_i es un nivel de sección genómica normalizado para la porción i, c_i son los recuentos (por ejemplo, representación de recuento) para la porción i, G_0, G_1 y G_2 son estimaciones de la curvatura específicas de la muestra (por ejemplo, de cero, primer y segundo orden respectivamente) y m_0, m_1 y m_2 son estimaciones de la curvatura específicas de la porción (por ejemplo, de orden cero, primer y segundo orden, respectivamente) para la porción i. La ecuación (40) es un ejemplo de una implementación de una normalización PERUN cuadrática.

En algunas implementaciones, se determinan los niveles de sección genómica normalizados para una porción de una muestra de prueba, según la siguiente ecuación (41):

$$l_i = \frac{c_i}{G_0 + a_0 + G_1 a_1} \quad (41),$$

donde l_i es un nivel de sección genómica normalizado para la porción i, c_i son los recuentos (por ejemplo, representación del recuento) para la porción i, G_0, y G_1 son estimaciones de la curvatura específicas de la muestra y a_0 y a_1 son estimaciones de la curvatura específicas de la porción definidas según la ecuación (36) y (37), respectivamente. La ecuación (41) es un ejemplo de una implementación de una normalización PERUN cuasicuadrática.

En algunas implementaciones, los niveles de sección genómica normalizados se determinan para una muestra de prueba, según la siguiente ecuación (42):

$$l_i = \frac{c_i}{G_0 + \mu_0 + X_1 \mu_1 + X_2 \mu_2} \quad (42),$$

donde l_i es un nivel de sección genómica normalizado para la porción i, c_i son los recuentos (por ejemplo, representación de recuento) para la porción i, G_0 es una estimación de la curvatura específica de la muestra, X_1 y X_2 son coordenadas canónicas. Los coeficientes μ_0, μ_1 y μ_2 están relacionados con el conjunto cuadrático de parámetros m_0, m_1 y m_2 por una inversa de una transformación de coordenadas lineales utilizada para generar coordenadas canónicas X_1 y X_2, y m_0, m_1 y m_2 son estimaciones de la curvatura específicas de la porción para una pluralidad de muestras. La ecuación (42) es un ejemplo de una implementación de una normalización PERUN cuadrática canónica.

En algunas implementaciones, se realiza un análisis de correlación para determinar el tipo de proceso de normalización (por ejemplo, LOESS, PERUN, PERUN lineal, PERUN no lineal, PERUN cuadrático, similares) que se utiliza para los recuentos y/o niveles normalizados. En determinadas implementaciones, se realiza un análisis de correlación para evaluar el grado de la curvatura de una correlación. En algunas implementaciones, evaluar el grado de la curvatura comprende realizar un análisis de correlación. En algunas implementaciones, se evalúa un grado de la curvatura para una relación entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia y (ii) una función de mapeo para las porciones de un genoma de referencia. En algunas implementaciones, un análisis de correlación comprende un análisis de regresión y/o una evaluación de la bondad de ajuste.

En algunas implementaciones, se utiliza un proceso de normalización PERUN lineal cuando una evaluación de la bondad de ajuste es igual o superior a un valor de corte del coeficiente de correlación predeterminado. En algunas implementaciones, se usa una normalización PERUN no lineal (por ejemplo, una normalización PERUN cuadrática, una normalización PERUN semicuadrática o similares) cuando una evaluación de la bondad de ajuste es inferior a un valor de corte del coeficiente de correlación predeterminado. En algunas implementaciones, un coeficiente de correlación en un valor de R^2. En algunas implementaciones, un valor de corte de coeficiente de correlación predeterminado es aproximadamente 0,5 o mayor, aproximadamente 0,55 o mayor, aproximadamente 0,6 o mayor, aproximadamente 0,65 o mayor, aproximadamente 0,7 o mayor, aproximadamente 0,75 o mayor, aproximadamente 0,8 o mayor o aproximadamente 0,85 o mayor.

Normalización de regresión híbrida

En algunas implementaciones, se utiliza un método de normalización híbrida. En algunas implementaciones, un método de normalización híbrida reduce el sesgo (por ejemplo, sesgo de GC). Una normalización híbrida, en algunas implementaciones, comprende (i) un análisis de una relación de dos variables (por ejemplo, recuentos y contenido de GC) y (ii) la selección y aplicación de un método de normalización según el análisis. Una normalización híbrida, en determinadas implementaciones, comprende (i) una regresión (por ejemplo, un análisis de regresión) y (ii) la selección y aplicación de un método de normalización según la regresión. En algunas implementaciones, los recuentos obtenidos para una primera muestra (por ejemplo, un primer conjunto de muestras) se normalizan mediante un método diferente al de los recuentos obtenidos de otra muestra (por ejemplo, un segundo conjunto de muestras). En algunas implementaciones, los recuentos obtenidos para una primera muestra (por ejemplo, un

primer conjunto de muestras) se normalizan mediante un primer método de normalización y los recuentos obtenidos a partir de una segunda muestra (por ejemplo, un segundo conjunto de muestras) se normalizan mediante un segundo método de normalización. En algunas implementaciones, la primera normalización es igual al segundo método de normalización. En algunas implementaciones, el primer método de normalización es diferente del segundo método de normalización. Por ejemplo, en determinadas implementaciones, un primer método de normalización comprende el uso de una regresión lineal y un segundo método de normalización comprende el uso de una regresión no lineal (por ejemplo, una regresión LOESS, GC-LOESS, LOWESS, suavizado LOESS).

En algunas implementaciones, se utiliza un método de normalización híbrida para normalizar las lecturas de secuencia mapeadas en porciones de un genoma o cromosoma (por ejemplo, recuentos, recuentos mapeados, lecturas mapeadas). En determinadas implementaciones, los recuentos sin procesar se normalizan y, en algunas implementaciones, los recuentos ajustados, ponderados, filtrados o previamente normalizados se normalizan mediante un método de normalización híbrida. En determinadas implementaciones, se normalizan los niveles de sección genómica o de porción o las puntuaciones Z. En algunas implementaciones, los recuentos mapeados en porciones seleccionadas de un genoma o cromosoma se normalizan mediante un enfoque de normalización híbrida. Los recuentos pueden referirse a una medida adecuada de lecturas de secuencia mapeadas en porciones de un genoma, cuyos ejemplos no limitativos incluyen recuentos sin manipular (por ejemplo, recuentos sin procesar), recuentos normalizados (por ejemplo, normalizados mediante PERUN o un método adecuado), niveles de porción (por ejemplo, niveles promedio, niveles medios, mediana de los niveles o similares), puntuaciones Z, similares o combinaciones de los mismos. Los recuentos pueden ser recuentos sin procesar o recuentos procesados de una o más muestras (por ejemplo, una muestra de prueba, una muestra de una mujer embarazada). En algunas implementaciones, los recuentos se obtienen de una o más muestras obtenidas de uno o más sujetos.

En algunas implementaciones, se selecciona un método de normalización (por ejemplo, el tipo de método de normalización) según una regresión (por ejemplo, un análisis de regresión) y/o un coeficiente de correlación. Un análisis de regresión se refiere a una técnica estadística para estimar una relación entre variables (por ejemplo, recuentos y contenido de GC). En algunas implementaciones, se genera una regresión según los recuentos y una medida del contenido de GC para cada porción de múltiples porciones de un genoma de referencia. Se puede utilizar una medida adecuada del contenido de GC, cuyos ejemplos no limitativos incluyen una medida del contenido de guanina, citosina, adenina, timina, purina (GC) o pirimidina (AT o ATU), temperatura de fusión (T_f) (por ejemplo, temperatura de desnaturalización, temperatura de combinación, temperatura de hibridación), una medida de energía libre, similares o combinaciones de los mismos. Una medida del contenido de guanina (G), citosina (C), adenina (A), timina (T), purina (GC) o pirimidina (AT o ATU) se puede expresar como una relación o un porcentaje. En algunas implementaciones se utiliza cualquier relación o porcentaje adecuado, cuyos ejemplos no limitativos incluyen GC/AT, GC/nucleótido total, GC/A, GC/T, AT/nucleótido total, AT/GC, AT/G, AT/ C, G/A, C/A, G/T, G/A, G/AT, C/T, similares o combinaciones de los mismos. En algunas implementaciones, una medida del contenido de GC es una relación o un porcentaje de GC con respecto al contenido total de nucleótidos. En algunas implementaciones, una medida del contenido de GC es una relación o un porcentaje de GC con respecto al contenido total de nucleótidos para lecturas de secuencia mapeadas en una porción del genoma de referencia. En determinadas implementaciones, el contenido de GC se determina según y/o a partir de lecturas de secuencia mapeadas en cada porción de un genoma de referencia y las lecturas de secuencia se obtienen de una muestra (por ejemplo, una muestra obtenida de una mujer embarazada). En algunas implementaciones, una medida del contenido de GC no se determina según y/o a partir de las lecturas de secuencia. En determinadas implementaciones, se determina una medida del contenido de GC para una o más muestras obtenidas de uno o más sujetos.

En algunas implementaciones, la generación de una regresión comprende generar un análisis de regresión o un análisis de correlación. Se puede usar una regresión adecuada, cuyos ejemplos no limitativos incluyen un análisis de regresión (por ejemplo, un análisis de regresión lineal), un análisis de bondad de ajuste, un análisis de correlación de Pearson, una correlación de rango, una fracción de varianza no explicada, un análisis de eficiencia del modelo de Nash-Sutcliffe, validación del modelo de regresión, reducción proporcional de la pérdida, desviación cuadrática media, similares o una combinación de los mismos. En algunas implementaciones, se genera una línea de regresión. En determinadas implementaciones, la generación de una regresión comprende generar una regresión lineal. En determinadas implementaciones, la generación de una regresión comprende generar una regresión no lineal (por ejemplo, una regresión LOESS, una regresión LOWESS).

En algunas implementaciones, una regresión determina la presencia o ausencia de una correlación (por ejemplo, una correlación lineal), por ejemplo, entre recuentos y una medida del contenido de GC. En algunas implementaciones, se genera una regresión (por ejemplo, una regresión lineal) y se determina un coeficiente de correlación. En algunas implementaciones, se determina un coeficiente de correlación adecuado, cuyos ejemplos no limitativos incluyen un coeficiente de determinación, un valor de R^2, un coeficiente de correlación de Pearson, o similar.

En algunas implementaciones, se determina la bondad de ajuste para una regresión (por ejemplo, un análisis de regresión, una regresión lineal). La bondad de ajuste a veces se determina mediante análisis visual o matemático. Una evaluación a veces incluye determinar si la bondad de ajuste es mayor para una regresión no lineal o para una regresión lineal. En algunas implementaciones, un coeficiente de correlación es una medida de la bondad de ajuste. En algunas implementaciones, una evaluación de la bondad de ajuste para una regresión se determina según un coeficiente de correlación y/o un valor de corte del coeficiente de correlación. En algunas implementaciones, una evaluación de la bondad de ajuste comprende comparar un coeficiente de correlación con un valor de corte del coeficiente de correlación. En algunas implementaciones, una evaluación de la bondad de ajuste de una regresión

es indicativa de una regresión lineal. Por ejemplo, en determinadas implementaciones, la bondad de ajuste es mayor para una regresión lineal que para una regresión no lineal, y la evaluación de la bondad de ajuste es indicativa de una regresión lineal. En algunas implementaciones, una evaluación es indicativa de una regresión lineal, y se utiliza una regresión lineal para normalizar los recuentos. En algunas implementaciones, una evaluación de la bondad de ajuste de una regresión es indicativa de una regresión no lineal. Por ejemplo, en determinadas implementaciones, la bondad de ajuste es mayor para una regresión no lineal que para una regresión lineal, y la evaluación de la bondad de ajuste es indicativa de una regresión no lineal. En algunas implementaciones, una evaluación es indicativa de una regresión no lineal, y se utiliza una regresión no lineal para normalizar los recuentos.

En algunas implementaciones, una evaluación de la bondad de ajuste es indicativa de una regresión lineal cuando un coeficiente de correlación es igual o superior a un valor de corte del coeficiente de correlación. En algunas implementaciones, una evaluación de la bondad de ajuste es indicativa de una regresión no lineal cuando un coeficiente de correlación es inferior a un valor de corte del coeficiente de correlación. En algunas implementaciones, se predetermina un valor de corte del coeficiente de correlación. En algunas implementaciones, un valor de corte del coeficiente de correlación es de aproximadamente 0,5 o mayor, aproximadamente 0,55 o mayor, aproximadamente 0,6 o mayor, aproximadamente 0,65 o mayor, aproximadamente 0,7 o mayor, aproximadamente 0,75 o mayor, aproximadamente 0,8 o mayor o aproximadamente 0,85 o mayor.

Por ejemplo, en determinadas implementaciones, se usa un método de normalización que comprende una regresión lineal cuando un coeficiente de correlación es igual o superior a aproximadamente 0,6. En determinadas implementaciones, los recuentos de una muestra (por ejemplo, recuentos por porción de un genoma de referencia, recuentos por porción, recuentos por bin) se normalizan según una regresión lineal cuando un coeficiente de correlación es igual o superior a un valor de corte del coeficiente de correlación. de 0,6; de lo contrario, los recuentos se normalizan según una regresión no lineal (por ejemplo, cuando el coeficiente es inferior a un valor de corte del coeficiente de correlación de 0,6). En algunas implementaciones, un proceso de normalización comprende generar una regresión lineal o una regresión no lineal para (i) los recuentos y (ii) el contenido de GC, para cada porción de múltiples porciones de un genoma de referencia. En determinadas implementaciones, se usa un método de normalización que comprende una regresión no lineal (por ejemplo, un método LOWESS o LOESS) cuando un coeficiente de correlación es inferior a un valor de corte del coeficiente de correlación de 0,6. En algunas implementaciones, se usa un método de normalización que comprende una regresión no lineal (por ejemplo, un método LOWESS) cuando un coeficiente de correlación (por ejemplo, un coeficiente de correlación) es inferior a un valor de corte del coeficiente de correlación de aproximadamente 0,7, inferior a aproximadamente 0,65, inferior a aproximadamente 0,6, inferior a aproximadamente 0,55 o inferior a aproximadamente 0,5. Por ejemplo, en algunas implementaciones se usa un método de normalización que comprende una regresión no lineal (por ejemplo, un método LOWESS o LOESS) cuando un coeficiente de correlación es inferior a un valor de corte del coeficiente de correlación de aproximadamente 0,6.

En algunas implementaciones, se selecciona un tipo específico de regresión (por ejemplo, una regresión lineal o no lineal) y, después de generar la regresión, los recuentos se normalizan restando la regresión a los recuentos. En algunas implementaciones, la resta de una regresión a los recuentos proporciona recuentos normalizados con un sesgo reducido (por ejemplo, sesgo de GC). En algunas implementaciones, se resta una regresión lineal a los recuentos. En algunas implementaciones, se resta una regresión no lineal (por ejemplo, una regresión LOESS, GC-LOESS, LOWESS) a los recuentos. Se puede utilizar cualquier método adecuado para restar una línea de regresión a los recuentos. Por ejemplo, si los recuentos x se derivan de la porción i (por ejemplo, una porción l, un bin i) que comprende un contenido de GC de 0,5 y una línea de regresión determina los recuentos y con un contenido de GC de 0,5, entonces x-y = recuentos normalizados para la porción i. En algunas implementaciones, los recuentos se normalizan antes y/o después de restar una regresión. En algunas implementaciones, se utilizan recuentos normalizados mediante un enfoque de normalización híbrida para generar niveles de porción, niveles de sección genómica, puntuaciones Z, elevaciones o niveles y/o perfiles de un genoma o un segmento del mismo. En determinadas implementaciones, los recuentos normalizados mediante un enfoque de normalización híbrida se analizan mediante métodos descritos en el presente documento para determinar la presencia o ausencia de una variación genética (por ejemplo, en un feto).

En algunas implementaciones, un método de normalización híbrida comprende filtrar o ponderar una o más porciones o secciones genómicas antes o después de la normalización. Se puede utilizar un método adecuado para filtrar porciones, incluidos los métodos de filtrado de porciones (por ejemplo, secciones genómicas, bins, porciones de un genoma de referencia) descritos en el presente documento. En algunas implementaciones, las porciones (por ejemplo, bins, secciones genómicas, porciones de un genoma de referencia) se filtran antes de aplicar un método de normalización híbrida. En algunas implementaciones, solo se normalizan mediante una normalización híbrida los recuentos de lecturas de secuenciación mapeadas en porciones seleccionadas (por ejemplo, porciones seleccionadas según la variabilidad del recuento). En algunas implementaciones, los recuentos de lecturas de secuenciación mapeadas en porciones filtradas de un genoma de referencia (por ejemplo, porciones filtradas según la variabilidad del recuento) se eliminan antes de utilizar un método de normalización híbrida. En algunas implementaciones, un método de normalización híbrida comprende seleccionar o filtrar porciones (por ejemplo, bins, porciones de un genoma de referencia) según un método adecuado (por ejemplo, un método descrito en el presente documento). En algunas implementaciones, un método de normalización híbrida comprende seleccionar o filtrar porciones (por ejemplo, bins, porciones de un genoma de referencia) según una medida de incertidumbre para los recuentos mapeados en cada una de las porciones para múltiples muestras de prueba. En algunas implementaciones, un método de normalización

híbrida comprende seleccionar o filtrar porciones (por ejemplo, bins o porciones de un genoma de referencia) según la variabilidad del recuento. En algunas implementaciones, un método de normalización híbrida comprende seleccionar o filtrar porciones (por ejemplo, porciones de un genoma de referencia) según el contenido de GC, elementos repetitivos, secuencias repetitivas, intrones, exones, similares o una combinación de los mismos.

Por ejemplo, en algunas implementaciones se analizan múltiples muestras de múltiples sujetos de sexo femenino gestantes y se selecciona un subconjunto de porciones (por ejemplo, bins o porciones de un genoma de referencia) según la variabilidad del recuento. En determinadas implementaciones se utiliza una regresión lineal para determinar un coeficiente de correlación para (i) recuentos y (ii) contenido de GC, para cada una de las porciones seleccionadas para una muestra obtenida de un sujeto de sexo femenino gestante. En algunas implementaciones, se determina un coeficiente de correlación que es superior a un valor de corte de correlación predeterminado (por ejemplo, de aproximadamente 0,6), una evaluación de la bondad de ajuste es indicativa de una regresión lineal y los recuentos se normalizan restando la regresión lineal a los condes. En determinadas implementaciones, se determina un coeficiente de correlación que es inferior a un valor de corte de correlación predeterminado (por ejemplo, de aproximadamente 0,6), una evaluación de la bondad de ajuste es indicativa de una regresión no lineal, se genera una regresión LOESS y los recuentos se normalizan restando la regresión LOESS a los recuentos.

Perfiles

En algunas implementaciones, una etapa de procesamiento puede comprender generar uno o más perfiles (por ejemplo, gráfico de perfiles) a partir de diversos aspectos de un conjunto de datos o derivación del mismo (por ejemplo, producto de una o más etapas de procesamiento de datos matemáticas y/o estadísticas conocidas en la técnica y/o descritas en el presente documento). El término "perfil", tal 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" incluye a menudo valores resultantes de una o más manipulaciones de datos o conjuntos de datos, basándose en uno o más criterios. Un perfil incluye a menudo múltiples puntos de datos. Cualquier número adecuado de puntos de datos puede incluirse en un perfil dependiendo de la naturaleza y/o complejidad de un conjunto de datos. En determinadas implementaciones, 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 algunas implementaciones, un perfil es representativo de la totalidad de un conjunto de datos y, en determinadas implementaciones, 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, a veces, 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 algunas implementaciones, un punto de datos en un perfil representa los resultados de la manipulación de datos para una porción. En determinadas implementaciones, un punto de datos en un perfil incluye resultados de la manipulación de datos para grupos de porciones. En algunas implementaciones, los grupos de porciones pueden ser adyacentes entre sí y, en determinadas implementaciones, 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 categorización de datos adecuada. Los ejemplos no limitativos de categorías en las que los datos pueden agruparse para generar puntos de datos de perfil incluyen: porciones basadas en el tamaño, porciones basadas en características de secuencia (por ejemplo, contenido de GC, 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 algunas implementaciones, puede generarse 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 determinadas implementaciones, un perfil generado a partir de los puntos de datos obtenidos a partir de otro perfil reduce el número de puntos de datos y/o la complejidad del conjunto de datos. Reducir el número de puntos de datos y/o la complejidad de un conjunto de datos facilita a menudo 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) suele ser una colección de recuentos normalizados o no normalizados para dos o más porciones. Un perfil suele incluir al menos una elevación o un nivel (por ejemplo, un nivel de sección genómica), y suele comprender dos o más elevaciones o niveles (por ejemplo, un perfil suele tener múltiples elevaciones o niveles). Una elevación o un nivel generalmente corresponde a un conjunto de porciones que tienen aproximadamente los mismos recuentos o recuentos normalizados. Las elevaciones o los niveles se describen con mayor detalle en el presente documento. En determinadas implementaciones, un perfil comprende una o más porciones, que pueden ser ponderadas, eliminadas, filtradas, normalizadas, ajustadas, promediadas, derivadas como media, sumadas, restadas, procesadas o transformadas mediante cualquier combinación de las mismas. Un perfil suele comprender recuentos normalizados mapeados en porciones que definen dos o más elevaciones o niveles, donde los recuentos se normalizan aún más según una de las elevaciones o de los niveles mediante un método adecuado. A menudo, los recuentos de un perfil (por ejemplo, una elevación o un nivel del perfil) se asocian a una medida de incertidumbre o un valor de incertidumbre.

Un perfil que comprende una o más elevaciones o niveles a veces se rellena (por ejemplo, relleno de huecos). El relleno (por ejemplo, relleno de huecos) se refiere a un proceso de identificación y ajuste de elevaciones o niveles en un perfil que se deben a microdeleciones maternas o duplicaciones maternas (por ejemplo, variaciones en el número de copias). En algunas implementaciones se rellenan elevaciones o niveles que se deben a microduplicaciones fetales o microdeleciones fetales. Las microduplicaciones o microdeleciones en un perfil pueden, en algunas implementaciones, aumentar o disminuir artificialmente el nivel general de un perfil (por ejemplo, un perfil de un cromosoma), lo que lleva a determinaciones positivas falsas o negativas falsas de una aneuploidía cromosómica (por ejemplo, una trisomía). En algunas implementaciones, las elevaciones o los niveles en un perfil que se deben a microduplicaciones y/o eliminaciones se identifican y ajustan (por ejemplo, se rellenan y/o se eliminan) mediante un proceso que a veces se denomina relleno o relleno de huecos. En determinadas implementaciones, un perfil comprende una o más primeras elevaciones o niveles que son significativamente diferentes de una segunda elevación o nivel dentro del perfil, cada una de las primeras elevaciones o niveles comprende una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal y una o más de las primeras elevaciones o niveles se ajustan.

Un perfil que comprende una o más elevaciones o niveles puede incluir una primera elevación o nivel y una segunda elevación o nivel. En algunas implementaciones una primera elevación o nivel es diferente (por ejemplo, significativamente diferente) que una segunda elevación o nivel. En algunas implementaciones, una primera elevación o nivel comprende un primer conjunto de porciones, una segunda elevación o nivel comprende un segundo conjunto de porciones, y el primer conjunto de porciones no es un subconjunto del segundo conjunto de porciones. En determinadas implementaciones, un primer conjunto de porciones es diferente de un segundo conjunto de porciones a partir del cual se determinan una primera y una segunda elevación o nivel. En algunas implementaciones, un perfil puede tener múltiples primeras elevaciones o niveles que son diferentes (por ejemplo, significativamente diferentes, por ejemplo, tienen un valor significativamente diferente) que una segunda elevación o nivel dentro del perfil. En algunas implementaciones, un perfil comprende una o más primeras elevaciones o niveles que son significativamente diferentes que una segunda elevación o nivel dentro del perfil, y se ajustan una o más de las primeras elevaciones o niveles. En algunas implementaciones, un perfil comprende una o más primeras elevaciones o niveles que son significativamente diferentes que una segunda elevación o nivel dentro del perfil, cada una de las una o más primeras elevaciones o niveles comprenden una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal, y se ajustan una o más de las primeras elevaciones o niveles. En algunas implementaciones, una primera elevación o nivel dentro de un perfil se elimina del perfil o se ajusta (por ejemplo, se rellena). Un perfil puede comprender múltiples elevaciones o niveles que incluyen una o más primeras elevaciones o niveles significativamente diferentes de una o más segundas elevaciones o niveles y, a menudo, la mayoría de elevaciones o niveles de un perfil son segundas elevaciones o niveles, cuyas segundas elevaciones o niveles son aproximadamente iguales entre sí. En algunas implementaciones, más del 50 %, más del 60 %, más del 70 %, más del 80 %, más del 90 % o más del 95 % de las elevaciones o niveles de un perfil son segundas elevaciones o niveles.

Algunas veces, un perfil se visualiza como una representación gráfica. Por ejemplo, pueden representarse gráficamente y visualizarse una o más elevaciones o niveles que representen recuentos (por ejemplo, recuentos normalizados) de porciones. Entre los ejemplos no limitativos de representaciones gráficas de perfiles que pueden generarse se incluyen el recuento sin procesar (por ejemplo, perfil de recuento sin procesar o perfil sin procesar), recuento normalizado, ponderado en bins, ponderado en porciones, puntuación z, valor de *p*, relación del área frente a la ploidía ajustada, mediana de elevación o del nivel frente a la relación entre la fracción fetal ajustada y medida, componentes principales, similares o combinaciones de los mismos. Los gráficos de perfiles permiten la visualización de los datos manipulados, en algunas implementaciones. En determinadas implementaciones, puede usarse una representación gráfica de perfiles para proporcionar un resultado (por ejemplo, relación de área frente a ploidía ajustada, mediana de elevación o del nivel frente a relación entre fracción fetal ajustada y medida, componentes principales). Las expresiones "representación gráfica de perfil de recuento en bruto" o "representación gráfica de perfil sin procesar", tal y como se utilizan en el presente documento, se refieren a una representación gráfica de recuentos en cada porción de una región normalizados con respecto a los recuentos totales de una región (por ejemplo, genoma, porción, cromosoma, bins cromosómicos, porciones cromosómicas de un genoma de referencia o un segmento de un cromosoma). En algunas implementaciones, puede generarse un perfil usando un procedimiento de ventana estática y, en determinadas implementaciones, puede generarse un perfil usando un procedimiento de ventana deslizante.

A veces, un perfil generado para un sujeto de prueba 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 algunas implementaciones, se genera un perfil basándose en una o más suposiciones 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 determinadas implementaciones, un perfil de prueba se centra a menudo alrededor de un valor predeterminado representativo de la ausencia de una variación genética y se desvía a menudo de un valor predeterminado en áreas correspondientes a la ubicación genómica en la que está ubicada la variación genética en el sujeto de prueba, si el sujeto de prueba presentase la variación genética. En los sujetos de prueba en riesgo de, o que padecen, una afección médica asociada con una variación genética, se espera que el valor numérico para una porción seleccionada varíe significativamente con respecto al valor predeterminado para ubicaciones genómicas no afectadas. Dependiendo de las suposiciones iniciales (por ejemplo, ploidía fija o ploidía optimizada, fracción fetal fija o fracción fetal optimizada o combinaciones de las mismas) el umbral predeterminado o valor de punto de corte o rango umbral de valores indicativos de la presencia o ausencia de una variación

genética puede variar mientras todavía proporciona un resultado útil para determinar la presencia o ausencia de una variación genética. En algunas implementaciones, un perfil es indicativo de y/o representativo de un fenotipo.

A modo de ejemplo no limitativo, los perfiles normalizados de recuento de muestras y/o de referencia pueden obtenerse a partir de datos de lectura de secuencias sin procesar mediante (a) calcular la mediana de los recuentos de referencia para cromosomas, porciones o segmentos de los mismos seleccionados a partir de un conjunto de referencias que se sabe que no portan una variación genética, (b) eliminar las porciones no informativas de los recuentos sin procesar de las muestras de referencia (por ejemplo, filtrado); (c) normalizar los recuentos de referencia para todos los bins o porciones restantes de un genoma de referencia con respecto al número residual total de recuentos (por ejemplo, la suma de los recuentos restantes después de eliminar los bins o porciones no informativas de un genoma de referencia) para el cromosoma seleccionado de la muestra de referencia o ubicación genómica seleccionada, generando de esta manera un perfil de sujeto de referencia normalizado; (d) retirar las porciones correspondientes de la muestra del sujeto de prueba; y (e) normalizar los recuentos de sujetos de prueba restantes para una o más ubicaciones genómicas seleccionadas con respecto a la suma de la mediana de los recuentos de referencia residuales para el cromosoma o cromosomas que contienen las ubicaciones genómicas seleccionadas, generando de esta manera un perfil de sujeto de prueba normalizado. En determinadas implementaciones, una etapa de normalización adicional con respecto a todo el genoma, reducida por las porciones filtradas en (b), puede incluirse entre (c) y (d).

Un perfil de conjunto de datos puede generarse mediante una o más manipulaciones de datos de lecturas de secuencia mapeadas contadas. Algunas implementaciones incluyen lo siguiente. Las lecturas de secuencia se mapean y se determina el número de recuentos o de etiquetas de secuencia que se mapean en cada bin genómico o porción (por ejemplo, se cuentan). Se genera un perfil de recuento sin procesar a partir de las lecturas de secuencia mapeadas que se cuentan. Se proporciona un resultado comparando un perfil de recuento sin procesar de un sujeto de prueba con una mediana de perfil de recuento de referencia para cromosomas, porciones o segmentos de los mismos de un conjunto de sujetos de referencia que se sabe que no presentan una variación genética, en determinadas implementaciones.

En algunas implementaciones, los datos de lectura de secuencia se filtran, opcionalmente, para eliminar los datos con ruido o las porciones no informativas. Después del filtrado, los recuentos restantes se suman normalmente para generar un conjunto de datos filtrado. Un perfil de recuento filtrado se genera a partir de un conjunto de datos filtrado, en determinadas implementaciones.

Después de contar los datos de lectura de secuencia y, opcionalmente, filtrarse, los conjuntos de datos pueden normalizarse para generar elevaciones o niveles, o perfiles. Un conjunto de datos puede normalizarse normalizando una o más porciones seleccionadas con respecto a un valor de referencia de normalización adecuado. En algunas implementaciones, un valor de referencia normalizado es representativo de los recuentos totales para el cromosoma o cromosomas de los cuales se seleccionan las porciones. En determinadas implementaciones, 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 referencia preparado a partir de un conjunto de sujetos de referencia que se sabe que no presentan una variación genética. En algunas implementaciones, un valor de referencia de normalización es representativo de una o más porciones correspondientes, porciones de cromosomas o cromosomas de un conjunto de datos del sujeto de prueba preparado a partir de un sujeto de prueba que se analiza para determinar la presencia o ausencia de una variación genética. En determinadas implementaciones, el procedimiento de normalización se realiza utilizando un enfoque de ventana estática y, en algunas implementaciones, el procedimiento de normalización se realiza utilizando un enfoque de ventana móvil o deslizante. En determinadas implementaciones, se genera un perfil que comprende recuentos normalizados para facilitar la clasificación y/o proporcionar un resultado. Puede proporcionarse un resultado basado en una representación gráfica de un perfil que comprende recuentos normalizados (por ejemplo, usando una representación gráfica de tal perfil).

Reescalado

Para eliminar la variabilidad residual específica de la muestra, a menudo provocada por diferencias biológicas (ploidía, duplicaciones/deleciones), a veces se vuelve a escalar un perfil (por ejemplo, un perfil normalizado). Un proceso de escalado, en algunas implementaciones, evalúa un nivel promedio, nivel medio o la mediana del nivel, y una medida de incertidumbre asociada (por ejemplo, una DAM) para porciones de un perfil normalizado. En algunas implementaciones, se evalúan todas las porciones y, a veces, solo se evalúan las porciones autosómicas. En algunas implementaciones, las porciones que quedan fuera de un rango predeterminado (por ejemplo, según un nivel promedio, nivel medio o mediana del nivel del perfil y una medida de incertidumbre para el nivel del perfil) se identifican, marcan y/o filtran. Por ejemplo, en algunas implementaciones, se eliminan y/o filtran porciones con un nivel de sección genómica normalizado que es superior a un nivel promedio, a la mediana del nivel o un nivel medio del perfil en aproximadamente 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5 o aproximadamente 6 veces la medida de incertidumbre (por ejemplo, una DAM determinada según un nivel de perfil normalizado) o mayores. En algunas implementaciones, una porción se identifica, se marca y/o se filtra cuando la desviación supera 3 veces la medida de incertidumbre (por ejemplo, tres DAM). En algunas implementaciones de un proceso de reescalado, un nivel promedio, un nivel medio o la mediana de un nivel de un perfil (por ejemplo, antes del filtrado) se divide entre un nivel promedio, un nivel medio o la mediana de un nivel de un perfil después del filtrado. En algunas implementaciones, este proceso de reescalado lleva todas las porciones euploides a un nivel de aproximadamente uno. En algunas implementaciones, el proceso de reescalado

minimiza el efecto de cualquier aneuploidía en el nivel de porciones euploides del genoma. Un proceso de reescalado se puede realizar antes o después de una normalización adecuada y se puede repetir, en algunas implementaciones.

Diferentes elevaciones o niveles

En algunas implementaciones, un perfil de recuentos normalizados comprende una elevación o un nivel (por ejemplo, una primera elevación o nivel) significativamente diferente de otra elevación o nivel (por ejemplo, una segunda elevación o nivel) dentro del perfil. Una primera elevación o nivel puede ser superior o inferior a una segunda elevación o nivel. En algunas implementaciones, una primera elevación o nivel es para un conjunto de porciones que comprenden una o más lecturas que comprenden una variación del número de copias (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal) y la segunda elevación o nivel es para un conjunto de porciones que comprenden lecturas que prácticamente no tienen variación del número de copias. En algunas implementaciones, significativamente diferente se refiere a una diferencia observable. En algunas implementaciones, significativamente diferente se refiere a estadísticamente diferente o a una diferencia estadísticamente significativa. Una diferencia estadísticamente significativa es, algunas veces, una evaluación estadística de una diferencia observada. Una diferencia estadísticamente significativa puede evaluarse mediante un método adecuado en la técnica. Puede usarse cualquier umbral o rango adecuado para determinar que dos elevaciones o niveles son significativamente diferentes. En determinadas implementaciones, dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) que difieren en aproximadamente 0,01 por ciento o más (por ejemplo, 0,01 por ciento de uno o cualquiera de los valores de elevación o nivel) son significativamente diferentes. En algunas implementaciones, dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) que difieren en aproximadamente un 0,1 por ciento o más son significativamente diferentes. En determinadas implementaciones, dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) que difieren en aproximadamente un 0,5 por ciento o más son significativamente diferentes. En algunas implementaciones, dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) que difieran en aproximadamente 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 el 10 % son significativamente diferentes. En algunas implementaciones dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) son significativamente diferentes y no hay solapamiento en ninguna de las elevaciones o niveles y/o no hay solapamiento en un rango definido por una medida de incertidumbre calculada para una o ambas elevaciones o niveles. En determinadas implementaciones, la medida de incertidumbre es una desviación estándar expresada como sigma. En algunas implementaciones dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) son significativamente diferentes y difieren en aproximadamente 1 o más veces la medida de incertidumbre (por ejemplo, 1 sigma). En algunas implementaciones, dos elevaciones o niveles (por ejemplo, elevaciones o niveles medios) son significativamente diferentes y difieren en aproximadamente 2 o más veces la medida 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 la medida de incertidumbre. En algunas implementaciones dos elevaciones o niveles (por ejemplo, elevaciones o 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 la medida de incertidumbre o más. En algunas implementaciones, el nivel de confianza aumenta a medida que aumenta la diferencia entre dos elevaciones o niveles. En determinadas implementaciones, el nivel de confianza disminuye a medida que disminuye la diferencia entre dos elevaciones o niveles y/o a medida que aumenta la medida de incertidumbre. Por ejemplo, algunas veces el nivel de confianza aumenta con la relación de la diferencia entre las elevaciones o niveles y la desviación estándar (por ejemplo, DAM).

Se pueden usar uno o más algoritmos de predicción para determinar la importancia o dar significado a los datos de detección recopilados en condiciones variables que se pueden ponderar de forma independiente o dependiente 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. Por ejemplo, una variable puede ser el diseño de un conjunto de especies de ácidos nucleicos amplificados, el número de conjuntos de especies de ácidos nucleicos amplificados, el porcentaje de contribución genética fetal probado, el porcentaje de contribución genética materna probado, el tipo de anomalía cromosómica probado, el tipo de trastorno genético probado, el tipo de anomalías ligadas al sexo probado, la edad de la madre y similares. El término “independiente” como se usa en el presente documento se refiere a no ser influenciado o controlado por otro. El término “dependiente”, como se usa en el presente documento, se refiere a ser influenciado o controlado por otro. Por ejemplo, un cromosoma en particular y un evento de trisomía que ocurre para ese cromosoma en particular que da como resultado un ser viable son variables que dependen una de la otra.

En algunas implementaciones, un primer conjunto de porciones suele incluir porciones que son diferentes (por ejemplo, que no se solapan con) un segundo conjunto de porciones. Por ejemplo, a veces una primera elevación o nivel de recuentos normalizados es significativamente diferente de una segunda elevación o nivel de recuentos normalizados en un perfil, y la primera elevación o nivel es para un primer conjunto de porciones, la segunda elevación o nivel es para un segundo conjunto de porciones, y las porciones no se solapan en el primer conjunto y el segundo conjunto de porciones. En determinadas implementaciones, un primer conjunto de porciones no es un subconjunto de un segundo conjunto de porciones a partir del cual se determinan una primera elevación o nivel y una segunda elevación o nivel, respectivamente. En algunas implementaciones, un primer conjunto de porciones es diferente y/o distinto de un segundo conjunto de porciones a partir del cual se determinan una primera elevación o nivel y una segunda elevación o nivel, respectivamente.

En algunas implementaciones, un primer conjunto de porciones es un subconjunto de un segundo conjunto de porciones de un perfil. Por ejemplo, a veces una segunda elevación o nivel de recuentos normalizados para un segundo conjunto de porciones en un perfil comprende recuentos normalizados de un primer conjunto de porciones para una primera elevación o nivel en el perfil, y el primer conjunto de porciones es un subconjunto del segundo conjunto de porciones en el perfil. En algunas implementaciones, una elevación o nivel medio, promedio o mediana se deriva de una segunda elevación o nivel donde la segunda elevación o nivel comprende una primera elevación o nivel. En algunas implementaciones, una segunda elevación o nivel comprende un segundo conjunto de porciones que representan un cromosoma entero, y una primera elevación o nivel comprende un primer conjunto de porciones donde el primer conjunto es un subconjunto del segundo conjunto de porciones y la primera elevación o nivel representa una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal que está presente en el cromosoma.

En algunas implementaciones, un valor de una segunda elevación o nivel está más cerca del valor medio, promedio o mediana de un perfil de recuento para un cromosoma, o segmento del mismo, que la primera elevación o nivel. En algunas implementaciones, una segunda elevación o nivel es una elevación o nivel medio de un cromosoma, una porción de un cromosoma o un segmento del mismo. En algunas implementaciones, una primera elevación o nivel es significativamente diferente de una elevación o nivel predominante (por ejemplo, una segunda elevación o nivel) que representa un cromosoma o un segmento del mismo. Un perfil puede incluir múltiples primeras elevaciones o niveles que difieren significativamente de una segunda elevación o nivel, y cada primera elevación o nivel independientemente puede ser mayor o menor que la segunda elevación o nivel. En algunas implementaciones, una primera elevación o nivel y una segunda elevación o nivel se derivan del mismo cromosoma y la primera elevación o nivel es mayor o menor que la segunda elevación o nivel, y la segunda elevación o nivel es la elevación o nivel predominante del cromosoma. En algunas implementaciones, una primera elevación o nivel y una segunda elevación o nivel se derivan del mismo cromosoma, una primera elevación o nivel es indicativa de una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal, deleción, inserción, duplicación), y una segunda elevación o nivel es una elevación o nivel medio o una elevación o nivel predominante de porciones para un cromosoma, o segmento del mismo.

En determinadas implementaciones, una lectura en un segundo conjunto de porciones para una segunda elevación o nivel no incluye sustancialmente una variación genética (por ejemplo, una variación del número de copias, una variación del número de copias materno y/o fetal). A menudo, un segundo conjunto de porciones para una segunda elevación o nivel incluye cierta variabilidad (por ejemplo, variabilidad en la elevación o nivel, variabilidad en los recuentos de las porciones). En algunas implementaciones, una o más porciones en un conjunto de porciones para una elevación o un nivel asociado con una variación sustancialmente nula del número de copias incluyen una o más lecturas que tienen una variación del número de copias presente en un genoma materno y/o fetal. Por ejemplo, a veces un conjunto de porciones incluye una variación del número de copias que está presente en un segmento pequeño de un cromosoma (por ejemplo, menos de 10 porciones) y el conjunto de porciones es para una elevación o un nivel asociado con una variación del número de copias sustancialmente nula. Por tanto, un conjunto de porciones que no incluya sustancialmente ninguna variación en el número de copias aún puede incluir una variación en el número de copias que esté presente en menos de aproximadamente 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 porciones de una elevación o un nivel.

En algunas implementaciones, una primera elevación o nivel corresponde a un primer conjunto de porciones y una segunda elevación o nivel corresponde a 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 algunas implementaciones, el primer conjunto de porciones y el segundo conjunto de porciones no son contiguos.

Pueden utilizarse lecturas de secuencias relativamente cortas de una mezcla de ácido nucleico fetal y materno para proporcionar recuentos que pueden transformarse en una elevación o un nivel y/o un perfil. Los recuentos, elevaciones o niveles y perfiles pueden representarse de forma electrónica o tangible y visualizarse. Los recuentos mapeados en porciones (por ejemplo, representados como elevaciones o 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 presente en un feto y/o una mujer embarazada.

Elevación o nivel de referencia y valor de referencia normalizado

En algunas implementaciones, un perfil comprende una elevación o un nivel de referencia (por ejemplo, una elevación o un nivel utilizado como referencia). A menudo, un perfil de recuentos normalizados proporciona una elevación o nivel de referencia a partir del cual se determinan las elevaciones o niveles esperados y los rangos esperados (véase más adelante la explicación sobre elevaciones o niveles esperados y rangos). Una elevación o nivel de referencia suele ser para recuentos normalizados de porciones que comprenden lecturas mapeadas tanto de una madre como de un feto. Una elevación o nivel de referencia suele ser la suma de los recuentos normalizados de lecturas cartografiadas de un feto y una madre (por ejemplo, una mujer embarazada). En algunas implementaciones, una elevación o nivel de referencia es para porciones que comprenden lecturas mapeadas de una madre euploide y/o un feto euploide. En algunas implementaciones, una elevación o 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 del número de copias, una

microduplicación, una microdeleción, una inserción). En algunas implementaciones, una elevación o nivel de referencia es para porciones que sustancialmente no incluyen variaciones genéticas maternas y/o fetales (por ejemplo, una aneuploidía (por ejemplo, una trisomía), una variación del número de copias, una microduplicación, una microdeleción, una inserción). En algunas implementaciones se utiliza una segunda elevación o nivel como elevación o nivel de referencia. En determinadas implementaciones, un perfil comprende una primera elevación o nivel de recuentos normalizados y una segunda elevación o nivel de recuentos normalizados, siendo la primera elevación o nivel significativamente diferente de la segunda elevación o nivel, y la segunda elevación o nivel es la elevación o nivel de referencia. En determinadas implementaciones un perfil comprende una primera elevación o nivel de recuentos normalizados para un primer conjunto de porciones, una segunda elevación o nivel de recuentos normalizados para un segundo conjunto de porciones, el primer conjunto de porciones incluye lecturas mapeadas que tienen una variación del número de copias materno y/o fetal, el segundo conjunto de porciones comprende lecturas mapeadas que sustancialmente no tienen variación del número de copias materno y/o fetal, y la segunda elevación o nivel es una elevación o nivel de referencia.

En algunas implementaciones, los recuentos mapeados en porciones para una o más elevaciones o niveles de un perfil se normalizan según los recuentos de una elevación o nivel de referencia. En algunas implementaciones, la normalización de los recuentos de una elevación o un nivel según los recuentos de una elevación o un nivel de referencia comprende dividir los recuentos de una elevación o un nivel entre los recuentos de una elevación o un nivel de referencia, o un múltiplo o fracción de los mismos. Los recuentos normalizados según los recuentos de una elevación o nivel de referencia se han normalizado a menudo según otro procedimiento (por ejemplo, PERUN) y además a menudo, los recuentos de una elevación o nivel de referencia se han normalizado (por ejemplo, mediante PERUN). En algunas implementaciones, los recuentos de una elevación o nivel se normalizan según los recuentos de una elevación o nivel de referencia, y los recuentos de la elevación o nivel de referencia son escalables a un valor adecuado antes o después de la normalización. El procedimiento de escalar los recuentos de una elevación o nivel de referencia puede comprender cualquier constante adecuada (es decir, número) y cualquier manipulación matemática adecuada puede aplicarse a los recuentos de una elevación o nivel de referencia.

Un valor de referencia normalizado (NRV) se determina a menudo según los recuentos normalizados de una elevación o nivel de referencia. Determinar un NRV puede comprender cualquier procedimiento de normalización adecuado (por ejemplo, manipulación matemática) aplicado a los recuentos de una elevación o nivel de referencia donde se usa el mismo procedimiento de normalización para normalizar los recuentos de otras elevaciones o niveles dentro del mismo perfil. Determinar un NRV comprende a menudo dividir una elevación o nivel de referencia entre sí misma. Determinar un NRV comprende a menudo dividir una elevación o nivel de referencia entre un múltiplo de sí misma. Determinar un NRV comprende a menudo dividir una elevación o nivel de referencia entre la suma o diferencia de la elevación o nivel de referencia y una constante (por ejemplo, cualquier número).

Un NRV se denomina, algunas veces, valor nulo. Un NRV puede ser cualquier valor adecuado. En algunas implementaciones, un NRV es cualquier valor distinto de cero. En algunas implementaciones, un NRV es un número entero. En algunas implementaciones, un NRV es un número entero positivo. En algunas implementaciones, un NRV es 1, 10, 100 o 1000. A menudo, un NRV es igual a 1. En algunas implementaciones, un NRV es igual a cero. Los recuentos de una elevación o nivel de referencia pueden normalizarse con respecto a cualquier NRV adecuado. En algunas implementaciones, los recuentos de una elevación o nivel de referencia se normalizan con respecto a un NRV de cero. A menudo, los recuentos de una elevación o nivel de referencia se normalizan con respecto a un NRV de 1.

Elevaciones o niveles esperados

Una elevación o nivel esperado es a veces una elevación o nivel predefinido (por ejemplo, una elevación o nivel teórico, una elevación o nivel predicho). Una “elevación o nivel esperado” se denomina a veces en el presente documento “valor de elevación o nivel predeterminado”. En algunas implementaciones, una elevación o nivel esperado es un valor predicho para una elevación o un nivel de recuentos normalizados para un conjunto de porciones que incluyen una variación del número de copias. En determinadas implementaciones, una elevación o nivel esperado se determina para un conjunto de porciones que no incluyen sustancialmente ninguna variación en el número de copias. Una elevación o nivel esperado puede determinarse para una ploidía cromosómica (por ejemplo, 0, 1, 2 (es decir, diploide), 3 o 4 cromosomas) o una microploidía (deleción homocigota o heterocigota, duplicación, inserción o ausencia de los mismos). A menudo, se determina una elevación o nivel esperado para una microploidía materna (por ejemplo, una variación del número de copias materno y/o fetal).

Una elevación o nivel esperado para una variación genética o una variación del número de copias puede determinarse de cualquier manera adecuada. A menudo, una elevación o un nivel esperados se determinan mediante una manipulación matemática adecuada de una elevación o un nivel (por ejemplo, recuentos mapeados en un conjunto de porciones para una elevación o un nivel). En algunas implementaciones, una elevación o nivel esperado se determina utilizando una constante a la que a veces se hace referencia como constante de elevación o nivel esperado. Una elevación o nivel esperado para una variación del número de copias a veces se calcula multiplicando una elevación o nivel de referencia, recuentos normalizados de una elevación o nivel de referencia o un NRV por una constante de elevación o nivel esperado, sumando una constante de elevación o nivel esperado, restando una constante de elevación o nivel esperado, dividiendo entre una constante de elevación o nivel esperado, o una combinación de los mismos. A menudo, una elevación o nivel esperado (por

ejemplo, una elevación o nivel esperado de una variación del número de copias materno y/o fetal) determinado para el mismo sujeto, muestra o grupo de prueba se determina según la misma elevación o nivel de referencia o NRV.

A menudo, una elevación o nivel esperado se determina multiplicando una elevación o nivel de referencia, recuentos normalizados de una elevación o nivel de referencia o un NRV por una constante de elevación o nivel esperado, donde la elevación o nivel de referencia, recuentos normalizados de una elevación o nivel de referencia o NRV no es igual a cero. En algunas implementaciones, una elevación o nivel esperado se determina sumando una constante de elevación o nivel esperado a la elevación o nivel de referencia, los recuentos normalizados de una elevación o nivel de referencia o un NRV que sea igual a cero. En algunas implementaciones, una elevación o nivel esperado, los recuentos normalizados de una elevación o nivel de referencia, el NRV y la constante de elevación o nivel esperado son escalables. El procedimiento de escalado puede comprender cualquier constante adecuada (es decir, número) y cualquier manipulación matemática adecuada en el que el mismo procedimiento de escalado se aplica a todos los valores considerados.

Constante de elevación o nivel esperado

Una constante de elevación o nivel esperado puede determinarse mediante un método adecuado. En algunas implementaciones, una constante de elevación o nivel esperado se determina arbitrariamente. A menudo, se determina empíricamente una constante de elevación o nivel esperado. En algunas implementaciones, una constante de elevación o nivel esperado se determina según una manipulación matemática. En algunas implementaciones, una constante de elevación o nivel esperado se determina según una referencia (por ejemplo, un genoma de referencia, una muestra de referencia, datos de prueba de referencia). En algunas implementaciones, una constante de elevación o nivel esperado se predetermina para una elevación un nivel representativo de la presencia o ausencia de una variación genética o variación del número de copias (por ejemplo, una duplicación, inserción o deleción). En algunas implementaciones, una constante de elevación o nivel esperado se predetermina para una elevación o un nivel representativo de la presencia o ausencia de una variación del número de copias materno, variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal. Una constante de elevación o nivel esperado para una variación del número de copias puede ser cualquier constante o conjunto de constantes adecuadas.

En algunas implementaciones, la constante de elevación o nivel esperado para una duplicación homocigota (por ejemplo, una duplicación homocigota) 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 algunas implementaciones, la constante de elevación o nivel esperado para una duplicación homocigota 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 elevación o nivel esperado para una duplicación homocigota es de 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 elevación o nivel esperado para una duplicación homocigota es de aproximadamente 2.

En algunas implementaciones, la constante de elevación o nivel esperado para una duplicación heterocigota (por ejemplo, una duplicación homocigota) 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 algunas implementaciones, la constante de elevación o nivel esperado para una duplicación heterocigota es de aproximadamente 1,2, 1,3, 1,4, 1,5, 1,6, 1,7 o aproximadamente 1,8. A menudo, la constante de elevación o nivel esperado para una duplicación heterocigota es de 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 algunas implementaciones, la constante de elevación o nivel esperado para una duplicación heterocigota es de aproximadamente 1,5.

En algunas implementaciones, la constante de elevación o nivel esperado para la ausencia de una variación del número de copias (por ejemplo, la ausencia de una variación del número de copias materno y/o variación del número de copias fetal) 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 algunas implementaciones, la constante de elevación o nivel esperado para la ausencia de una variación del 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 elevación o nivel esperado para la ausencia de una variación del número de copias es de aproximadamente 1,09, 1,08, 1,06, 1,04, 1,02, 1,0, 0,98, 0,96, 0,94 o aproximadamente 0,92. En algunas implementaciones, la constante de elevación o nivel esperado para la ausencia de una variación del número de copias es de aproximadamente 1.

En algunas implementaciones, la constante de elevación o nivel esperado para una deleción heterocigota (por ejemplo, una deleción materna, fetal, o una deleción materna y una fetal heterocigota) es 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. En algunas implementaciones, la constante de elevación o nivel esperado para una deleción heterocigota es de aproximadamente 0,2, 0,3, 0,4, 0,5, 0,6, 0,7 o aproximadamente 0,8. A menudo, la constante de elevación o nivel esperado para una deleción heterocigota es de aproximadamente 0,40, 0,42, 0,44, 0,46, 0,48, 0,5, 0,52, 0,54, 0,56, 0,58 o aproximadamente 0,60. En algunas implementaciones, la constante de elevación o nivel esperado para una deleción heterocigota es de aproximadamente 0,5.

En algunas implementaciones, la constante de elevación o nivel esperado para una deleción homocigota (por ejemplo, una deleción homocigota) puede ser 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 algunas implementaciones, la constante de elevación o nivel esperado para una deleción homocigota 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 elevación o nivel esperado para una deleción homocigota es de aproximadamente -0,1, -0,08, -0,06, -0,04, -0,02, 0,0, 0,02, 0,04, 0,06, 0,08 o aproximadamente 0,10. A menudo, la constante de elevación o nivel esperado para una deleción homocigota es de aproximadamente 0.

Rango de elevación o nivel esperado

En algunas implementaciones, la presencia o ausencia de una variación genética o de una variación del número de copias (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal) se determina mediante una elevación o un nivel que cae dentro o fuera de un rango de elevación o nivel esperado. Un rango de elevación o nivel esperado se determina a menudo según una elevación o nivel esperado. En algunas implementaciones, un rango de elevación o nivel esperado se determina para una elevación o un nivel que no comprende sustancialmente ninguna variación genética o sustancialmente ninguna variación en el número de copias. Se puede utilizar un método adecuado para determinar un rango de elevación o nivel esperado.

En algunas implementaciones, un rango de elevación o nivel esperado se define según una medida adecuada de incertidumbre calculada para una elevación o un nivel. Los ejemplos no limitantes de una medida de incertidumbre son una desviación estándar, un error estándar, una varianza calculada, un valor de *p* y una desviación absoluta media (DAM). En algunas implementaciones, un rango de elevación o nivel esperado para una variación genética o una variación del número de copias se determina, en parte, calculando la medida de incertidumbre para una elevación o un nivel (por ejemplo, una primera elevación o nivel, una segunda elevación o nivel, una primera elevación o nivel y una segunda elevación o nivel). En algunas implementaciones, un rango de elevación o nivel esperado se define según una medida de incertidumbre calculada para un perfil (por ejemplo, un perfil de recuentos normalizados para un cromosoma o segmento del mismo). En algunas implementaciones, se calcula una medida de incertidumbre para una elevación o un nivel que comprenda sustancialmente ninguna variación genética o sustancialmente ninguna variación del número de copias. En algunas implementaciones, se calcula una medida de incertidumbre para una primera elevación o nivel, una segunda elevación, o nivel o una primera elevación o nivel y una segunda elevación o nivel. En algunas implementaciones, se determina una medida de incertidumbre para una primera elevación o nivel, una segunda elevación o nivel, o una segunda elevación o nivel que comprende una primera elevación o nivel.

A veces se calcula un rango de elevación o nivel esperado, en parte, multiplicando, sumando, restando o dividiendo una medida de incertidumbre por una constante (por ejemplo, una constante predeterminada) n. Puede utilizarse un procedimiento matemático adecuado o una combinación de procedimientos. La constante n (por ejemplo, constante predeterminada n) a veces se denomina intervalo de confianza. Un intervalo de confianza seleccionado se determina según la constante n que se selecciona. La constante n (por ejemplo, la constante predeterminada n, el intervalo de confianza) puede determinarse de manera adecuada. La constante n puede ser un número o fracción de un número mayor de cero. La constante n puede ser un número entero. A menudo, la constante n es un número menor de 10. En algunas implementaciones, la constante n es un número menor de aproximadamente 10, menor de aproximadamente 9, menor de aproximadamente 8, menor de aproximadamente 7, menor de aproximadamente 6, menor de aproximadamente 5, menor de aproximadamente 4, menor de aproximadamente 3 o menor de aproximadamente 2. En algunas implementaciones, la constante n es de aproximadamente 10, 9,5, 9, 8,5, 8, 7,5, 7, 6,5, 6, 5,5, 5, 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 mujer embarazada y/o un feto) con una disposición genética conocida.

A menudo, una medida de incertidumbre y una constante n definen un rango (por ejemplo, un límite de incertidumbre). Por ejemplo, a veces una medida de incertidumbre es una desviación estándar (por ejemplo, +/- 5) y se multiplica por una constante n (por ejemplo, un intervalo de confianza) definiendo de este modo un rango o valora de corte de incertidumbre (por ejemplo, de 5n a -5n).

En algunas implementaciones, un rango de elevación o nivel esperado para una variación genética (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal) es la suma de una elevación o nivel esperado más una constante n veces la incertidumbre (por ejemplo, n x sigma (por ejemplo, 6 sigma)).

En algunas implementaciones, el rango de elevación esperado para una variación genética o de número de copias designada por k puede definirse mediante la fórmula:

$$\text{Fórmula R: (rango de elevación esperado)}_k = \text{(elevación esperada)}_k + n\sigma$$

donde σ es un valor de incertidumbre, n es una constante (por ejemplo, una constante predeterminada) y el rango de elevación esperado y la elevación esperada son para la variación genética k (por ejemplo, k = una deleción heterocigota,

por ejemplo, k = la ausencia de una variación genética). Por ejemplo, para una elevación esperada igual a 1 (por ejemplo, la ausencia de una variación del número de copias), un valor de incertidumbre (es decir, σ) igual a +/- -0,05, y n = 3, el rango de elevación esperado se define como de 1,15 a 0,85. En algunas implementaciones, el rango de elevación esperado para una duplicación heterocigota se determina como de 1,65 a 1,35 cuando la elevación esperada para una duplicación heterocigota es 1,5, n = 3, y el valor de incertidumbre σ es +/- 0,05. En algunas implementaciones el rango de elevación esperado para una deleción heterocigota se determina como de 0,65 a 0,35 cuando la elevación esperada para una duplicación heterocigota es 0,5, n = 3, y el valor de incertidumbre σ es +/- 0,05. En algunas implementaciones el rango de elevación esperado para una duplicación homocigota se determina como de 2,15 a 1,85 cuando la elevación esperada para una duplicación heterocigota es de 2,0, n = 3 y el valor de incertidumbre σ es +/- 0,05. En algunas implementaciones el rango de elevación esperado para una deleción homocigota se determina como de 0,15 a -0,15 cuando la elevación esperada para una duplicación heterocigota es de 0,0, n = 3 y el valor de incertidumbre σ es +/- 0,05.

En algunas implementaciones, el rango de nivel esperado para una variación genética o de número de copias designada por k puede definirse mediante la fórmula:

Fórmula R: $(\text{rango de elevación esperado})_k = (\text{elevación esperada})_k + n\sigma$

donde σ es una medida de incertidumbre, *n* es una constante (por ejemplo, una constante predeterminada) y el rango de nivel esperado y el nivel esperado son para la variación genética k (por ejemplo, k = una deleción heterocigota, 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), una medida de incertidumbre (es decir, σ) igual a +/- 0,05, y *n* = 3, el rango de nivel esperado se define como de 1,15 a 0,85. En algunas implementaciones, el rango de nivel esperado para una duplicación heterocigota se determina como de 1,65 a 1,35 cuando el nivel esperado para una duplicación heterocigota es 1,5, *n* = 3, y la medida de incertidumbre σ es +/-0,05. En algunas implementaciones el rango de nivel esperado para una deleción heterocigota se determina como de 0,65 a 0,35 cuando el nivel esperado para una duplicación heterocigota es 0,5, *n* = 3, y la medida de incertidumbre σ es +/- 0,05. En algunas implementaciones, el rango de nivel esperado para una duplicación homocigota se determina como de 2,15 a 1,85 cuando el nivel esperado para una duplicación heterocigota es 2,0, *n* = 3 y la medida de incertidumbre σ es +/- 0,05. En algunas implementaciones, el rango de nivel esperado para una deleción homocigota se determina como de 0,15 a -0,15 cuando el nivel esperado para una duplicación heterocigota es 0,0, *n* = 3 y la medida de incertidumbre σ es +/- 0,05.

En algunas implementaciones, un rango de elevación o nivel esperado para una variación del número de copias homocigota (por ejemplo, una variación del número de copias homocigota materna, fetal o materna y fetal) se determina, en parte, según un rango de elevación o nivel esperado para una variación del número de copias heterocigota correspondiente. Por ejemplo, a veces un rango de elevación o nivel esperado para una duplicación homocigota comprende todos los valores mayores que un límite superior de un rango de elevación o nivel esperado para una duplicación heterocigota. En algunas implementaciones un rango de elevación o nivel esperado para una duplicación homocigota comprende todos los valores mayores o iguales a un límite superior de un rango de elevación o nivel esperado para una duplicación heterocigota. En algunas implementaciones un rango de elevación o nivel esperado para una duplicación homocigota comprende todos los valores mayores que un límite superior de un rango de elevación o nivel esperado para una duplicación heterocigota y menores que el límite superior definido por la fórmula R, donde σ es una medida de incertidumbre y es un valor positivo, n es una constante y k es una duplicación homocigota. En algunas implementaciones, un rango de elevación o nivel esperado para una duplicación homocigota comprende todos los valores mayores o iguales que un límite superior de un rango de elevación o nivel esperado para una duplicación heterocigota y menores o iguales que el límite superior definido por la fórmula R, donde σ es una medida de incertidumbre, σ es un valor positivo, n es una constante y k es una duplicación homocigota.

En algunas implementaciones, un rango de elevación o nivel esperado para una deleción homocigota comprende todos los valores menores que un límite inferior de un rango de elevación o nivel esperado para una deleción heterocigota. En algunas implementaciones un rango de elevación o nivel esperado para una deleción homocigota comprende todos los valores menores o iguales a un límite inferior de un rango de elevación o nivel esperado para una deleción heterocigota. En algunas implementaciones un rango de elevación o nivel esperado para una deleción homocigota comprende todos los valores menores que un límite inferior de un rango de elevación o nivel esperado para una deleción heterocigota y mayores que el límite inferior definido por la fórmula R, donde σ es una medida de incertidumbre, σ es un valor negativo, n es una constante y k es una deleción homocigota. En algunas implementaciones, un rango de elevación o nivel esperado para una deleción homocigota comprende todos los valores menores o iguales que un límite inferior de un rango de elevación o nivel esperado para una deleción heterocigota y mayores o iguales que el límite inferior definido por la fórmula R, donde σ es una medida de incertidumbre, σ es un valor negativo, n es una constante y k es una deleción homocigota.

Se puede utilizar una medida de incertidumbre (también denominada en el presente documento valor de incertidumbre) para determinar un valor umbral. En algunas implementaciones, un rango (por ejemplo, un rango umbral) se obtiene calculando la medida de incertidumbre determinado a partir de recuentos sin procesar, filtrados y/o normalizados. Un rango puede determinarse multiplicando la medida de incertidumbre para una elevación o un nivel (por ejemplo, recuentos normalizados de una elevación o un nivel) por una constante predeterminada (por ejemplo, 1, 2, 3, 4, 5, 6, etc.) que represente el múltiplo de incertidumbre (por ejemplo, número de desviaciones estándar) seleccionado como umbral de corte (por ejemplo, multiplicar por 3 para 3 desviaciones estándar), donde se genera un rango, en algunas

implementaciones. Un rango puede determinarse sumando y/o restando un valor (por ejemplo, un valor predeterminado, una medida de incertidumbre, una medida de incertidumbre multiplicada por una constante predeterminada) a y/o de una elevación o un nivel, done se genera un rango, en algunas implementaciones. Por ejemplo, para una elevación o un nivel igual a 1, una desviación estándar de +/-0,2, donde una constante predeterminada es 3, el intervalo puede calcularse como (1 + 3(0,2)) a (1 + 3(-0,2)), o de 1,6 a 0,4. Un rango a veces puede definir un rango esperado o rango de elevación o nivel esperado para una variación del número de copias. En determinadas implementaciones, algunas o todas las porciones que superan un valor umbral, que se encuentran fuera de un rango o que se encuentran dentro de un rango de valores, se eliminan como parte de, antes de, o después de un procedimiento de normalización. En algunas implementaciones, algunas o todas las porciones que superan un valor umbral calculado, que se encuentran fuera de un rango o que se encuentran dentro de un rango se ponderan o ajustan como parte de, o antes del procedimiento de normalización o clasificación. En el presente documento se describen ejemplos de ponderación. Las expresiones "datos redundantes", y "lecturas mapeadas redundantes", tal como se usan en el presente documento, se refieren a lecturas de secuencia derivadas de muestras que se identifican que ya se han asignado a una ubicación genómica (por ejemplo, posición de base) y/o contado para una porción.

En algunas implementaciones, una medida de incertidumbre (o valor de incertidumbre) se determina según la siguiente fórmula:

$$Z = \frac{L_A - L_o}{\sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_o^2}{N_o}}}$$

donde Z representa la desviación normalizada entre dos elevaciones o niveles, L es la elevación o nivel medio (o mediana) y sigma es la desviación estándar (o DAM). El subíndice O denota un segmento de un perfil (por ejemplo, una segunda elevación o nivel, un cromosoma, un NRV, un "nivel euploide", un nivel ausente de una variación del número de copias), y A denota otro segmento de un perfil (por ejemplo, una primera elevación o nivel, una elevación o un nivel que representa una variación del número de copias, una elevación o un nivel que representa una aneuploidía (por ejemplo, una trisomía). La variable N_o representa el número total de porciones Del segmento del perfil indicado por el subíndice O. N_A representa el número total de porciones del segmento del perfil indicado por el subíndice A.

Categorización de una variación del número de copias

Una elevación o un nivel (por ejemplo, una primera elevación o nivel) que difiere significativamente de otra elevación o nivel (por ejemplo, una segunda elevación o nivel) puede categorizarse a menudo como una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal, una variación del número de copias fetal, una deleción, duplicación, inserción) según un rango de elevación o nivel esperado. En algunas implementaciones, la presencia de una variación del número de copias se categoriza cuando una primera elevación o nivel es significativamente diferente de una segunda elevación o nivel y la primera elevación o nivel se encuentra dentro del rango de elevación o nivel esperado para una variación del número de copias. Por ejemplo, una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal, una variación del número de copias fetal) puede categorizarse cuando una primera elevación o nivel es significativamente diferente de una segunda elevación o nivel y la primera elevación o nivel se encuentra dentro del rango de elevación o nivel esperado para una variación del número de copias. En algunas implementaciones, una duplicación heterocigota (por ejemplo, una duplicación heterocigota materna o fetal, o materna y fetal) o deleción heterocigota (por ejemplo, una deleción materna o fetal, o materna y fetal, heterocigota) se categoriza cuando una primera elevación o nivel es significativamente diferente de una segunda elevación o nivel y la primera elevación o nivel se encuentra dentro del rango de elevación o nivel esperado para una duplicación heterocigota o deleción heterocigota, respectivamente. En algunas implementaciones, una duplicación homocigota o deleción homocigota se categoriza cuando una primera elevación o nivel es significativamente diferente de una segunda elevación o nivel y la primera elevación o nivel se encuentra dentro del rango de elevación o nivel esperado para una duplicación homocigota o deleción homocigota, respectivamente.

Ajustes de nivel

En algunas implementaciones, se ajustan uno o más niveles. Un procedimiento para ajustar un nivel se denomina a menudo relleno. En algunas implementaciones, se ajustan múltiples niveles en un perfil (por ejemplo, un perfil de un genoma, un perfil cromosómico, un perfil de una porción o un segmento de un cromosoma). En algunas implementaciones, se ajustan de aproximadamente 1 a aproximadamente 10.000 o más niveles en un perfil. En algunas implementaciones, se ajustan de aproximadamente 1 a aproximadamente 1000, de 1 a aproximadamente 900, de 1 a aproximadamente 800, de 1 a aproximadamente 700, de 1 a aproximadamente 600, de 1 a aproximadamente 500, de 1 a aproximadamente 400, de 1 a aproximadamente 300, de 1 a aproximadamente 200, de 1 a aproximadamente 100, de 1 a aproximadamente 50, de 1 a aproximadamente 25, de 1 a aproximadamente 20, de 1 a aproximadamente 15, de 1 a aproximadamente 10 o de 1 a aproximadamente 5 niveles en un perfil. En algunas implementaciones se ajusta un nivel. En algunas implementaciones, se ajusta un nivel(por ejemplo, un primer nivel de un perfil de recuento normalizado) que difiere significativamente de un segundo nivel. En algunas implementaciones se ajusta un nivel categorizado como variación del número de copias. En algunas

implementaciones, un nivel (por ejemplo, un primer nivel de un perfil de recuento normalizado) que difiere significativamente de un segundo nivel se categoriza como una variación del número de copias (por ejemplo, una variación del número de copias, por ejemplo, una variación del número de copias materno) y se ajusta. En algunas implementaciones, un nivel (por ejemplo, un primer nivel) se encuentra dentro de un rango de nivel esperado para una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal, y el nivel se ajusta. En algunas implementaciones, uno o más niveles (por ejemplo, los niveles de un perfil) no se ajustan. En algunas implementaciones, un nivel (por ejemplo, un primer nivel) está fuera de un rango de nivel esperado para una variación del número de copias, y el nivel no se ajusta. A menudo, un nivel dentro de un rango de nivel esperado para la ausencia de una variación del número de copias no se ajusta. Puede realizarse cualquier número adecuado de ajustes a uno o más niveles en un perfil. En algunas implementaciones, se ajustan uno o más niveles. En algunas implementaciones, se ajustan 2 o más, 3 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más y, algunas veces, 10 o más niveles.

En algunas implementaciones, el valor de un primer nivel se ajusta según el valor de un segundo nivel. En algunas implementaciones, un primer nivel, identificado como representativo de una variación del número de copias, se ajusta al valor de un segundo nivel, donde el segundo nivel suele estar asociado a ninguna variación del número de copias. En ciertas implementaciones, un valor de un primer nivel, identificado como representativo de una variación del número de copias, se ajusta de modo que el valor del primer nivel sea aproximadamente igual al valor de un segundo nivel.

Un ajuste puede comprender una operación matemática adecuada. En algunas implementaciones un ajuste comprende una o más operaciones matemáticas. En algunas implementaciones un nivel se ajusta normalizando, filtrando, promediando, multiplicando, dividiendo, sumando o restando o una combinación de los mismos. En algunas implementaciones, un nivel se ajusta mediante un valor predeterminado o una constante. En algunas implementaciones 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, valor medio, normalizado y similares).

En algunas implementaciones, un nivel se categoriza como una variación del número de copias (por ejemplo, una variación del número de copias materno) y se ajusta según un valor predeterminado denominado en el presente documento valor de ajuste predeterminado (PAV). A menudo, se determina un PAV para una variación específica del número de copias. A menudo, un PAV determinado para una variación del número de copias específica (por ejemplo, duplicación homocigota, deleción homocigota, duplicación heterocigota, deleción heterocigota) se usa para ajustar un nivel categorizado como una variación del número de copias específica (por ejemplo, duplicación homocigota, deleción homocigota, duplicación heterocigota, deleción heterocigota). En determinadas implementaciones, un nivel se categoriza como una variación del número de copias y, después, se ajusta según un PAV específico para el tipo de variación del número de copias categorizada. En algunas implementaciones, un nivel (por ejemplo, un primer nivel) se categoriza como una variación del número de copias materno, variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal y se ajusta sumando o restando un PAV del nivel. A menudo, un nivel (por ejemplo, un primer nivel) se categoriza como una variación del número de copias materno y se ajusta mediante la suma de un PAV al nivel. Por ejemplo, un nivel categorizado como una duplicación (por ejemplo, una duplicación homocigota materna, fetal o materna y fetal) puede ajustarse sumando un PAV determinado para una duplicación específica (por ejemplo, una duplicación homocigota) proporcionando de ese modo un nivel ajustado. A menudo, un PAV determinado para una duplicación del número de copias es un valor negativo. En algunas implementaciones, proporcionar un ajuste a un nivel representativo de una duplicación usando un PAV determinado para una duplicación da como resultado una reducción del valor del nivel. En algunas implementaciones, un nivel (por ejemplo, un primer nivel) que difiere significativamente de un segundo nivel se categoriza como una deleción del número de copias (por ejemplo, una deleción homocigota, deleción heterocigota, duplicación homocigota, duplicación homocigota) y el primer nivel se ajusta mediante la suma de un PAV determinado para una deleción del número de copias. A menudo, un PAV determinado para una deleción del número de copias es un valor de positivo. En algunas implementaciones, el proporcionar un ajuste a un nivel representativo de una deleción utilizando un PAV determinado para una deleción da como resultado un aumento del valor del nivel.

Un PAV puede ser cualquier valor adecuado. A menudo, un PAV se determina según una variación del número de copias y es específico para esta (por ejemplo, una variación del número de copias categorizada). En determinadas implementaciones, un PAV se determina según un nivel esperado para una variación del número de copias (por ejemplo, una variación del número de copias categorizada) y/o un factor de PAV. A veces, un PAV se determina multiplicando un nivel esperado por un factor de PAV. Por ejemplo, un PAV para una variación del número de copias puede determinarse multiplicando un nivel esperado determinado para una variación del número de copias (por ejemplo, una deleción heterocigota) por un factor de PAV determinado para la misma variación del número de copias (por ejemplo, una deleción heterocigota). Por ejemplo, un PAV puede determinarse mediante la siguiente fórmula:

$$PAV_K = (\text{nivel esperado})_k \times (\text{factor de PAV})_k$$

para la variación del número de copias k (por ejemplo, k = una deleción heterocigota)

Un factor de PAV puede ser cualquier valor adecuado. En algunas implementaciones, un factor de PAV para una duplicación homocigota es de entre aproximadamente -0,6 y aproximadamente -0,4. En algunas implementaciones, un factor de PAV para una duplicación homocigota 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 de PAV para una duplicación homocigota es de aproximadamente -0,5.

Por ejemplo, para un NRV de aproximadamente 1 y un nivel esperado de una duplicación homocigota igual a aproximadamente 2, el PAV para la duplicación homocigota se determina como de aproximadamente -1 según la fórmula anterior. En este caso, un primer nivel categorizado como una duplicación homocigota se ajusta sumando aproximadamente -1 al valor del primer nivel, por ejemplo.

En algunas implementaciones, un factor de PAV para una duplicación heterocigota es de entre aproximadamente -0,4 y aproximadamente -0,2. En algunas implementaciones, un factor de PAV para una duplicación heterocigota 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 de PAV para una duplicación heterocigota es de aproximadamente -0,33.

Por ejemplo, para un NRV de aproximadamente 1 y un nivel esperado de una duplicación heterocigota igual a aproximadamente 1,5, el PAV para la duplicación homocigota se determina como de aproximadamente -0,495 según la fórmula anterior. En este caso, un primer nivel categorizado como una duplicación heterocigota se ajusta sumando aproximadamente -0,495 al valor del primer nivel, por ejemplo.

En algunas implementaciones, un factor de PAV para una deleción heterocigota es de entre aproximadamente 0,4 y aproximadamente 0,2. En algunas implementaciones, un factor de PAV para una deleción heterocigota 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 de PAV para una deleción heterocigota es de aproximadamente 0,33.

Por ejemplo, para un NRV de aproximadamente 1 y un nivel esperado de una deleción heterocigota igual a aproximadamente 0,5, el PAV para la deleción heterocigota se determina como de aproximadamente 0,495 según la fórmula anterior. En este caso, un primer nivel categorizado como una deleción heterocigota se ajusta sumando aproximadamente 0,495 al valor del primer nivel, por ejemplo.

En algunas implementaciones, un factor de PAV para una deleción homocigota es de entre aproximadamente 0,6 y aproximadamente 0,4. En algunas implementaciones, un factor de PAV para una deleción homocigota 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 de PAV para una deleción homocigota es de aproximadamente 0,5.

Por ejemplo, para un NRV de aproximadamente 1 y un nivel esperado de una deleción homocigota igual a aproximadamente 0, el PAV para la deleción homocigota se determina como de aproximadamente 1 según la fórmula anterior. En este caso, un primer nivel categorizado como una deleción homocigota se ajusta sumando aproximadamente 1 al valor del primer nivel, por ejemplo.

En determinadas implementaciones, un PAV es aproximadamente igual o igual a un nivel esperado para una variación del número de copias (por ejemplo, el nivel esperado de una variación del número de copias).

En algunas implementaciones, los recuentos de un nivel se normalizan antes de realizar un ajuste. En determinadas implementaciones, los recuentos de algunos o todos los niveles de un perfil se normalizan antes de realizar un ajuste. Por ejemplo, los recuentos de un nivel pueden normalizarse según los recuentos de un nivel de referencia o de un NRV. En determinadas implementaciones, 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 todos los demás niveles (por ejemplo, un primer nivel) de un perfil se normalizan en relación con los recuentos del mismo nivel de referencia o VRN antes de realizar un ajuste.

En algunas implementaciones, un nivel de un perfil resulta de uno o más ajustes. En determinadas implementaciones, un nivel de un perfil se determina después de ajustar uno o más niveles del perfil. En algunas implementaciones, un nivel de un perfil se vuelve a calcular después de realizar uno o más ajustes.

En algunas implementaciones, una variación del número de copias (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal) se determina (por ejemplo, se determina directa o indirectamente) a partir de un ajuste. Por ejemplo, un nivel de un perfil que se haya ajustado (por ejemplo, un primer nivel ajustado) puede identificarse como una variación del número de copias materno. En algunas implementaciones, la magnitud del ajuste indica el tipo de variación del número de copias (por ejemplo, deleción heterocigota, duplicación homocigota, y similares). En determinadas implementaciones, un nivel ajustado en un perfil puede identificarse como representativo de una variación del número de copias según el valor de un PAV para la variación del número de copias. Por ejemplo, para un perfil dado, PAV es de aproximadamente -1 para una duplicación homocigota, aproximadamente -0,5 para una duplicación heterocigota, aproximadamente 0,5 para una deleción heterocigota y aproximadamente 1 para una deleción homocigota. En el ejemplo anterior, un nivel ajustado en aproximadamente -1 puede identificarse como una duplicación homocigota, por ejemplo. En algunas implementaciones, una o más variaciones del número de copias pueden determinarse a partir de un perfil o un nivel que comprende uno o más ajustes.

En determinadas implementaciones, se comparan los niveles ajustados dentro de un perfil. En algunas implementaciones, las anomalías y los errores se identifican comparando los niveles ajustados. Por ejemplo, a menudo se comparan uno o más niveles ajustados de un perfil y un nivel concreto puede identificarse como una anomalía o un error. En algunas implementaciones, una anomalía o error se identifica dentro de una o más porciones que componen un nivel. Una anomalía o error puede identificarse dentro del mismo nivel (por ejemplo, en un perfil) o en uno o más niveles que representan porciones que son adyacentes, contiguas, anexas o colindantes. En algunas implementaciones, uno o más niveles ajustados son niveles de porciones que son adyacentes, contiguas, anexas o colindantes donde se comparan los uno o más niveles ajustados y se identifica una anomalía o error. Una anomalía o un error puede ser un pico o depresión en un perfil o nivel donde se conoce o desconoce una causa del pico o la depresión. En determinadas implementaciones se comparan los niveles ajustados y se identifica una anomalía o error donde la anomalía o error se debe a un error estocástico, sistemático, aleatorio o del usuario. En algunas implementaciones se comparan los niveles ajustados y se elimina una anomalía o error de un perfil. En determinadas implementaciones, se comparan los niveles ajustados y se ajusta una anomalía o error.

Ajustes de elevación

En algunas implementaciones, se ajustan una o más elevaciones. Un procedimiento para ajustar una elevación se denomina a menudo relleno. En algunas implementaciones, se ajustan múltiples elevaciones en un perfil (por ejemplo, un perfil de un genoma, un perfil cromosómico, un perfil de una porción o un segmento de un cromosoma). En algunas implementaciones, se ajustan de aproximadamente 1 a aproximadamente 10.000 o más elevaciones en un perfil. En algunas implementaciones, se ajustan de aproximadamente 1 a aproximadamente 1000, de 1 a aproximadamente 900, de 1 a aproximadamente 800, de 1 a aproximadamente 700, de 1 a aproximadamente 600, de 1 a aproximadamente 500, de 1 a aproximadamente 400, de 1 a aproximadamente 300, de 1 a aproximadamente 200, de 1 a aproximadamente 100, de 1 a aproximadamente 50, de 1 a aproximadamente 25, de 1 a aproximadamente 20, de 1 a aproximadamente 15, de 1 a aproximadamente 10 o de 1 a aproximadamente 5 elevaciones en un perfil. En algunas implementaciones, se ajusta una elevación. En algunas implementaciones, se ajusta una elevación (por ejemplo, una primera elevación de un perfil de recuento normalizado) que difiere significativamente de una segunda elevación. En algunas implementaciones, se ajusta una elevación categorizada como variación del número de copias. En algunas implementaciones, una elevación (por ejemplo, una primera elevación de un perfil de recuento normalizado) que difiere significativamente de una segunda elevación se categoriza como una variación del número de copias (por ejemplo, una variación del número de copias, por ejemplo, una variación del número de copias materno) y se ajusta. En algunas implementaciones, una elevación (por ejemplo, una primera elevación) está dentro de un rango de elevación esperado para una variación del número de copias materno, la variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal y se ajusta la elevación. En algunas implementaciones, una o más elevaciones (por ejemplo, las elevaciones de un perfil) no se ajustan. En algunas implementaciones, una elevación (por ejemplo, una primera elevación) está fuera de un rango de elevación esperado para una variación del número de copias y la elevación no se ajusta. A menudo, no se ajusta una elevación dentro de un rango de elevación esperado para la ausencia de una variación del número de copias. Puede realizarse cualquier número adecuado de ajustes a una o más elevaciones en un perfil. En algunas implementaciones, se ajustan una o más elevaciones. En algunas implementaciones, se ajustan 2 o más, 3 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más y, algunas veces, 10 o más elevaciones.

En algunas implementaciones, un valor de una primera elevación se ajusta según un valor de una segunda elevación. En algunas implementaciones, una primera elevación, identificada como representativa de una variación del número de copias, se ajusta al valor de una segunda elevación, donde la segunda elevación suele estar asociada a una variación nula del número de copias. En determinadas implementaciones, un valor de una primera elevación, identificado como representativo de una variación del número de copias, se ajusta de modo que el valor de la primera elevación es aproximadamente igual a un valor de una segunda elevación. En algunas implementaciones, una elevación se ajusta normalizando, filtrando, promediando, multiplicando, dividiendo, sumando o restando o una combinación de los mismos. Algunas veces, se ajusta un nivel mediante un valor predeterminado o una constante. En algunas implementaciones, se ajusta una elevación modificando el valor de la elevación al valor de otra elevación. Por ejemplo, una primera elevación puede ajustarse modificando su valor al valor de una segunda elevación. Un valor en tales casos puede ser un valor procesado (por ejemplo, valor medio, normalizado y similares).

En algunas implementaciones, una elevación se categoriza como una variación del número de copias (por ejemplo, una variación del número de copias materno) y se ajusta según un valor predeterminado denominado en el presente documento valor de ajuste predeterminado (PAV). A menudo, se determina un PAV para una variación específica del número de copias. A menudo, un PAV determinado para una variación del número de copias específica (por ejemplo, duplicación homocigota, deleción homocigota, duplicación heterocigota, deleción heterocigota) se usa para ajustar una elevación categorizada como una variación del número de copias específica (por ejemplo, duplicación homocigota, deleción homocigota, duplicación heterocigota, deleción heterocigota). En determinadas implementaciones, una elevación se categoriza como una variación del número de copias y, después, se ajusta según un PAV específico para el tipo de variación del número de copias categorizada. En algunas implementaciones, una elevación (por ejemplo, una primera elevación) se categoriza como una variación del número de copias materno, variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal y se ajusta sumando o restando un PAV de la elevación. A menudo, una elevación (por ejemplo, una primera elevación) se categoriza como una variación del número de copias materno y se ajusta mediante la suma de un PAV a la elevación. Por ejemplo, una elevación categorizada como una duplicación (por ejemplo, una

duplicación homocigota materna, fetal o materna y fetal) puede ajustarse sumando un PAV determinado para una duplicación específica (por ejemplo, una duplicación homocigota) proporcionando de ese modo una elevación ajustada. A menudo, un PAV determinado para una duplicación del número de copias es un valor negativo. En algunas implementaciones, proporcionar un ajuste a una elevación representativa de una duplicación usando un PAV determinado para una duplicación da como resultado una reducción del valor de la elevación. En algunas implementaciones, una elevación (por ejemplo, una primera elevación) que difiere significativamente de una segunda elevación se categoriza como una deleción del número de copias (por ejemplo, una deleción homocigota, deleción heterocigota, duplicación homocigota, duplicación homocigota) y la primera elevación se ajusta mediante la suma de un PAV determinado para una deleción del número de copias. A menudo, un PAV determinado para una deleción del número de copias es un valor de positivo. En algunas implementaciones el proporcionar un ajuste a una elevación representativa de una deleción utilizando un PAV determinado para una deleción da como resultado un aumento del valor de la elevación.

Un PAV puede ser cualquier valor adecuado. A menudo, un PAV se determina según una variación del número de copias y es específico para esta (por ejemplo, una variación del número de copias categorizada). En determinadas implementaciones, un PAV se determina según una elevación esperada para una variación del número de copias (por ejemplo, una variación del número de copias categorizada) y/o un factor de PAV. A veces, un PAV se determina multiplicando una elevación esperada por un factor de PAV. Por ejemplo, un PAV para una variación del número de copias puede determinarse multiplicando una elevación esperada determinada para una variación del número de copias (por ejemplo, una deleción heterocigota) por un factor de PAV determinado para la misma variación del número de copias (por ejemplo, una deleción heterocigota). Por ejemplo, un PAV puede determinarse mediante la siguiente fórmula:

$$PAV_k = (\text{elevación esperada})_k \times (\text{factor de PAV})_x$$

para la variación del número de copias k (por ejemplo, k = una deleción heterocigota)

Un factor de PAV puede ser cualquier valor adecuado. En algunas implementaciones, un factor de PAV para una duplicación homocigota es de entre aproximadamente -0,6 y aproximadamente -0,4. En algunas implementaciones, un factor de PAV para una duplicación homocigota 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 de PAV para una duplicación homocigota es de aproximadamente -0,5.

Por ejemplo, para un NRV de aproximadamente 1 y una elevación esperada de una duplicación homocigota igual a aproximadamente 2, el PAV para la duplicación homocigota se determina como de aproximadamente -1 según la fórmula anterior. En este caso, una primera elevación categorizada como una duplicación homocigota se ajusta sumando aproximadamente -1 al valor de la primera elevación, por ejemplo.

En algunas implementaciones, un factor de PAV para una duplicación heterocigota es de entre aproximadamente -0,4 y aproximadamente -0,2. En algunas implementaciones, un factor de PAV para una duplicación heterocigota 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 de PAV para una duplicación heterocigota es de aproximadamente -0,33.

Por ejemplo, para un NRV de aproximadamente 1 y una elevación esperada de una duplicación heterocigota igual a aproximadamente 1,5, el PAV para la duplicación homocigota se determina como de aproximadamente -0,495 según la fórmula anterior. En este caso, una primera elevación categorizada como una duplicación heterocigota se ajusta sumando aproximadamente -0,495 al valor de la primera elevación, por ejemplo.

En algunas implementaciones, un factor de PAV para una deleción heterocigota es de entre aproximadamente 0,4 y aproximadamente 0,2. En algunas implementaciones, un factor de PAV para una deleción heterocigota 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 de PAV para una deleción heterocigota es de aproximadamente 0,33.

Por ejemplo, para un NRV de aproximadamente 1 y una elevación esperada de una deleción heterocigota igual a aproximadamente 0,5, el PAV para la deleción heterocigota se determina como de aproximadamente 0,495 según la fórmula anterior. En este caso, una primera elevación categorizada como una deleción heterocigota se ajusta sumando aproximadamente 0,495 al valor de la primera elevación, por ejemplo.

En algunas implementaciones, un factor de PAV para una deleción homocigota es de entre aproximadamente 0,6 y aproximadamente 0,4. En algunas implementaciones, un factor de PAV para una deleción homocigota 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 de PAV para una deleción homocigota es de aproximadamente 0,5.

Por ejemplo, para un NRV de aproximadamente 1 y una elevación esperada de una deleción homocigota igual a aproximadamente 0, el PAV para la deleción homocigota se determina como de aproximadamente 1 según la fórmula anterior. En este caso, una primera elevación categorizada como una deleción homocigota se ajusta sumando aproximadamente 1 al valor de la primera elevación, por ejemplo.

En determinadas implementaciones, un PAV es aproximadamente igual o igual a una elevación esperada para una variación del número de copias (por ejemplo, la elevación esperada de una variación del número de copias).

En algunas implementaciones, los recuentos de una elevación se normalizan antes de realizar un ajuste. En determinadas implementaciones, los recuentos de algunas o todas las elevaciones de un perfil se normalizan antes de realizar un ajuste. Por ejemplo, los recuentos de una elevación pueden normalizarse según los recuentos de una elevación de referencia o un NRV. En determinadas implementaciones, los recuentos de una elevación (por ejemplo, una segunda elevación) se normalizan según los recuentos de una elevación de referencia o un NRV, y los recuentos de todas las demás elevaciones (por ejemplo, una primera elevación) en un perfil se normalizan en relación con los recuentos de la misma elevación de referencia o NRV antes de realizar un ajuste.

En algunas implementaciones, una elevación de un perfil resulta de uno o más ajustes. En determinadas implementaciones, una elevación de un perfil se determina después de ajustar una o más elevaciones en el perfil. En algunas implementaciones, una elevación de un perfil se calcula de nuevo después de realizar uno o más ajustes.

En algunas implementaciones, una variación del número de copias (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal) se determina (por ejemplo, se determina directa o indirectamente) a partir de un ajuste. Por ejemplo, una elevación en un perfil que se ajustó (por ejemplo, una primera elevación ajustada) puede identificarse como una variación del número de copias materno. En algunas implementaciones, la magnitud del ajuste indica el tipo de variación del número de copias (por ejemplo, deleción heterocigota, duplicación homocigota, y similares). En determinadas implementaciones, una elevación ajustada en un perfil puede identificarse como representativa de una variación del número de copias según el valor de un PAV para la variación del número de copias. Por ejemplo, para un perfil dado, PAV es de aproximadamente -1 para una duplicación homocigota, aproximadamente -0,5 para una duplicación heterocigota, aproximadamente 0,5 para una deleción heterocigota y aproximadamente 1 para una deleción homocigota. En el ejemplo anterior, una elevación ajustada en aproximadamente -1 puede identificarse como una duplicación homocigota, por ejemplo. En algunas implementaciones, una o más variaciones del número de copias pueden determinarse a partir de un perfil o una elevación que comprende uno o más ajustes.

En determinadas implementaciones, se comparan las elevaciones ajustadas dentro de un perfil. En algunas implementaciones, las anomalías y los errores se identifican comparando las elevaciones ajustadas. Por ejemplo, a menudo se comparan una o más elevaciones ajustadas en un perfil y una elevación particular puede identificarse como una anomalía o un error. En algunas implementaciones, una anomalía o error se identifica dentro de una o más porciones que componen una elevación. Una anomalía o error puede identificarse dentro de la misma elevación (por ejemplo, en un perfil) o en una o más elevaciones que representan porciones que son adyacentes, contiguas, anexas o colindantes. En algunas implementaciones, una o más elevaciones ajustadas son niveles de porciones que son adyacentes, contiguas, anexas o colindantes donde se comparan los uno o más elevaciones ajustadas y se identifica una anomalía o error. Una anomalía o un error puede ser un pico o depresión en un perfil o elevación en el que se conoce o desconoce una causa del pico o la depresión. En determinadas implementaciones, se comparan las elevaciones ajustadas y se identifica una anomalía o error donde la anomalía o error se debe a un error estocástico, sistemático, aleatorio o del usuario. En algunas implementaciones se comparan las elevaciones ajustadas y se elimina una anomalía o error de un perfil. En determinadas implementaciones, se comparan las elevaciones ajustadas y se ajusta una anomalía o error.

Determinación de la fracción fetal en función de una elevación o un nivel

En algunas implementaciones, una fracción fetal se determina según una elevación o un nivel categorizados como representativos de una variación del número de copias materno y/o fetal. Por ejemplo, determinar la fracción fetal comprende a menudo evaluar una elevación o un nivel esperado para una variación del número de copias materno y/o fetal utilizada para la determinación de fracción fetal. En algunas implementaciones, se determina una fracción fetal para una elevación o un nivel (por ejemplo, una primera elevación o nivel) categorizado como representativo de una variación del número de copias según un rango de elevación o nivel esperado determinado para el mismo tipo de variación del número de copias. A menudo, una fracción fetal se determina según una elevación o nivel observado que cae dentro de un rango de elevación o nivel esperado y, por lo tanto, se categoriza como una variación del número de copias materno y/o fetal. En algunas implementaciones, se determina una fracción fetal cuando una elevación o nivel observado (por ejemplo, una primera elevación o nivel) categorizado como una variación del número de copias materno y/o fetal es diferente de la elevación o nivel esperado determinado para el mismo tipo de variación del número de copias materno y/o fetal.

En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado), es significativamente diferente de una segunda elevación o nivel, la primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones, una primera elevación o nivel es una elevación o nivel observado y/u obtenido experimentalmente que es significativamente diferente de una segunda elevación o nivel en un perfil y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones, la primera elevación o nivel es una elevación o nivel promedio, medio o sumado y se determina una fracción fetal según la primera elevación o nivel. En determinadas implementaciones, una primera elevación o nivel y una segunda elevación o nivel son elevaciones o niveles

observados y/u obtenidos experimentalmente y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones, una primera elevación o nivel comprende recuentos normalizados para un primer conjunto de porciones y una segunda elevación o nivel comprende recuentos normalizados para un segundo conjunto de porciones, y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones, un primer conjunto de porciones de una primera elevación o nivel incluye una variación del número de copias (por ejemplo, la primera elevación o nivel es representativo de una variación del número de copias) y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones, el primer conjunto de porciones de una primera elevación o nivel incluye una variación del número de copias materno homocigota o heterocigota y se determina una fracción fetal según la primera elevación o nivel. En algunas implementaciones un perfil comprende una primera elevación o nivel para un primer conjunto de porciones y una segunda elevación o nivel para un segundo conjunto de porciones, el segundo conjunto de porciones no incluye sustancialmente ninguna variación del número de copias (por ejemplo, una variación del número de copias materno, una variación del número de copias fetal, o una variación del número de copias materno y una variación del número de copias fetal), y se determina una fracción fetal según la primera elevación o nivel.

En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado), es significativamente diferente de una segunda elevación o nivel, la primera elevación o nivel se categoriza como para una variación del número de copias materno y/o fetal, y se determina una fracción fetal según la primera elevación o nivel y/o una elevación o nivel esperado de la variación del número de copias. En algunas implementaciones, una primera elevación o un nivel se categoriza como para una variación del número de copias según una elevación o un nivel esperado para una variación del número de copias y una fracción fetal se determina según una diferencia entre la primera elevación o un nivel y la elevación o el nivel esperado. En determinadas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado) se categoriza como una variación del número de copias materno y/o fetal, y se determina una fracción fetal como el doble de la diferencia entre la primera elevación o nivel y la elevación o nivel esperado de la variación del número de copias. En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado) se categoriza como una variación del número de copias materno y/o fetal, la primera elevación o nivel se resta de la elevación o nivel esperado proporcionando de este modo una diferencia, y se determina una fracción fetal como el doble de la diferencia. En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado) se categoriza como una variación del número de copias materno y/o fetal, una elevación o nivel esperado se resta de una primera elevación o nivel proporcionando de este modo una diferencia, y se determina la fracción fetal como el doble de la diferencia.

A menudo, se proporciona una fracción fetal como un porcentaje. Por ejemplo, una fracción fetal puede dividirse entre 100 proporcionando de ese modo un valor porcentual. Por ejemplo, para una primera elevación o nivel representativo de una duplicación homocigota materna y que tenga una elevación o un nivel de 155 y una elevación o nivel esperado para una duplicación homocigota materna que tenga una elevación o un nivel de 150, puede determinarse una fracción fetal como del 10 % (por ejemplo, (fracción fetal = 2 x (155-150)).

En algunas implementaciones, se determina una fracción fetal a partir de dos o más elevaciones o niveles dentro de un perfil que se categorizan como variaciones del número de copias. Por ejemplo, a veces dos o más elevaciones o niveles (por ejemplo, dos o más primeras elevaciones o niveles) en un perfil se identifican como significativamente diferentes de una elevación o nivel de referencia (por ejemplo, una segunda elevación o nivel, una elevación o un nivel que no incluye sustancialmente ninguna variación del número de copias), las dos o más elevaciones o niveles se categorizan como representativos de una variación del número de copias materno y/o fetal, y se determina una fracción fetal a partir de cada una de las dos o más elevaciones o niveles. En algunas implementaciones, se determina una fracción fetal 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 algunas implementaciones, se determina una fracción fetal 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 algunas implementaciones, se determina una fracción fetal 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 algunas implementaciones, se determina una fracción fetal a partir de aproximadamente 10 a aproximadamente 1000, de aproximadamente 20 a aproximadamente 900, de aproximadamente 30 a aproximadamente 700, de aproximadamente 40 a aproximadamente 600, de aproximadamente 50 a aproximadamente 500, de aproximadamente 50 a aproximadamente 400, de aproximadamente 50 a aproximadamente 300, de aproximadamente 50 a aproximadamente 200 o de aproximadamente 50 a aproximadamente 100 determinaciones de fracción fetal dentro de un perfil.

En algunas implementaciones, se determina una fracción fetal como el promedio o la media de múltiples determinaciones de fracciones fetales dentro de un perfil. En determinadas implementaciones, una fracción fetal determinada a partir de múltiples determinaciones de fracciones fetales es una media (por ejemplo, un promedio, un promedio estándar, una mediana, o similares) de múltiples determinaciones de fracción fetal. A menudo, una fracción fetal determinada a partir de múltiples determinaciones de fracción fetal es un valor medio determinado mediante un método adecuado conocido en la técnica o descrito en el presente documento. En algunas implementaciones, un valor

medio de una determinación de fracción fetal es una media ponderada. En algunas implementaciones, un valor medio de una determinación de fracción fetal es una media no ponderada. Una determinación de fracción fetal media, mediana o promedio (es decir, un valor medio, de mediana o promedio de determinación de fracción fetal) generada a partir de múltiples determinaciones de fracción fetal está algunas veces asociada con una medida de incertidumbre (por ejemplo, una varianza, desviación estándar, DAM, o similares). Antes de determinar un valor medio, de mediana o promedio de fracción fetal a partir de múltiples determinaciones, se eliminan una o más determinaciones desviadas en algunas implementaciones (descritas con mayor detalle en el presente documento).

Algunas determinaciones de fracción fetal dentro de un perfil a veces no se incluyen en la determinación global de una fracción fetal (por ejemplo, determinación media o promedio de fracción fetal). En algunas implementaciones, la determinación de una fracción fetal se deriva de una primera elevación o nivel (por ejemplo, una primera elevación o nivel que es significativamente diferente de una segunda elevación o nivel) en un perfil y la primera elevación o nivel no es indicativo de una variación genética. Por ejemplo, algunas primeras elevaciones o niveles (por ejemplo, aumentos bruscos o depresiones) en un perfil se generan a partir de anomalías o causas desconocidas. Tales valores generan a menudo determinaciones de fracción fetal que difieren significativamente de otras determinaciones de fracción fetal obtenidas a partir de variaciones verdaderas del número de copias. En algunas implementaciones, las determinaciones de fracción fetal que difieren significativamente de otras determinaciones de fracción fetal en un perfil se identifican y eliminan de una determinación de fracción fetal. Por ejemplo, algunas determinaciones de fracción fetal obtenidas a partir de depresiones y aumentos bruscos anómalos se identifican comparándolas con otras determinaciones de fracción fetal dentro de un perfil y se excluyen de la determinación global de fracción fetal.

En algunas implementaciones, una determinación de fracción fetal independiente que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio es una diferencia identificada, reconocida y/u observable. En determinadas implementaciones, 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 algunas implementaciones, una única determinación) a partir de una elevación o nivel específico categorizado como variación del número de copias. Puede usarse cualquier umbral o rango adecuado para determinar que una determinación de fracción fetal difiere significativamente de una determinación de fracción fetal media, en mediana o promedio. En determinadas implementaciones, una determinación de fracción fetal difiere significativamente de una determinación de fracción fetal media, en mediana o promedio y la determinación puede expresarse como una desviación porcentual del valor promedio o medio. En determinadas implementaciones, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio difiere en aproximadamente el 10 por ciento o más. En algunas implementaciones, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio difiere en aproximadamente el 15 por ciento o más. En algunas implementaciones, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio difiere en aproximadamente el 15 % a aproximadamente el 100 % o más.

En determinadas implementaciones, una determinación de fracción fetal difiere significativamente de una determinación de fracción fetal media, mediana o promedio según un múltiplo de una medida de incertidumbre asociada a la determinación de fracción fetal media o promedio. A menudo, una medida de incertidumbre y una constante n (por ejemplo, un intervalo de confianza) definen un intervalo (por ejemplo, un valor de corte de incertidumbre). Por ejemplo, a veces una medida de incertidumbre es una desviación estándar para las determinaciones de la fracción fetal (por ejemplo, +/- 5) y se multiplica por una constante n (por ejemplo, un intervalo de confianza), definiendo de este modo un rango o valor de corte de incertidumbre (por ejemplo, de 5n a -5n, a veces denominado 5 sigma). En algunas implementaciones, una determinación de fracción fetal independiente se encuentra fuera de un rango definido por el punto de corte de incertidumbre y se considera significativamente diferente de una determinación de fracción fetal media, en mediana o promedio. Por ejemplo, para un valor medio de 10 y un punto de corte de incertidumbre de 3, una fracción fetal independiente mayor de 13 o menor de 7 es significativamente diferente. En algunas implementaciones, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio difiere en más de n veces la medida de incertidumbre (por ejemplo, n x sigma), donde n es aproximadamente igual o superior a 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10. En algunas implementaciones, una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio difiere en más de n veces la medida de incertidumbre (por ejemplo, n x sigma), donde n es aproximadamente igual o superior a 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.

En algunas implementaciones, una elevación o un nivel es representativo de una microploidía fetal y/o materna. En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado), es significativamente diferente de una segunda elevación o nivel, la primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, y la primera elevación o nivel y/o la segunda elevación o nivel es representativo de una microploidía fetal y/o una microploidía materna. En determinadas implementaciones, una primera elevación o nivel es representativo de una microploidía fetal. En algunas implementaciones, una primera elevación o nivel es representativo de una microploidía materna. A menudo, una primera elevación o nivel es representativo de una microploidía fetal y de una microploidía materna. En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado), es significativamente diferente de una segunda elevación o nivel, la primera elevación o nivel se categoriza como una variación del número de copias

materno y/o fetal, la primera elevación o nivel es representativo de una microploidía fetal y/o materna, y se determina una fracción fetal según la microploidía fetal y/o materna. En algunos casos, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, la primera elevación o nivel es representativo de una microploidía fetal y se determina una fracción fetal según la microploidía fetal. En algunas implementaciones, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, la primera elevación o nivel es representativo de una microploidía materna y se determina una fracción fetal según la microploidía materna. En algunas implementaciones, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, la primera elevación o nivel es representativo de una microploidía materna y una microploidía fetal, y se determina una fracción fetal según la microploidía materna y fetal.

En algunas implementaciones, una determinación de una fracción fetal comprende determinar una microploidía fetal y/o materna. En algunas implementaciones, una elevación o un nivel (por ejemplo, una primera elevación o nivel, una elevación o nivel observado), es significativamente diferente de una segunda elevación o nivel, la primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, se determina una microploidía fetal y/o materna según la primera elevación o nivel y/o la segunda elevación o nivel y se determina una fracción fetal. En algunas implementaciones, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, se determina una microploidía fetal según la primera elevación o nivel y/o la segunda elevación o nivel y se determina una fracción fetal según la microploidía fetal. En determinadas implementaciones, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, se determina una microploidía materna según la primera elevación o nivel y/o la segunda elevación o nivel y se determina una fracción fetal según la microploidía materna. En algunas implementaciones, una primera elevación o nivel se categoriza como una variación del número de copias materno y/o fetal, se determina una microploidía materna y fetal según la primera elevación o nivel y/o la segunda elevación o nivel y se determina una fracción fetal según la microploidía materna y fetal.

A menudo, se determina una fracción fetal cuando la microploidía de la madre es diferente (por ejemplo, no igual) a la microploidía del feto para una elevación o nivel determinado o para una elevación o un nivel categorizado como variación del número de copias. En algunas implementaciones, se determina una fracción fetal cuando la madre es homocigota para una duplicación (por ejemplo, una microploidía de 2) y el feto es heterocigoto para la misma duplicación (por ejemplo, una microploidía de 1,5). En algunas implementaciones, se determina una fracción fetal cuando la madre es heterocigota para una duplicación (por ejemplo, una microploidía de 1,5) y el feto es homocigoto 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 algunas implementaciones, se determina una fracción fetal cuando la madre es homocigota para una deleción (por ejemplo, una microploidía de 0) y el feto es heterocigoto para la misma deleción (por ejemplo, una microploidía de 0,5). En algunas implementaciones, se determina una fracción fetal cuando la madre es heterocigota para una deleción (por ejemplo, una microploidía de 0,5) y el feto es homocigoto para la misma deleción (por ejemplo, una microploidía de 0) o la deleción está ausente en el feto (por ejemplo, una microploidía de 1).

En determinadas implementaciones, no puede determinarse una fracción fetal cuando la microploidía de la madre es la misma (por ejemplo, identificada como la misma) que la microploidía del feto para una elevación o nivel dado identificado como una variación del número de copias. Por ejemplo, para una elevación o nivel dado donde tanto la madre como el feto portan el mismo número de copias de una variación del número de copias, no se determina una fracción fetal, en algunas implementaciones. Por ejemplo, no puede determinarse una fracción fetal para una elevación o un nivel categorizado como una variación del número de copias cuando tanto la madre como el feto son homocigotos para la misma deleción u homocigotos para la misma duplicación. En determinadas implementaciones, no puede determinarse una fracción fetal para una elevación o un nivel categorizado como una variación del número de copias cuando tanto la madre como el feto son heterocigotos para la misma deleción o heterocigotos para la misma duplicación. En implementaciones en las que se realizan múltiples determinaciones de fracción fetal para una muestra, las determinaciones que se desvían significativamente de una media, mediana o valor promedio pueden resultar de una variación del número de copias para la cual la ploidía materna es igual a la ploidía fetal, y tales determinaciones pueden eliminarse de la consideración.

En algunas implementaciones se desconoce la microploidía de una variación del número de copias materno y la variación del número de copias fetal. En algunas implementaciones, en casos done no hay determinación de microploidía fetal y/o materna para una variación del número de copias, se genera una fracción fetal y se compara con una determinación de fracción fetal media, en mediana o promedio. Una determinación de fracción fetal para una variación del número de copias que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio se debe a que a veces la microploidía de la madre y el feto son las mismas para la variación del número de copias. Una determinación de fracción fetal que difiere significativamente de una determinación de fracción fetal media, en mediana o promedio se excluye a menudo de una determinación global de fracción fetal independientemente de la fuente o causa de la diferencia. En algunas implementaciones, 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).

Determinación de correspondencias

En algunas implementaciones, una correspondencia es una correspondencia geométrica y/o gráfica. En algunas implementaciones, una correspondencia es una correspondencia matemática. En algunas implementaciones, se representa gráficamente una correspondencia. En algunas implementaciones, una correspondencia es una

correspondencia lineal. En determinadas implementaciones, una correspondencia es una correspondencia no lineal. En determinadas implementaciones, una correspondencia es una regresión (por ejemplo, una línea de regresión). Una regresión puede ser una regresión lineal o una regresión no lineal. Una correspondencia se puede expresar mediante una ecuación matemática. A menudo, una correspondencia se define, en parte, por una o más constantes.

Determinación de la ploidía fetal

Una determinación de ploidía fetal, en algunas implementaciones, se usa, en parte, para determinar la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica, una trisomía). Una ploidía fetal puede determinarse, 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 algunas implementaciones, la ploidía fetal se determina según una determinación de la fracción fetal y la ecuación (8), (20), (21) o una variación o derivación de la misma (Ejemplo 2). En algunas implementaciones, la ploidía fetal se determina mediante un método que se describe a continuación. En algunas implementaciones, 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, un bin o una porción, i) de un genoma para múltiples muestras donde la ploidía del feto para la porción i del genoma es euploide. En algunas implementaciones, se determina una medida de incertidumbre (por ejemplo, una desviación estándar, σ) para el recuento de referencia f_i. En algunas implementaciones, se utilizan un recuento de referencia f_i, una medida de incertidumbre, un recuento de muestra de prueba y/o una fracción fetal medida (F) para determinar la ploidía fetal según un método descrito a continuación. En algunas implementaciones, se normaliza un recuento de referencia (por ejemplo, un recuento de referencia promedio, medio o en mediana) mediante un método descrito en el presente documento (por ejemplo, normalización en bins, normalización en porciones, normalización por contenido de GC, regresión lineal y no lineal de mínimos cuadrados, LOESS, GC de LOESS, LOWESS, PERUN, RM, GCRM y/o combinaciones de los mismos). En algunas implementaciones, un recuento de referencia de un segmento de un genoma que es euploide es igual a 1 cuando el recuento de referencia se normaliza mediante PERUN. En algunas implementaciones, tanto el recuento de referencia (por ejemplo, para un feto que se sabe que es euploide) como los recuentos de una muestra de prueba para una porción o segmento de un genoma están normalizados mediante PERUN y el recuento de referencia es igual a 1. Asimismo, en algunas implementaciones, 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, se dividen entre) mediante una mediana del recuento de referencia. Por ejemplo, en algunas implementaciones, tanto el recuento de referencia (por ejemplo, para un feto que es euploide) como los recuentos de una muestra de prueba para una porción o segmento de un genoma se normalizan mediante la mediana de un recuento de referencia, el recuento de referencia normalizado es igual a 1 y el recuento de la muestra de prueba se normaliza (por ejemplo, se divide entre) mediante la mediana del recuento de referencia. En algunas implementaciones, tanto el recuento de referencia (por ejemplo, para un feto que es euploide) como los recuentos de una muestra de prueba para una porción o segmento de un genoma se normalizan mediante GCRM, GC, RM o un método adecuado. En algunas implementaciones, un recuento de referencia es un recuento de referencia promedio, medio o en mediana. Un recuento de referencia suele ser un recuento normalizado para un bin o una porción (por ejemplo, una porción normalizada o un nivel de sección genómica). En algunas implementaciones, un recuento de referencia y los recuentos de una muestra de prueba son recuentos sin procesar. Un recuento de referencia, en algunas implementaciones, se determina a partir de un perfil de recuento promedio, medio o en mediana. En algunas implementaciones, un recuento de referencia es un nivel de porción o sección genómica calculado. En algunas implementaciones, un recuento de referencia de una muestra de referencia y un recuento de una muestra de prueba (por ejemplo, una muestra de paciente, por ejemplo, y_i) se normalizan mediante el mismo método o proceso.

En algunas implementaciones, se determina una medida de la fracción fetal (F). Este valor de la fracción fetal se utiliza luego para determinar la ploidía fetal según la ecuación (8), una derivación o una variación de la misma. En algunas implementaciones, se obtiene un valor negativo si el feto es euploide y un valor positivo si el feto no es euploide. En algunas implementaciones, un valor negativo indica que el feto es euploide para el segmento del genoma considerado. En determinadas implementaciones, un valor que no es negativo indica que el feto comprende una aneuploidía (por ejemplo, una duplicación). En determinadas implementaciones, un valor que no es negativo indica que el feto comprende una trisomía. En determinadas implementaciones, cualquier valor positivo indica que el feto comprende una aneuploidía (por ejemplo, una trisomía, una duplicación).

En algunas implementaciones se determina una suma de residuos al cuadrado. Por ejemplo, en la ecuación (18) se ilustra una ecuación que representa la suma de los residuos al cuadrado derivados de la ecuación (8). En algunas implementaciones, se determina una suma de residuos 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 algunas implementaciones, la suma de los residuos al cuadrado (ecuaciones (9) y (13)) se determina para un segmento de un genoma o cromosoma (por ejemplo, para todos los bins o porciones de un genoma de referencia i en un segmento del genoma). Por ejemplo, la suma de los residuos al cuadrado (por ejemplo, ecuaciones (9) y (13)) puede determinarse para los cromosomas 21, 13, 18 o una porción de los mismos. En algunas implementaciones, para determinar el estado de ploidía de un feto, se resta el resultado de la ecuación (13) de la ecuación (9) para llegar a un valor, phi (por ejemplo, consulte la ecuación (14)). En determinadas implementaciones, el signo (es decir, positivo o negativo) del valor de phi determina la presencia o ausencia de una aneuploidía fetal. En determinadas implementaciones, un valor de phi (por ejemplo, de la ecuación (14)) que es negativo indica la ausencia de una aneuploidía (por ejemplo, el feto es euploide para bins o porciones de un genoma de referencia i) y un valor de phi que no es negativo indica la presencia de una aneuploidía (por ejemplo, una trisomía).

En algunas implementaciones, el recuento de referencia f_i, la medida de incertidumbre para el recuento de referencia σ y/o la fracción fetal medida (F) se utilizan en las ecuaciones (9) y (13) para determinar la suma de los residuos al cuadrado para la suma de todos los bins i o porciones de un genoma de referencia i. En algunas implementaciones, el recuento de referencias f_i, la medida de incertidumbre para el recuento de referencia σ y/o la fracción fetal medida (F) se utilizan en las ecuaciones (9) y (13) para determinar la ploidía fetal. En algunas implementaciones, los recuentos (por ejemplo, recuentos normalizados, por ejemplo, nivel de porción o de sección genómica calculado), representados por y_i para el bin i o la porción i, para una muestra de prueba se utilizan para determinar el estado de ploidía de un feto para el bin i o la porción i. Por ejemplo, en determinadas implementaciones, el estado de ploidía de un segmento de un genoma se determina según un recuento de referencia f_i, una medida de incertidumbre (por ejemplo, del recuento de referencia), una fracción feta (F) determinada para una muestra de prueba y los recuentos y_i determinados para la muestra de prueba, donde el estado de ploidía se determina según la ecuación (14) o una derivación o variación de la misma. En algunas implementaciones, los recuentos y_i y/o los recuentos de referencia se normalizan mediante un método descrito en el presente documento (por ejemplo, normalización en bins, normalización en porciones, normalización por contenido de GC, regresión lineal y no lineal de mínimos cuadrados, LOESS, GC de LOESS, LOWESS, PERUN, RM, GCRM yo combinaciones de los mismos). En algunas implementaciones, un estado de ploidía fetal (por ejemplo, euploide, aneuploide, trisomía) para una porción o un segmento de un genoma o cromosoma se determina mediante el ejemplo no limitativo descrito anteriormente y en la sección de ejemplos.

En algunas implementaciones, se determina una fracción fetal a partir de una muestra de prueba, se determinan los recuentos y para una muestra de prueba, y ambos se utilizan para determinar una ploidía para un feto a partir de una muestra de prueba. En determinadas implementaciones del método descrito en el presente documento, el valor de la ploidía fetal representado por X no es fijo ni supuesto. En determinadas implementaciones del método descrito en el presente documento, la fracción fetal F es fija. En algunas implementaciones, se determina una ploidía (por ejemplo, un valor de ploidía) para una porción o segmento de un genoma según la ecuación (20) o (21) (Ejemplo 2). En algunas implementaciones de este método, se determina un valor de ploidía, donde el valor es cercano a 1, 3/2 o 5/4. En algunas implementaciones, 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 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 sobre la determinación de la presencia o ausencia de una aneuploidía fetal a partir de una determinación de ploidía fetal se analiza en otro apartado a continuación.

En algunas implementaciones, se determina la fracción fetal, se fija en su valor determinado, y se determina la ploidía fetal a partir de una regresión. Puede utilizarse cualquier regresión adecuada, cuyos ejemplos no limitativos incluyen una regresión lineal, una regresión no lineal (por ejemplo, una regresión polinomial) y similares. En algunas implementaciones se utiliza una regresión lineal según la ecuación (8), (20), (21) y/o una derivación o variación de la misma. En algunas implementaciones, la regresión lineal utilizada es según una suma de residuos al cuadrado derivados de la ecuación (8), (20), (21) y/o una derivación o variación de la misma. En algunas implementaciones, la ploidía fetal se determina según la ecuación (8), (20), (21) y/o una derivación o variación de la misma y no se utiliza una regresión. En algunas implementaciones, la ploidía fetal se determina según una suma de residuos al cuadrado derivados de la ecuación (8), (20), (21) y/o una derivación o variación de la misma para múltiples bins o porciones de un genoma de referencia i y no se utiliza una regresión. Una derivación de una ecuación es cualquier variación de la ecuación obtenida a partir de una prueba matemática de una ecuación.

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

La medida de incertidumbre σ puede ser una medida de error adecuada, cuyos ejemplos no limitativos incluyen la desviación estándar, el error estándar, la varianza calculada, el valor de *p* y/o la desviación absoluta media (DAM). La

medida de incertidumbre σ se puede determinar para cualquier medida adecuada, cuyos ejemplos no limitativos incluyen puntuaciones Z, valores de Z, valores de t, valores de *p*, error de validación cruzada, nivel de sección genómica o porción, niveles de sección genómica calculados, elevaciones o niveles, recuentos, similares, o combinaciones de los mismos. En algunas implementaciones, σ se establece en un valor de 1. En algunas implementaciones, σ no se establece en un valor de 1. En algunas implementaciones, se estima el valor de σ y a veces se mide y/o se calcula.

En algunas implementaciones, M_i es la ploidía de la madre (es decir, ploidía materna) para una porción del genoma i. En algunas implementaciones, M_i se determina para el mismo paciente (por ejemplo, la misma muestra de prueba) de la que se ha determinado y_i. En algunas implementaciones, la ploidía materna M_i se conoce o se determina según un método descrito en el presente documento. En algunas implementaciones, la ploidía materna se determina antes o después del relleno (por ejemplo, después de realizar ajustes de elevación o nivel). En determinadas implementaciones, M_i se estima o determina a partir de la visualización de un perfil. En algunas implementaciones, la ploidía materna M_i es desconocida. En algunas implementaciones, la ploidía materna M_i se supone. Por ejemplo, en algunas implementaciones, se supone o se sabe que la madre no tiene deleciones y/o duplicaciones en el segmento del genoma que se está evaluando. En algunas implementaciones, se supone o se sabe que la ploidía materna es 1. En algunas implementaciones, la ploidía materna se establece en un valor de 1 después del relleno (por ejemplo, después de realizar ajustes de elevaciones o niveles). En algunas implementaciones, la ploidía materna se ignora y se establece en un valor de 1. En algunas implementaciones, la ecuación (21) se deriva de la ecuación (20) suponiendo que la madre no tiene deleciones y/o duplicaciones en el segmento del genoma que se está evaluando.

En algunas implementaciones, un método para determinar la ploidía fetal es según las lecturas de secuencia de ácidos nucleicos para una muestra de prueba obtenida de una mujer embarazada. En algunas implementaciones, las lecturas de secuencia son lecturas de ácido nucleico circulante libre de células de una muestra (por ejemplo, una muestra de prueba). En algunas implementaciones, un método para determinar la ploidía fetal comprende obtener recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia. En algunas implementaciones, las lecturas de secuencia se mapean en un subconjunto de porciones del genoma de referencia. En algunas implementaciones, la determinación de la ploidía fetal comprende determinar una fracción fetal. En algunas implementaciones, la determinación de la ploidía fetal comprende calcular o determinar los niveles de sección genómica o porción. En determinadas implementaciones, la determinación de la ploidía fetal comprende determinar una fracción fetal y calcular o determinar los niveles de porción o sección genómica. En algunas implementaciones, la fracción fetal y los niveles de sección genómica o porción se determinan a partir de la misma muestra de prueba (por ejemplo, la misma parte de la muestra de prueba). En algunas implementaciones, la fracción fetal y los niveles de sección genómica o porción se determinan a partir de las mismas lecturas obtenidas de la misma muestra de prueba (por ejemplo, la misma parte de la muestra de prueba). En algunas implementaciones, la fracción fetal y los niveles de sección genómica o porción se determinan a partir de las mismas lecturas obtenidas del mismo proceso de secuenciación y/o de la misma celda de flujo. En algunas implementaciones, la fracción fetal y los niveles de sección genómica o porción se determinan a partir del mismo equipo y/o máquina (por ejemplo, aparato de secuenciación, celda de flujo o similar).

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

Métodos de análisis de decisiones para determinar una aneuploidía cromosómica

En algunas implementaciones, se realiza una determinación de un resultado (por ejemplo, realización de una identificación) o una determinación de la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica según un análisis de decisiones. Por ejemplo, un análisis de decisiones a veces comprende aplicar uno o más métodos que producen uno o más resultados, una evaluación de los resultados y una serie de decisiones basadas en resultados, evaluaciones y/o las posibles consecuencias de las decisiones y terminando en alguna coyuntura del proceso donde se toma una decisión final. En algunas implementaciones, un análisis de decisiones es un árbol de decisiones. Un análisis de decisiones, en algunas implementaciones, comprende el uso coordinado de uno o más procesos (por ejemplo, etapas de proceso, por ejemplo, algoritmos). Un análisis de decisiones puede ser realizado por una persona, un sistema, una máquina, un *software* (por ejemplo, un módulo), un ordenador, un procesador (por ejemplo, un microprocesador), similares o una combinación de los mismos. En algunas implementaciones, un análisis de decisiones comprende un método para determinar la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica en un feto con determinaciones reducidas de

falsos negativos y falsos positivos. En algunas implementaciones, un análisis de decisiones comprende determinar la presencia o ausencia de una afección asociada con una o más microduplicaciones o microdeleciones. Por ejemplo, en algunas implementaciones, un análisis de decisiones comprende determinar la presencia o ausencia de variaciones genéticas asociadas con un síndrome de DiGeorge. En algunas implementaciones, un análisis de decisiones comprende determinar la presencia o ausencia de un síndrome de DiGeorge.

En algunas implementaciones, un análisis de decisiones comprende generar un perfil para un genoma o un segmento de un genoma (por ejemplo, un cromosoma o parte del mismo). Se puede generar un perfil mediante cualquier método adecuado, conocido o descrito en el presente documento, y a menudo incluye la obtención de recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, recuentos de normalización, niveles de normalización, relleno, similares o combinaciones de los mismos. La obtención de recuentos de lecturas de secuencia mapeadas en un genoma de referencia puede incluir la obtención de una muestra (por ejemplo, de un sujeto de sexo femenino gestante), la secuenciación de ácidos nucleicos de una muestra (por ejemplo, ácidos nucleicos circulantes libres de células), la obtención de lecturas de secuencia, el mapeo de lecturas de secuencia en porciones de un genoma de referencia, similares y combinaciones de los mismos. En algunas implementaciones, la generación de un perfil comprende la normalización de recuentos mapeados en porciones de un genoma de referencia, lo que proporciona niveles de sección genómica calculados.

En algunas implementaciones, un análisis de decisiones comprende la segmentación. En algunas implementaciones, la segmentación modifica y/o transforma un perfil proporcionando de este modo una o más representaciones de descomposición de un perfil. Una representación de descomposición de un perfil suele ser una transformación de un perfil. Una representación de descomposición de un perfil es a veces una transformación de un perfil en una representación de un genoma, cromosoma o segmento del mismo. En determinadas implementaciones, la segmentación localiza e identifica uno o más niveles en un perfil (por ejemplo, ondículas) que son diferentes (por ejemplo, sustancialmente o significativamente diferentes) a uno o más niveles dentro de un perfil. Un nivel identificado en un perfil según una segmentación, donde ambos bordes del nivel son diferentes de otro nivel del perfil, se denomina en el presente documento ondícula.

En algunas implementaciones, la segmentación localiza e identifica los bordes de las ondículas dentro de un perfil. En determinadas implementaciones, se identifican uno o ambos bordes de una o más ondículas. Por ejemplo, un proceso de segmentación puede identificar la ubicación (por ejemplo, las coordenadas genómicas, por ejemplo, la ubicación de la porción) de los bordes derecho y/o izquierdo de una ondícula en un perfil. Una ondícula a menudo comprende dos bordes. Por ejemplo, una ondícula puede comprender un borde izquierdo y un borde derecho. En algunas implementaciones, dependiendo de la representación o vista, un borde izquierdo puede ser un borde 5' y un borde derecho puede ser un borde 3' de un segmento o perfil de ácido nucleico. En algunas implementaciones, un borde izquierdo puede ser un borde 3' y un borde derecho puede ser un borde 5'. A menudo, los bordes de un perfil se conocen antes de la segmentación y, por lo tanto, en algunas implementaciones, los bordes de un perfil determinan qué borde de un nivel es un borde 5' y qué borde es un borde 3'. En algunas implementaciones, uno o ambos bordes de un perfil y/u ondícula es un borde de un cromosoma.

En algunas implementaciones, los bordes de una ondícula se determinan según una representación de descomposición generada para una muestra de referencia (por ejemplo, un perfil de referencia). En algunas implementaciones, se determina una distribución de las alturas de los bordes nula según una representación de descomposición de un perfil de referencia (por ejemplo, un perfil de un cromosoma o segmento del mismo). En determinadas implementaciones, los bordes de una ondícula en un perfil se identifican cuando el nivel de la ondícula está fuera de una distribución de las alturas de los bordes nula. En algunas implementaciones, los bordes de una ondícula de un perfil se identifican según una puntuación Z calculada según una representación de descomposición para un perfil de referencia.

A veces, la segmentación genera dos o más ondículas (por ejemplo, dos o más niveles fragmentados, dos o más segmentos fragmentados) en un perfil. En algunas implementaciones, una representación de descomposición derivada de una segmentación está sobresegmentada o fragmentada y comprende múltiples ondículas. A veces, las ondículas generadas por la segmentación son sustancialmente diferentes y, a veces, las ondículas generadas por la segmentación son sustancialmente similares. Las ondículas sustancialmente similares (por ejemplo, niveles sustancialmente similares) a menudo se refieren a dos o más ondículas adyacentes de un perfil segmentado con un nivel de sección genómica (por ejemplo, un nivel) que difiere en menos de una medida de incertidumbre predeterminada. En algunas implementaciones, las ondículas sustancialmente similares son adyacentes entre sí y no están separadas por una ondícula intermedia. En algunas implementaciones, las ondículas sustancialmente similares están separadas por una o más ondículas más pequeñas. En algunas implementaciones, las ondículas sustancialmente similares están separadas por de aproximadamente 1 a aproximadamente 20, de aproximadamente 1 a aproximadamente 15, de aproximadamente 1 a aproximadamente 10 o de aproximadamente 1 a aproximadamente 5 porciones, donde una o más de las porciones tienen niveles significativamente diferentes al nivel de cada una de las ondículas sustancialmente similares. En algunas implementaciones, el nivel de pequeñas ondas sustancialmente similares difiere en menos de aproximadamente 3 veces, menos de aproximadamente 2 veces, menos de aproximadamente 1 vez o menos de aproximadamente 0,5 veces una medida de incertidumbre. Las ondículas sustancialmente similares, en algunas implementaciones, comprenden una mediana del nivel de sección genómica que difiere en menos de 3 DAM (por ejemplo, menos de 3 sigma), menos de 2 DAM, menos de 1 DAM o menos de aproximadamente 0,5 DAM, donde una DAM se calcula a partir de la mediana del nivel de sección genómica de cada una de las ondículas. Las ondículas sustancialmente diferentes, en algunas implementaciones no son adyacentes o están separadas por 10 o más, 15 o más o 20 o más porciones. En algunas implementaciones, las ondículas sustancialmente diferentes comprenden niveles sustancialmente diferentes. En determinadas implementaciones, las ondículas

sustancialmente diferentes comprenden niveles que difieren en más de aproximadamente 2,5 veces, más de aproximadamente 3 veces, más de aproximadamente 4 veces, más de aproximadamente 5 veces, más de aproximadamente 6 veces una medida de incertidumbre. Las ondículas sustancialmente diferentes, en algunas implementaciones, comprenden una mediana del nivel de sección genómica que difiere en más de 2,5 DAM (por ejemplo, más de 2,5 sigma), más de 3 DAM, más de 4 DAM, más de aproximadamente 5 DAM o más de aproximadamente 6 DAM, donde se calcula una DAM a partir de la mediana del nivel de sección genómica de cada una de las ondículas.

En algunas implementaciones un proceso de segmentación comprende determinar (por ejemplo, calcular) un nivel (por ejemplo, un valor cuantitativo, por ejemplo, un nivel medio o en mediana), una medida de incertidumbre (por ejemplo, una medida de incertidumbre), puntuación Z, valor de Z, valor de *p*, similares o combinaciones de los mismos para una o más ondículas (por ejemplo, niveles) en un perfil o segmento del mismo. En algunas implementaciones, se determinan (por ejemplo, se calculan) un nivel (por ejemplo, un valor cuantitativo, por ejemplo, un nivel medio o en mediana), una medida de incertidumbre (por ejemplo, una medida de incertidumbre), una puntuación Z, un valor de Z, un valor de *p*, similares o combinaciones de los mismos para una ondícula.

En algunas implementaciones, la segmentación comprende uno o más subprocesos, cuyos ejemplos no limitativos incluyen un proceso de generación de descomposición de ondículas, umbralización, nivelación, suavizado, similares o una combinación de los mismos.

En algunas implementaciones, la segmentación se realiza según un proceso de generación de descomposición de ondículas. En algunas implementaciones, la segmentación se realiza según dos o más procesos de generación de descomposición de ondículas. En algunas implementaciones, un proceso de generación de descomposición de ondículas identifica una o más ondículas en un perfil. En algunas implementaciones, un proceso de generación de descomposición de ondículas proporciona una representación de descomposición de un perfil. La segmentación se puede realizar, en su totalidad o en parte, mediante cualquier proceso de generación de descomposición de ondículas adecuado descrito en el presente documento o conocido en la técnica. Los ejemplos no limitativos de un proceso de generación de descomposición de ondículas incluyen una segmentación de ondículas de Haar (Haar, Alfred (1910) "Zur Theorie der orthogonalen Funktionensysteme", Mathematische Annalen 69 (3): 331-371; Nason, G. P. (2008) "Wavelet methods in Statistics", R. Springer, Nueva York.)(por ejemplo, WaveThresh), Wavethresh, un proceso de segmentación binaria recursiva adecuado, la segmentación binaria circular (CBS)(Olshen, A. B., Venkatraman, E. S., Lucito, R., Wigler, M. (2004) "Circular binary segmentation for the analysis of array-based DNA copy number data", *Biostatistics*, 5, 4:557-72; Venkatraman, E. S., Olshen, A. B. (2007) "A faster circular binary segmentation algorithm for the analysis of array CGH data", *Bioinformatics*, 23, 6:657-63), transformada de ondículas difrenciadas de solapamiento máximo (MODWT)(L. Hsu, S. Self, D. Grove, T. Randolph, K. Wang, J. Delrow, L. Loo, y P. Porter, "Denoising array-based comparative genomic hybridization data using wavelets", *Biostatistics* (Oxford, Inglaterra), vol. 6, n.° 2, págs. 211-226, 2005), ondícula estacionaria (SWT) (Y. Wang y S. Wang, "A novel stationary wavelet denoising algorithm for array-based DNA copy number data", *International Journal of Bioinformatics Research and Applications*, vol. 3, n.° 2, págs. 206-222, 2007), transformada de ondículas compleja de doble árbol (DTCWT)(Nha, N., H. Heng, S. Oraintara y W. Yuhang (2007) "Denoising of Array-Based DNA Copy Number Data Using The Dual-tree Complex Wavelet Transform". 137-144), segmentación de máxima entropía, convolución con núcleo de detección de bordes, divergencia de Jensen Shannon, divergencia de Kullback-Leibler, segmentación recursiva binaria, una transformada de Fourier, similares o combinaciones de los mismos.

Un proceso de generación de descomposición de ondículas puede representarse mediante un *software*, módulo y/o código adecuado escrito en un lenguaje adecuado (por ejemplo, un lenguaje de programación informático conocido en la técnica) y/o sistema operativo, cuyos ejemplos no limitativos incluyen UNIX, Linux, Oracle, Windows, Ubuntu, ActionScript, C, C++, C#, Haskell, Java, JavaScript, Objective-C, Perl, Python, Ruby, Smalltalk, SQL, Visual Basic, COBOL, Fortran, UML, HTML (por ejemplo, con PHP), PGP, G, R, S, similares o combinaciones de los mismos. En algunas implementaciones, un proceso de generación de descomposición de ondículas adecuado se representa en código S o R o mediante un paquete (por ejemplo, un paquete R). R, el código fuente de R, los programas de R, los paquetes de R y la documentación de R para los procesos de generación de descomposición de ondículas están disponibles para su descarga desde un CRAN o un sitio espejo de CRAN (The Comprehensive R Archive Network (CRAN) [en línea], [recuperado el 24-04-2013], obtenido de Internet <URL:*>http://cran.us.r-project.org/<>). CRAN es una red de servidores ftp y web en todo el mundo que almacenan versiones idénticas y actualizadas de código y documentación para R. Por ejemplo, WaveThresh (WaveThresh: Estadísticas y transformaciones de ondículas [en línea], [obtenido el 24-04-2013], obtenido de Internet <URL:*>http://cran.r-project.org/web/packages/wavethresh/index.html<>) y una descripción detallada de WaveThresh (Paquete "wavethresh"' [en línea, PDF], 2 de abril de 2013, [obtenido el 24-04-2013], obtenido de Internet <URL:*>http://cran.r-project.org/web/packages/wavethresh/wavethresh.pdf<>) están disponibles para su descarga. En algunas implementaciones, el código R para un proceso de generación de descomposición de ondículas (por ejemplo, segmentación de máxima entropía). Se puede descargar fácilmente un ejemplo de código R para un método CBS (por ejemplo, DNAcopy [en línea], [obtenido el 24-4-2013], obtenido de Internet <URL:* >http://bioconductor.org/packages/2.12/bioc/html/DNAcopy.html<> o el paquete informático 'DNAcopy' [en línea, PDF], 24 de abril de 2013, [obtenido el 24-04-2013], obtenido de Internet <URL:*>http://www.bioconductor.org/packages/release/bioc/manuals/DNAcopy/man/DNAcopy. pdf <>).

En algunas implementaciones, un proceso de generación de la descomposición de ondículas (por ejemplo, una segmentación de ondículas de Haar, por ejemplo, WaveThresh) comprende la umbralización. En algunas implementaciones, la umbralización distingue las señales del ruido. En determinadas implementaciones, la umbralización determina qué coeficientes de ondícula (por ejemplo, nodos) son indicativos de señales y deben conservarse y qué coeficientes de ondícula son indicativos de un reflejo de ruido y deben eliminarse. En algunas implementaciones, la umbralización comprende uno o más parámetros variables donde un usuario establece el valor del parámetro. En algunas implementaciones, los parámetros de la umbralización (por ejemplo, un parámetro de la umbralización, un parámetro de política) pueden describir o definir la cantidad de segmentación utilizada en un proceso de generación de la descomposición de ondículas. Se puede utilizar cualquier valor de parámetro adecuado. En algunas implementaciones, se utiliza un parámetro de umbralización. En algunas implementaciones, un valor de parámetro de umbralización es una umbralización suave. En determinadas implementaciones, se utiliza una umbralización suave para eliminar coeficientes pequeños y no significativos. En determinadas implementaciones, se utiliza una umbralización estricta. En determinadas implementaciones, una umbralización comprende un parámetro de política. Se puede utilizar cualquier valor de política adecuado. En algunas implementaciones, la política utilizada es "universal" y en algunas implementaciones, la política utilizada es "segura".

En algunas implementaciones, un proceso de generación de la descomposición de ondículas (por ejemplo, una segmentación de ondículas de Haar, por ejemplo, WaveThresh) comprende la nivelación. En algunas implementaciones, después de la umbralización, quedan algunos coeficientes de alto nivel. Estos coeficientes representan cambios bruscos o grandes picos en la señal original y, en determinadas implementaciones, se eliminan mediante la nivelación. En algunas implementaciones, la nivelación incluye la asignación de un valor a un parámetro conocido como nivel de descomposición (c). En determinadas implementaciones, se determina un nivel de descomposición óptimo según uno o más valores determinados, como la longitud del cromosoma (por ejemplo, la longitud del perfil), la longitud de ondícula deseada para la detección, la fracción fetal, la cobertura de la secuencia (por ejemplo, el nivel de plexación) y el nivel de ruido de un perfil normalizado. Para una longitud dada de un segmento de un genoma, cromosoma o perfil (N_{cr}), el nivel de descomposición de ondículas c a veces se relaciona con la longitud de la ondícula mínima N_{micro} según la ecuación $N_{micro} = N_{cr} / 2^{c+1}$. En algunas implementaciones, para detectar una microdeleción de tamaño N_{micro} o mayor, el nivel de descomposición deseado c se determina según la siguiente ecuación: $c = \log2 (N_{cr}/ N_{micro})-1$. Por ejemplo, si $N_{cr} = 4096$ porciones de un genoma de referencia y $N_{micro} = 128$ porciones de un genoma de referencia, entonces el nivel de descomposición es de aproximadamente 3 a aproximadamente 5. En algunas implementaciones, un nivel de descomposición (c) es aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10. En algunas implementaciones, la longitud de la ondícula mínima deseada para detectar, N_{micro} es de aproximadamente 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 10 Mb, 15 Mb o más de aproximadamente 20 Mb. En algunas implementaciones, N_{micro} está predeterminado. En algunas implementaciones, la cantidad de cobertura de secuencia (por ejemplo, el nivel de plexación) y la fracción fetal son inversamente proporcionales a N_{micro}. Por ejemplo, la longitud de ondícula mínima deseada para la detección disminuye (es decir, aumenta la resolución) a medida que aumenta la cantidad de fracción fetal en una muestra. En algunas implementaciones, la longitud de ondícula mínima deseada para la detección disminuye (es decir, aumenta la resolución) a medida que aumenta la cobertura (es decir, disminuye el nivel plexación). Por ejemplo, para una muestra que comprende aproximadamente el 10 % de fracción fetal, una 4-plexación produce un N_{micro} de aproximadamente 1 Mb o más y un nivel de 12 veces produce un N_{micro} de aproximadamente 3 Mb o más. En algunas implementaciones, la umbralización se realiza antes de la nivelación y, a veces, la umbralización se realiza después de la nivelación.

Proceso de segmentación de máxima entropía

En algunas implementaciones, un proceso de generación de la descomposición de ondículas adecuado comprende una segmentación de máxima entropía. En algunas implementaciones, una segmentación de máxima entropía comprende determinar una representación de descomposición. En algunas implementaciones, una segmentación de máxima entropía comprende determinar la presencia o ausencia de anomalías subcromosómicas (por ejemplo, una microduplicación, una microdeleción).

En determinadas implementaciones, una segmentación de máxima entropía comprende la partición recursiva de un segmento de un genoma (por ejemplo, un conjunto de porciones, un perfil). En determinadas implementaciones, una segmentación de máxima entropía divide un segmento de un genoma según niveles (por ejemplo, niveles de sección genómica). En determinadas implementaciones, una segmentación de máxima entropía comprende determinar un nivel para las partes segmentadas de un perfil. En algunas implementaciones, una segmentación de máxima entropía divide un segmento de un genoma en dos segmentos (por ejemplo, dos conjuntos de porciones) y calcula un nivel para los dos segmentos. En algunas implementaciones, el nivel de los dos segmentos se calcula antes o después de realizar una división (por ejemplo, una segmentación). En algunas implementaciones, se selecciona un sitio de partición (por ejemplo, la ubicación de la segmentación, la ubicación de la división) para maximizar la diferencia entre el nivel de los dos segmentos resultantes. En algunas implementaciones, una segmentación de máxima entropía determina la diferencia de nivel entre dos segmentos hipotéticos que resultaría de una segmentación hipotética para cada sitio de partición posible en un perfil (por ejemplo, segmento), selecciona el sitio donde se predice la diferencia máxima de nivel y luego divide (por ejemplo, particiona) el perfil en dos segmentos. En algunas implementaciones, dos segmentos adyacentes que se dividieron recientemente se determinan como significativamente diferentes o no significativamente diferentes mediante un método estadístico adecuado, cuyos ejemplos no limitativos incluyen una prueba de la *t*, un criterio basado en la *t* o similares. En algunas implementaciones, una segmentación de máxima entropía comprende particionar un primer y un segundo subconjunto de porciones cuando el

nivel del primer subconjunto de porciones es significativamente diferente del nivel del segundo subconjunto de porciones. En algunas implementaciones, el primer y el segundo subconjunto de porciones son adyacentes entre sí.

En algunas implementaciones, dos segmentos adyacentes que se dividieron recientemente se determinan como significativamente diferentes y cada uno de los segmentos se particiona nuevamente según la segmentación de máxima entropía (por ejemplo, según un sitio de partición que da como resultado una diferencia máxima de nivel). En algunas implementaciones, una segmentación de máxima entropía comprende particionar un conjunto de porciones (por ejemplo, un perfil) de forma recursiva, proporcionando de este modo dos o más subconjuntos de porciones donde cada uno de los subconjuntos resultantes comprende niveles que son significativamente diferentes al nivel de un subconjunto de porciones adyacente. En algunas implementaciones, una segmentación de máxima entropía comprende identificar una o más ondículas. En algunas implementaciones, una segmentación de máxima entropía comprende identificar un primer nivel significativamente diferente de un segundo nivel. Una ondícula es a menudo un primer nivel que es significativamente diferente de un segundo nivel (por ejemplo, un nivel de referencia). En determinadas implementaciones, una ondícula se determina según un nivel de referencia (por ejemplo, un nivel nulo, un perfil nulo). En algunas implementaciones, un nivel de referencia es un nivel de un perfil completo o de una parte del mismo. En algunas implementaciones, un nivel de referencia es un perfil de referencia (por ejemplo, o segmento) que se conoce como euploide o que carece de una variación del número de copias (por ejemplo, una microduplicación o una microdeleción). En algunas implementaciones, una ondícula es un primer nivel (por ejemplo, una ondícula) significativamente diferente de un segundo nivel (por ejemplo, un nivel de referencia) y el segundo nivel es un nivel de referencia. En algunas implementaciones, una segmentación de máxima entropía comprende determinar la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica en un feto para una muestra con determinaciones reducidas de falsos negativos y falsos positivos según una ondícula identificada y/o según un primer nivel significativamente diferente a un segundo nivel.

En algunas implementaciones, una segmentación de máxima entropía comprende volver a unir dos subconjuntos de porciones que fueron segmentadas (por ejemplo, divididas). En algunas implementaciones, dos segmentos que se dividieron no son significativamente diferentes y los dos segmentos se vuelven a unir. En algunas implementaciones, el nivel de cada uno de los dos subconjuntos de porciones que se segmentaron no son significativamente diferentes (por ejemplo, según un umbral predefinido, por ejemplo, una puntuación Z y/o una medida de incertidumbre, por ejemplo, una DAM) y los subconjuntos se vuelven a unir. En algunas implementaciones, los segmentos vueltos a unir no se vuelven a particionar.

En algunas implementaciones, un análisis de decisiones comprende dos o más procesos de segmentación que dan lugar a dos o más representaciones de descomposición. En determinadas implementaciones, un análisis de decisiones comprende dos o más procesos de segmentación diferentes (por ejemplo, procesos de generación de descomposición de ondículas) que generan de forma independiente diferentes representaciones de descomposición. En determinadas implementaciones, un análisis de decisiones comprende dos o más procesos de segmentación diferentes que generan de forma independiente representaciones de descomposición que son sustancialmente iguales (por ejemplo, sustancialmente similares). En algunas implementaciones, un análisis de decisiones comprende un primer proceso de segmentación y un segundo proceso de segmentación, y el primer y el segundo proceso de segmentación se realizan en paralelo. En determinadas implementaciones, se realizan en serie un primer y un segundo proceso de segmentación. En algunas implementaciones, cada uno de dos o más procesos de segmentación comprende un proceso de generación de descomposición de ondículas diferente. Por ejemplo, en algunas implementaciones, un primer proceso de segmentación comprende un proceso Haar Wavelet y un segundo proceso de segmentación comprende un proceso de segmentación binaria circular. En algunas implementaciones, cada uno de los dos o más procesos de segmentación son diferentes y comprenden los mismos procesos de generación de descomposición de ondículas. En determinadas implementaciones, dos procesos diferentes de generación de descomposición de ondículas generan independientemente dos representaciones de descomposición diferentes. En determinadas implementaciones, dos procesos diferentes de generación de descomposición de ondículas generan de forma independiente dos representaciones de descomposición que son sustancialmente iguales y/o comprenden una ondícula que es sustancialmente la misma. En algunas implementaciones, un primer proceso de segmentación comprende un primer proceso de generación de descomposición de ondículas y un segundo proceso de segmentación comprende un segundo proceso de generación de descomposición de ondículas, y los procesos de generación de descomposición de ondículas primero y segundo se aplican en paralelo. En algunas implementaciones, se realiza en serie un primer y un segundo proceso de generación de descomposición de ondículas.

Pulido

En algunas implementaciones, se pule una representación de descomposición, proporcionándose una representación de descomposición pulida. En algunas implementaciones, una representación de descomposición se pule dos o más veces. En algunas implementaciones, una representación de descomposición se pule antes y/o después de una o más etapas de un proceso de segmentación. En algunas implementaciones, un análisis de decisiones comprende dos o más procesos de segmentación y cada proceso de segmentación comprende uno o más procesos de pulido. Una representación de descomposición puede referirse a una representación de descomposición pulida o una representación de descomposición que no está pulida.

En algunas implementaciones, un proceso de segmentación comprende el pulido. En algunas implementaciones, un proceso de pulido identifica dos o más ondículas sustancialmente similares (por ejemplo, en una representación de descomposición) y las fusiona en una sola ondícula. En algunas implementaciones, un proceso de pulido identifica dos o más ondículas adyacentes que son sustancialmente similares y las fusiona en un solo nivel u ondícula. En algunas implementaciones, un proceso de pulido comprende un proceso de fusión. En determinadas implementaciones, las ondículas fragmentadas adyacentes se fusionan según sus niveles de sección genómica. En algunas implementaciones, la fusión de dos o más ondículas adyacentes comprende el cálculo de un nivel medio para las dos o más ondículas adyacentes que se fusionan. En algunas implementaciones, dos o más ondículas adyacentes que son sustancialmente similares se fusionan o pulen lugar a una sola ondícula o nivel. En determinadas implementaciones, dos o más ondículas adyacentes se fusionan mediante un proceso descrito por Willenbrock y Fridly (Willenbrock H, Fridlyand J. "A comparison study: applying segmentation to array CGH data for downstream analyses". *Bioinformatics* (2005) 15 de noviembre; 21(22):4084-91). En algunas implementaciones, dos o más ondículas adyacentes se fusionan mediante un proceso conocido como GLAD y descrito en Hupe, P. *y col.* (2004) "Analysis of array CGH data: from signal ratio to gain and loss of DNA regions", *Bioinformatics*, 20, 3413-3422.

Identificación de un evento de ondícula

En algunas implementaciones, un análisis de decisiones comprende la identificación de un evento de ondícula en una representación de descomposición. Un evento de ondícula es la ondícula más significativa identificada en una representación de descomposición (por ejemplo, un perfil). Un evento de ondícula suele ser la ondícula más grande de un perfil definido como el que tiene el mayor número de porciones en comparación con otras ondículas del perfil. Un evento de ondícula suele ser más grande y, a veces, sustancialmente superior a otras ondículas en una representación de descomposición. En algunas implementaciones, solo se identifica un evento de ondícula en una representación de descomposición. En algunas implementaciones, una o más ondículas se identifican en una representación de descomposición y una de las una o más ondículas se identifica como un evento de ondícula. En algunas implementaciones, un evento de ondícula es una primera ondícula (por ejemplo, un nivel) sustancialmente más grande que una segunda ondícula (por ejemplo, un segundo nivel), donde la primera ondícula es el nivel más grande en una representación de descomposición. Un evento de ondícula puede identificarse mediante un método adecuado. Una primera ondícula (por ejemplo, un evento de ondícula) que es sustancialmente más grande que otra ondícula a menudo comprende la mayor cantidad de porciones de un genoma de referencia y/o pares de bases en comparación con otras ondículas de una representación de descomposición. En algunas implementaciones, un evento de ondícula se identifica mediante un análisis del área bajo la curva (AUC). En algunas implementaciones, un análisis de decisiones comprende un análisis AUC. En determinadas implementaciones, una primera ondícula que es sustancialmente más grande que otra ondícula comprende un AUC mayor. En determinadas implementaciones, un AUC se determina como un valor absoluto de un AUC calculado (por ejemplo, un valor positivo resultante). En determinadas implementaciones, una vez que se identifica un evento de ondícula (por ejemplo, mediante un análisis AUC o mediante un método adecuado) se selecciona para un cálculo de puntuación z, o similar, para determinar si el evento de ondícula representa una variación genética (por ejemplo, una aneuploidía, microdeleción o microduplicación).

Comparación

En algunas implementaciones, un análisis de decisiones comprende una comparación. En algunas implementaciones, una comparación comprende comparar al menos dos representaciones de descomposición. En algunas implementaciones, una comparación comprende comparar al menos dos eventos de ondícula. En determinadas implementaciones, cada uno de los al menos dos eventos de ondícula proviene de una representación de descomposición diferente. Por ejemplo, un primer evento de ondícula puede provenir de una primera representación de descomposición y un segundo evento de ondícula puede provenir de una segunda representación de descomposición. En algunas implementaciones, una comparación comprende determinar si dos representaciones de descomposición son sustancialmente iguales o diferentes. En algunas implementaciones, una comparación comprende determinar si dos eventos de ondícula son sustancialmente iguales o diferentes.

En algunas implementaciones, dos representaciones de descomposición son sustancialmente iguales cuando cada representación comprende un evento de ondícula y los eventos de ondícula de cada representación de descomposición se determinan como sustancialmente iguales. Dos eventos de ondícula se pueden determinar como sustancialmente iguales o diferentes mediante un método de comparación adecuado, cuyos ejemplos no limitativos incluyen examen visual, comparando niveles o puntuaciones Z de los dos eventos de ondícula, comparando los bordes de los dos eventos de ondícula, superponiendo los dos eventos de ondícula o sus correspondientes representaciones de descomposición, similares o combinaciones de los mismos. En algunas implementaciones, los bordes de dos eventos de ondícula son sustancialmente iguales y los dos eventos de ondícula son sustancialmente iguales. En determinadas implementaciones, el borde de un evento de ondícula es sustancialmente el mismo que el borde de otro evento de ondícula, y los dos bordes están separados por menos de 10, menos de 9, menos de 8, menos de 7, menos de 6, menos de 5, menos de 4, menos de 3, menos de 2 o menos de 1 porción. En algunas implementaciones, dos bordes son sustancialmente iguales y están en la misma ubicación (por ejemplo, la misma porción). En algunas implementaciones, dos eventos de ondícula que son sustancialmente iguales comprenden niveles, puntuaciones Z o similares que son sustancialmente iguales (por ejemplo, dentro de una medida de incertidumbre, por ejemplo, aproximadamente 3, 2, 1 o menos veces una medida de incertidumbre). En algunas implementaciones, dos eventos

de ondícula comprenden bordes sustancialmente diferentes y/o niveles sustancialmente diferentes, y se determina, según una comparación, que no son sustancialmente iguales (por ejemplo, diferentes).

En determinadas implementaciones, una comparación comprende generar un evento de ondícula compuesto. En algunas implementaciones, dos o más eventos de ondícula son sustancialmente iguales y se genera un evento de ondícula compuesto. Un evento de ondícula compuesto puede generarse mediante cualquier método adecuado. En algunas implementaciones, se genera un evento de ondícula compuesto promediando dos o más eventos de ondícula (por ejemplo, los niveles, AUC y/o bordes) que son sustancialmente iguales. En algunas implementaciones, se genera un evento de ondícula compuesto superponiendo dos o más eventos de ondícula que son sustancialmente iguales. En algunas implementaciones, dos eventos de ondícula son diferentes y no se genera un evento de ondícula compuesto.

En determinadas implementaciones, una comparación comprende determinar la presencia o ausencia de un evento de ondícula compuesto a partir de eventos de ondícula identificados en dos o más representaciones de descomposición. En algunas implementaciones, dos o más eventos de ondícula (por ejemplo, derivados de dos o más representaciones de descomposición) son sustancialmente iguales y se determina la presencia de un evento de ondícula compuesto. La presencia o ausencia de un evento de ondícula compuesto puede determinarse mediante cualquier método adecuado. En algunas implementaciones, la presencia o ausencia de un evento de ondícula compuesto se determina promediando dos o más eventos de ondícula (por ejemplo, los niveles, AUC y/o bordes). En algunas implementaciones, la presencia o ausencia de un evento de ondícula compuesto se determina superponiendo dos o más eventos de ondícula. En determinadas implementaciones, la presencia de un evento de ondícula compuesto se determina cuando dos o más eventos de ondícula son sustancialmente iguales.

En algunas implementaciones, dos o más eventos de ondícula (por ejemplo, derivados de dos o más representaciones de descomposición) son diferentes (por ejemplo, sustancialmente diferentes) y se determina la ausencia de un evento de ondícula compuesto. En algunas implementaciones, la ausencia de un evento de ondícula compuesto indica la ausencia de una aneuploidía, microduplicación o microdeleción cromosómica.

Métodos adicionales de un análisis de decisiones

En algunas implementaciones, un análisis de decisiones comprende determinar un resultado (por ejemplo, determinar la presencia o ausencia de una variación genética, por ejemplo, en un feto). En algunas implementaciones, un análisis de decisiones comprende un método para determinar la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica. En algunas implementaciones, un análisis de decisiones comprende un método para determinar la presencia o ausencia de una variación genética (por ejemplo, en un feto) con determinaciones reducidas de falsos negativos y falsos positivos. En algunas implementaciones, un análisis de decisiones comprende una serie de métodos o etapas de métodos. En el presente documento, se describen ejemplos no limitativos de un análisis de decisión. En determinadas implementaciones, un análisis de decisiones comprende la obtención de recuentos y la generación y/u obtención de un perfil. En algunas implementaciones, un análisis de decisiones comprende cuantificar un perfil, o un segmento del mismo (por ejemplo, un segmento que representa un cromosoma). En algunas implementaciones de un análisis de decisiones, un perfil y/o un segmento del mismo (por ejemplo, un segmento que representa un cromosoma, un nivel, una ondícula, un evento de ondícula, una ondícula compuesta) se cuantifica mediante un método adecuado. Los ejemplos no limitativos de métodos de cuantificación adecuados se conocen en la técnica y se describen, en parte, en el presente documento e incluyen, por ejemplo, métodos para determinar una puntuación Z, valor de *p*, valor de t, nivel o nivel, AUC, ploidía, medida de incertidumbre, similares o combinaciones de los mismos.

En algunas implementaciones, un análisis de decisiones comprende segmentar un perfil mediante dos o más métodos. En algunas implementaciones, un análisis de decisiones comprende 50 o más métodos de segmentación. En determinadas implementaciones, un análisis de decisiones comprende 50 o menos, 40 o menos, 30 o menos, 20 o menos, 10 o menos, o aproximadamente 5 o menos métodos de segmentación. En determinadas implementaciones, un análisis de decisiones comprende aproximadamente 10, 9, 8, 7, 6, 5, 4, 3 o 2 métodos de segmentación. En algunas implementaciones, cada método de segmentación (por ejemplo, cuando se utilizan dos métodos) proporciona una representación de descomposición de un perfil. En algunas implementaciones, las representaciones de descomposición proporcionadas por dos o más métodos de segmentación son iguales, sustancialmente iguales o diferentes.

En algunas implementaciones, un pulido sigue a una segmentación. En algunas implementaciones, una o más representaciones de descomposición derivadas de una o más etapas de segmentación se pulen a veces mediante el mismo método de pulido. En algunas implementaciones, una o más representaciones de descomposición derivadas de una o más etapas de segmentación se pulen mediante un método de pulido diferente. En algunas implementaciones, una representación de descomposición se pule mediante uno, dos, tres o más métodos de pulido. En algunas implementaciones, cada representación de descomposición se pule mediante un método, y el método es el mismo para cada representación de descomposición.

En algunas implementaciones, la presencia o ausencia de un evento de ondícula se identifica siguiendo una segmentación o pulido. En algunas implementaciones, se omite una etapa de pulido, y se identifica un evento de ondícula directamente a partir de una representación de descomposición derivada de la segmentación. En algunas

implementaciones, un evento de ondícula se identifica en y/o a partir de una representación de descomposición pulida. En algunas implementaciones, no se identifica un evento de ondícula en una o más representaciones de descomposición y se determina la ausencia de una variación genética. En algunas implementaciones, cuando un evento de ondícula no se identifica en una o más representaciones de descomposición (por ejemplo, representaciones de descomposición pulidas), se termina un análisis de decisiones.

En algunas implementaciones, se cuantifica un evento de ondícula, una vez identificado. Un evento de ondícula se puede cuantificar mediante un método adecuado, cuyos ejemplos no limitativos incluyen el cálculo de una puntuación Z, el cálculo de un valor de *p*, la determinación de un valor t, la determinación de un nivel o nivel, la determinación de una ploidía, el cálculo de una medida de incertidumbre, similares o combinaciones de los mismos.

En algunas implementaciones, un análisis de decisiones comprende una comparación. En algunas implementaciones, una comparación sigue a una cuantificación. En algunas implementaciones, una comparación sigue a una identificación de ondícula. A veces, una comparación sigue a una determinación de perfil.

En algunas implementaciones, una comparación compara dos o más valores (por ejemplo, valores derivados de una cuantificación, por ejemplo, una cuantificación de un perfil y/o una cuantificación de un evento de ondícula). En algunas implementaciones, una comparación compara una cuantificación de un evento o perfil de ondícula con un valor o umbral predeterminado. En algunas implementaciones, una comparación comprende comparar puntuaciones Z. En determinadas implementaciones, una comparación comprende comparar una puntuación Z para un perfil de un cromosoma (es decir, $|Z_{cr}|$) a un valor o umbral predeterminado. El término $|Z_{cr}|$ representa el valor absoluto de una puntuación Z para un cromosoma. En algunas implementaciones, el umbral o valor predeterminado utilizado para la comparación de una puntuación Z es de aproximadamente 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,75, 3,8, 3,85, 3,9, 3,95, 4,0, 4,05, 4,1, 4,15, 4,2, 4,3, 4,4 o aproximadamente 4,5.

En algunas implementaciones, el resultado de una comparación es una decisión. En algunas implementaciones, el resultado de una comparación es un resultado. En algunas implementaciones, el resultado de una primera comparación es una decisión que determina la siguiente comparación en una serie de comparaciones. Por ejemplo, una primera comparación puede determinar que $|Z_{cr}|$ es superior o igual a un valor predeterminado y una segunda comparación compara $|Z_{cr}|$ con $|Z_{A4}|$ y/o $|Z_{B4}|$. Como alternativa, una primera comparación puede determinar que $|Z_{cr}|$ es inferior a un valor predeterminado y una segunda comparación determina si los eventos de ondícula identificados previamente en el análisis de decisiones son sustancialmente iguales o diferentes.

En algunas implementaciones, el resultado de una primera comparación es una decisión que determina una segunda comparación de una serie, y una decisión derivada de la segunda comparación determina una tercera comparación y así sucesivamente. En algunas implementaciones una primera comparación puede determinar que $|Z_{cr}|$ es superior o igual a un valor predeterminado y una segunda comparación puede determinar que $|Z_{cr}|$ es superior a $|Z_{A4}|$ y/o $|Z_{B4}|$ o una fracción del mismo (por ejemplo, $|Z_{A4}|$ y/o $|Z_{B4}|$ multiplicado por un valor predeterminado a) y se determina la presencia de una aneuploidía cromosómica completa. La trisomía y la monosomía pueden discernirse mediante un método adecuado.

Como alternativa, una primera comparación puede determinar que $|Z_{cr}|$ es superior o igual a un valor predeterminado y una segunda comparación puede determinar que $|Z_{cr}|$ es inferior a $|Z_{A4}|$ y/o $|Z_{B4}|$, o una fracción del mismo (por ejemplo, $|Z_{A4}|$ y/o $|Z_{B4}|$ se multiplica por un valor predeterminado a) y se realiza una tercera comparación. En determinadas implementaciones una primera comparación puede determinar que $|Z_{cr}|$ es inferior a un valor predeterminado, todos los eventos de ondícula identificados son sustancialmente iguales, una tercera comparación determina que $|Z_{A4}|$ y $|Z_{B4}|$ son superiores o iguales a un valor predeterminado (por ejemplo, 3,95) y se determina la presencia de una microduplicación y/o microdeleción. Una microduplicación y una microdeleción pueden discernirse mediante un método adecuado. Por ejemplo, una microduplicación puede tener una puntuación Z positiva y una microdeleción puede tener una puntuación Z negativa.

En algunas implementaciones, una comparación puede determinar que dos o más eventos de ondícula no son sustancialmente iguales (por ejemplo, sustancialmente diferentes) y que no existe variación genética en el perfil. En algunas implementaciones, una comparación puede determinar que dos o más eventos de ondículas (por ejemplo, todos los eventos de ondículas identificados en una o más representaciones de descomposición) son sustancialmente iguales y se determina la presencia o ausencia de una microduplicación o microdeleción. En algunas implementaciones, la presencia o ausencia de una microduplicación o microdeleción se determina según la cuantificación de un evento de ondícula compuesto.

En algunas implementaciones, un análisis de decisiones comprende dos o más de entre una segmentación, un pulido y la identificación de un evento de ondícula. En algunas implementaciones, un análisis de decisiones puede comprender una cuantificación de dos o más eventos de ondícula. En algunas implementaciones, un análisis de decisiones puede comprender la cuantificación de un perfil de un cromosoma. En algunas implementaciones, un análisis de decisiones comprende una o más comparaciones. En algunas implementaciones, un análisis de decisiones comprende una determinación de la presencia o ausencia de una variación genética.

En algunas implementaciones, un análisis de decisiones comprende y/o consiste en segmentar, pulir, identificar un evento de ondícula, comparar y determinar la presencia o ausencia de una variación genética. En algunas implementaciones, un análisis de decisiones comprende y/o consiste en una segmentación, un pulido, la identificación de un evento de

ondícula, una cuantificación, una comparación y una determinación de la presencia o ausencia de una variación genética. En algunas implementaciones, un análisis de decisiones comprende y/o consiste en una segmentación, un pulido, la identificación de un evento de ondícula, una comparación, una determinación de la presencia o ausencia de un evento de ondícula compuesto, una cuantificación de un evento de ondícula compuesto y una determinación de la presencia o ausencia de una variación genética. En algunas implementaciones, un análisis de decisiones comprende y/o consiste en una segmentación, un pulido, la identificación de un evento de ondícula, una cuantificación de un evento de ondícula, una cuantificación de un perfil de un cromosoma, una comparación y una determinación de la presencia o ausencia de una variación genética. En algunas implementaciones, un análisis de decisiones comprende una validación.

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 ese modo un resultado (por ejemplo, proporcionando de ese modo un resultado determinante de la presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal)). Una variación genética incluye a menudo 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 translocadas, secuencias de nucleótidos alteradas, similares o combinaciones de los anteriores) que dan como resultado un cambio detectable en el genoma o la información genética de un sujeto de prueba con respecto a una referencia. La presencia o ausencia de una variación genética puede determinarse transformando, analizando y/o manipulando lecturas de secuencias que hayan sido mapeadas en porciones (por ejemplo, recuentos, recuentos de porciones genómicas de un genoma de referencia). La determinación de un resultado, en algunas implementaciones, comprende analizar el ácido nucleico de una mujer embarazada. En determinadas implementaciones, se determina un resultado según los recuentos (por ejemplo, recuentos normalizados) obtenidos de una mujer embarazada donde los recuentos son de ácido nucleico obtenido de la mujer embarazada.

Los métodos descritos en el presente documento determinan, algunas veces, 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 prueba de una mujer embarazada que porta un feto. En determinadas implementaciones, los métodos descritos en el presente documento detectan euploidía o falta de euploidía (no euploidía) para una muestra de una mujer embarazada que porta un feto. Los métodos descritos en el presente documento detectan, algunas veces, trisomía para uno o más cromosomas (por ejemplo, el cromosoma 13, el cromosoma 18, el cromosoma 21 o combinación de los mismos) o segmento de los mismos.

En algunas implementaciones, 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 mediante una combinación de los mismos. La presencia o ausencia de una variación genética se determina generalmente a partir de recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia. Los recuentos de lecturas de secuencia usados para determinar la presencia o ausencia de una variación genética son, algunas veces, recuentos sin procesar y/o recuentos filtrados y a menudo recuentos normalizados. Pueden usarse uno o varios procedimientos de normalización adecuados para generar recuentos normalizados, los ejemplos no limitativos de los cuales incluyen normalización basada en porciones, normalización por contenido de GC, regresión lineal y no lineal por mínimos cuadrados, 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 de porciones en particular. Algunas veces, los recuentos normalizados se ajustan o rellenan antes de determinar la presencia o ausencia de una variación genética.

En algunas implementaciones, un resultado se determina según uno o más niveles. En algunas implementaciones, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según uno o más niveles ajustados. En algunas implementaciones, la determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según un perfil que comprende de 1 a aproximadamente 10.000 niveles ajustados. A menudo, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según un perfil que comprende de aproximadamente 1 a aproximadamente 1000, de 1 a aproximadamente 900, de 1 a aproximadamente 800, de 1 a aproximadamente 700, de 1 a aproximadamente 600, de 1 a aproximadamente 500, de 1 a aproximadamente 400, de 1 a aproximadamente 300, de 1 a aproximadamente 200, de 1 a aproximadamente 100, de 1 a aproximadamente 50, de 1 a aproximadamente 25, de 1 a aproximadamente 20, de 1 a aproximadamente 15, de 1 a aproximadamente 10 o de 1 a aproximadamente 5 ajustes. En algunas implementaciones, la determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según un perfil que comprende aproximadamente 1 ajuste (por ejemplo, un nivel ajustado). En algunas implementaciones, se determina un resultado según 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 algunas implementaciones, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según un perfil donde algunos niveles de un perfil no se ajustan. En algunas implementaciones, una determinación de la presencia o ausencia de una variación genética (por ejemplo, una aneuploidía cromosómica) se determina según un perfil donde no se realizan ajustes.

En algunas implementaciones, un ajuste de un nivel (por ejemplo, un primer nivel) en un perfil reduce una determinación falsa o un resultado falso. En algunas implementaciones, un ajuste de un nivel (por ejemplo, un primer nivel) en un perfil reduce la frecuencia y/o probabilidad (por ejemplo, probabilidad estadística, probabilidad) de una determinación falsa o un resultado falso. Una determinación o un resultado falso puede ser una determinación o un resultado que no sea exacto. Una determinación o un resultado falso puede ser una determinación o un resultado que no refleja 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 mujer embarazada, un feto y/o una combinación de los mismos). En algunas implementaciones, una determinación o un resultado falso es una determinación de falso negativo. En algunas implementaciones, una determinación negativa o un resultado negativo es la ausencia de una variación genética (por ejemplo, aneuploidía, variación del número de copias). En algunas implementaciones, una determinación falsa o un resultado falso es una determinación de falso positivo o un resultado de falso positivo. En algunas implementaciones, una determinación positiva o un resultado positivo es la presencia de una variación genética (por ejemplo, aneuploidía, variación del número de copias). En algunas implementaciones, se utiliza una determinación o un resultado en un diagnóstico. En algunas implementaciones, una determinación o un resultado es para un feto.

Algunas veces se determina la presencia o ausencia de una variación genética (por ejemplo, aneuploidía fetal) sin comparar recuentos de un conjunto de porciones con una referencia. Los recuentos medidos para una muestra de prueba y que se encuentran en una región de prueba (por ejemplo, un conjunto de porciones de interés) se denominan en el presente documento "recuentos de prueba". Los recuentos de prueba son, algunas veces, recuentos procesados, recuentos promediados o sumados, una representación, recuentos normalizados o uno o más niveles o niveles, tal como se describe en el presente documento. En determinadas implementaciones, los recuentos de prueba se promedian o suman (por ejemplo, se calcula un promedio, una media, una mediana, una moda o una suma) para un conjunto de porciones, y los recuentos promediados o sumados se comparan con un umbral o rango. Los recuentos de la prueba se expresan a veces como una representación, que puede expresarse como una relación o porcentaje de recuentos para un primer conjunto de porciones con respecto a los recuentos para un segundo conjunto de porciones. En determinadas implementaciones, el primer conjunto de porciones corresponde a uno o más cromosomas de prueba (por ejemplo, el cromosoma 13, el cromosoma 18, el cromosoma 21 o una combinación de los mismos) y, a veces, el segundo conjunto de porciones corresponde al genoma o a una parte del genoma (por ejemplo, los autosomas o los autosomas y los cromosomas sexuales). En determinadas implementaciones, una representación se compara con un umbral o rango. En determinadas implementaciones, los recuentos de prueba 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 rango. Los recuentos de prueba (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 rango particular o fuera de un rango particular, a veces determinan la presencia de una variación genética o falta de euploidía (por ejemplo, no euploidía). Los recuentos de prueba (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 rango particular o fuera de un rango 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 comparando recuentos, cuyos ejemplos no limitativos incluyen recuentos de prueba, recuentos de referencia, recuentos sin procesar, recuentos filtrados, recuentos promediados o sumados, representaciones (por ejemplo, representaciones cromosómicas), recuentos normalizados, uno o más niveles o niveles (por ejemplo, para un conjunto de porciones, por ejemplo, niveles de sección genómica, perfiles), puntuaciones Z, similares o combinaciones de los mismos. En algunas implementaciones, los recuentos de prueba 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 limitativos incluyen recuentos sin procesar, recuentos filtrados, recuentos promediados o sumados, representaciones (por ejemplo, representaciones cromosómicas), recuentos normalizados, uno o más niveles o niveles (por ejemplo, para un conjunto de porciones, por ejemplo, niveles de sección genómica, perfiles), puntuaciones Z, similares o combinaciones de los mismos. Los recuentos de referencia a menudo son recuentos de una región de prueba euploide o de un segmento de un genoma o cromosoma que es euploide. En algunas implementaciones, los recuentos de referencia y los recuentos de prueba se obtienen de la misma muestra y/o del mismo sujeto. En algunas implementaciones, los recuentos de referencia son de diferentes muestras y/o de diferentes sujetos. En algunas implementaciones, los recuentos de referencia se determinan a partir de y/o se comparan con un segmento correspondiente del genoma del que se derivan y/o determinan los recuentos de prueba. Un segmento correspondiente se refiere a un segmento, porción o conjunto de porciones que se mapean la misma ubicación de un genoma de referencia. En algunas implementaciones, los recuentos de referencia se determinan a partir de y/o se comparan con un segmento diferente del genoma del que se derivan y/o determinan los recuentos de prueba.

En determinadas implementaciones, los recuentos de prueba a veces son para un primer conjunto de porciones, y una referencia incluye recuentos para un segundo conjunto de porciones diferentes del primer conjunto de porciones. Algunas veces, los recuentos de referencia son para una muestra de ácido nucleico de la misma mujer embarazada de la cual se obtiene la muestra de prueba. En determinadas implementaciones, los recuentos de referencia son para una muestra de ácido nucleico de una o más mujeres embarazadas diferentes a las mujeres de las que se obtuvo la muestra de prueba. En algunas implementaciones, un primer conjunto de porciones se encuentra en el cromosoma 13, el cromosoma 18, el cromosoma 21, un segmento del mismo o una combinación de los anteriores, y el segundo conjunto de porciones se encuentra en otro cromosoma o cromosomas o segmento del mismo. En un ejemplo no

limitativo, cuando un primer conjunto de porciones se encuentra en el cromosoma 21 o en un segmento del mismo, un segundo conjunto de porciones suele encontrarse en otro cromosoma (por ejemplo, el cromosoma 1, el cromosoma 13, el cromosoma 14, el cromosoma 18, el cromosoma 19, un segmento del mismo o una combinación de los anteriores). Una referencia está ubicada a menudo en un cromosoma o segmento del mismo que es normalmente euploide. Por ejemplo, el cromosoma 1 y el cromosoma 19 son a menudo euploides en fetos debido a una alta tasa de mortalidad fetal precoz asociada con aneuploidías del cromosoma 1 y del cromosoma 19. Puede generarse una medida de desviación entre los recuentos de prueba y los recuentos de referencia.

En determinadas implementaciones, una referencia comprende recuentos para el mismo conjunto de porciones que para los recuentos de prueba, donde los recuentos para la referencia son de una o más muestras de referencia (por ejemplo, a menudo múltiples muestras de referencia de múltiples sujetos de referencia). Una muestra de referencia es a menudo de una o más mujeres embarazadas diferentes a una mujer de la cual se obtiene una muestra de prueba. Puede generarse una medida de la desviación (por ejemplo, una medida de la incertidumbre, una medida de la incertidumbre) entre los recuentos de prueba y los recuentos de referencia. En algunas implementaciones, se determina una medida de desviación a partir de los recuentos de prueba. En algunas implementaciones, se determina una medida de desviación a partir de los recuentos de referencia. En algunas implementaciones, se determina una medida de desviación a partir de un perfil completo o un subconjunto de porciones dentro de un perfil.

Puede seleccionarse una medida de desviación adecuada, cuyos ejemplos no limitativos incluyen desviación estándar, desviación absoluta promedio, mediana de desviación absoluta, desviación absoluta máxima, puntuación estándar (por ejemplo, valor de z, puntuación z, puntuación normal, variable normalizada) y similares. En algunas implementaciones, las muestras de referencia son euploides para una región de prueba y se evalúa la desviación entre los recuentos de prueba y los recuentos de referencia. En algunas implementaciones, la determinación de la presencia o ausencia de una variación genética es según el número de desviaciones (por ejemplo, medidas de desviaciones, DAM) entre los recuentos de prueba y los recuentos de referencia para un segmento o porción de un genoma o cromosoma. En algunas implementaciones, la presencia de una variación genética se determina cuando el número de desviaciones entre los recuentos de prueba y los recuentos de referencia es superior a aproximadamente 1, superior a aproximadamente 1,5, superior a aproximadamente 2, superior a aproximadamente 2,5, superior a aproximadamente 2 6, superior a aproximadamente 2,7, superior a aproximadamente 2,8, superior a aproximadamente 2,9, superior a aproximadamente 3, superior a aproximadamente 3,1, superior a aproximadamente 3,2, superior a aproximadamente 3,3, superior a aproximadamente 3,4, superior a aproximadamente 3,5, superior a aproximadamente 4, superior a aproximadamente 5 o superior a aproximadamente 6. Por ejemplo, a veces un recuento de prueba difiere de un recuento de referencia en más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 DAM) y se determina la presencia de una variación genética. En algunas implementaciones, un recuento de prueba obtenido de una mujer embarazada es superior a un recuento de referencia en más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 DAM) y se determina la presencia de una aneuploidía cromosómica fetal (por ejemplo, una trisomía fetal). A menudo, una desviación mayor de tres entre los recuentos de prueba y los recuentos de referencia es indicativa de una región de prueba no euploide (por ejemplo, presencia de una variación genética). Los recuentos de prueba significativamente por encima de los recuentos de referencia, recuentos de referencia que son indicativos de euploidía, algunas veces son determinantes de una trisomía. En algunas implementaciones, un recuento de prueba obtenido de una mujer embarazada es inferior a un recuento de referencia en más de 3 medidas de desviación (por ejemplo, 3 sigma, 3 DAM) y se determina la presencia de una aneuploidía cromosómica fetal (por ejemplo, una monosomía fetal). Los recuentos de prueba significativamente por debajo de los recuentos de referencia, recuentos de referencia que son indicativos de euploidía, algunas veces son determinantes de una monosomía.

En algunas implementaciones, la ausencia de una variación genética se determina cuando el número de desviaciones entre los recuentos de prueba y los recuentos de referencia es inferior a aproximadamente 3,5, inferior a aproximadamente 3,4, inferior a aproximadamente 3,3, inferior a aproximadamente 3,2, inferior a aproximadamente 3,1, inferior a aproximadamente 3,0, inferior a aproximadamente 2,9, inferior a aproximadamente 2,8, inferior a aproximadamente 2,7, inferior a aproximadamente 2,6, inferior a aproximadamente 2,5, inferior a aproximadamente 2,0, inferior a aproximadamente 1,5 o inferior a aproximadamente 1,0. Por ejemplo, a veces un recuento de prueba difiere de un recuento de referencia en menos de 3 medidas de desviación (por ejemplo, 3 sigma, 3 DAM) y se determina la ausencia de una variación genética. En algunas implementaciones, un recuento de prueba obtenido de una mujer embarazada difiere de un recuento de referencia en menos de 3 medidas de desviación (por ejemplo, 3 sigma, 3 DAM) y se determina la ausencia de aneuploidía cromosómica fetal (por ejemplo, euploidía fetal). En algunas implementaciones (por ejemplo, una desviación inferior a tres entre los recuentos de prueba y los recuentos de referencia (por ejemplo, 3-sigma para la desviación estándar) suele ser indicativa de una región de prueba euploide (por ejemplo, ausencia de una variación genética). Puede representarse gráficamente y visualizarse una medida de la desviación entre los recuentos de prueba de una muestra de prueba y los recuentos de referencia de uno o más sujetos de referencia (por ejemplo, una representación gráfica de la puntuación z).

Cualquier otra referencia adecuada puede factorizarse con recuentos de prueba para determinar la presencia o ausencia de una variación genética (o determinación de euploides o no euploides) para una región de prueba de una muestra de prueba. Por ejemplo, una determinación de fracción fetal puede factorizarse con recuentos de pruebas para determinar la presencia o ausencia de una variación genética. Puede usarse un procedimiento adecuado para cuantificar la fracción fetal, los ejemplos no limitativos de los mismos incluyen un procedimiento de espectrometría de masas, procedimiento de secuenciación o combinación de los mismos.

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

En algunas implementaciones, una ploidía fetal inferior a aproximadamente 1,35, inferior a aproximadamente 1,30, inferior a aproximadamente 1,25, inferior a aproximadamente 1,20 o inferior a aproximadamente 1,15 indica la ausencia de una aneuploidía fetal (por ejemplo, la ausencia de una trisomía fetal, por ejemplo, euploide). En algunas implementaciones, una determinación de la ploidía fetal de aproximadamente 0,7 a aproximadamente 1,35, de aproximadamente 0,7 a aproximadamente 1,30, de aproximadamente 0,7 a aproximadamente 1,25, de aproximadamente 0,7 a aproximadamente 1,20, de aproximadamente 0,7 a aproximadamente 1,15, de aproximadamente 0,75 a aproximadamente 1,35, de aproximadamente 0,75 a aproximadamente 1,30, de aproximadamente 0,75 a aproximadamente 1,25, de aproximadamente 0,75 a aproximadamente 1,20, de aproximadamente 0,75 a aproximadamente 1,15, de aproximadamente 0,8 a aproximadamente 1,35, de aproximadamente 0,8 a aproximadamente 1,30, de aproximadamente 0,8 a aproximadamente 1,25, de aproximadamente 0,8 a aproximadamente 1,20 o de 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 algunas implementaciones, una ploidía fetal inferior a aproximadamente 0,8, inferior a aproximadamente 0,75, inferior a aproximadamente 0,70 o inferior a aproximadamente 0,6 indica la presencia de una aneuploidía fetal (por ejemplo, la presencia de una deleción cromosómica). En algunas implementaciones, una determinación de una ploidía fetal de aproximadamente 0 a aproximadamente 0,8, de aproximadamente 0 a aproximadamente 0,75, de aproximadamente 0 a aproximadamente 0,70, de aproximadamente 0 a aproximadamente 0,65, de aproximadamente 0 a aproximadamente 0,60, de aproximadamente 0,1 a aproximadamente 0,8, de aproximadamente 0,1 a aproximadamente 0,75, de aproximadamente 0,1 a aproximadamente 0,70, de aproximadamente 0,1 a aproximadamente 0,65, de aproximadamente 0,1 a aproximadamente 0,60, de aproximadamente 0,2 a aproximadamente 0,8, de aproximadamente 0,2 a aproximadamente 0,75, de aproximadamente 0,2 a aproximadamente 0,70, de aproximadamente 0,2 a aproximadamente 0,65, de aproximadamente 0,2 a aproximadamente 0,60, de aproximadamente 0,25 a aproximadamente 0,8, de aproximadamente 0,25 a aproximadamente 0,75, de aproximadamente 0,25 a aproximadamente 0,70, de aproximadamente 0,25 a aproximadamente 0,65, de aproximadamente 0,25 a aproximadamente 0,60, de aproximadamente 0,3 a aproximadamente 0,8, de aproximadamente 0,3 a aproximadamente 0,75, de aproximadamente 0,3 a aproximadamente 0,70, de aproximadamente 0,3 a aproximadamente 0,65, de aproximadamente 0,3 a aproximadamente 0,60 indica la presencia de una aneuploidía cromosómica fetal (por ejemplo, la presencia de una deleción cromosómica). En algunas implementaciones, la aneuploidía fetal determinada es una deleción cromosómica completa.

En algunas implementaciones, la determinación de la presencia o ausencia de una aneuploidía fetal (por ejemplo, según uno o más de los rangos de determinación de la ploidía anteriores) se determina según una zona de identificación. En determinadas implementaciones, se realiza una identificación (por ejemplo, una identificación que determina 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, una medida de incertidumbre) o colección de valores cae dentro de un rango predefinido (por ejemplo, una zona, una zona de identificación). En algunas implementaciones, una zona de identificación se define según una colección de valores que se obtienen de la misma muestra de paciente. En determinadas implementaciones, una zona de identificación se define según una colección de valores que se derivan del mismo cromosoma o segmento del mismo. En algunas implementaciones, una zona de identificación basada en una determinación de ploidía se define según un nivel de confianza (por ejemplo, alto nivel de confianza, por ejemplo, baja medida de incertidumbre) y/o una fracción fetal. En algunas implementaciones, una zona de identificación se define según una determinación de ploidía y una fracción fetal de aproximadamente 2,0 % o superior, aproximadamente 2,5 % o superior, aproximadamente 3 % o superior, aproximadamente 3,25 % o superior, aproximadamente 3,5 % o superior, aproximadamente 3,75 % o superior, o aproximadamente 4,0 % o superior. Por ejemplo, en algunas implementaciones se hace una identificación de que un feto comprende una trisomía 21 basándose en una determinación de ploidía superior a 1,25 con una determinación de fracción fetal del 2 % o superior o del 4 % o superior para una muestra obtenida de una mujer embarazada portadora de un feto. En determinadas

implementaciones, por ejemplo, se hace una identificación de que un feto es euploide basándose en una determinación de ploidía inferior a 1,25 con una determinación de fracción fetal de 2 % o superior, o del 4 % o superior para una muestra obtenida de una mujer embarazada portadora de un feto. En algunas implementaciones, una zona de identificación se define por un nivel de confianza de aproximadamente el 99 % o superior, aproximadamente el 99,1 % o superior, aproximadamente el 99,2 % o superior, aproximadamente el 99,3 % o superior, aproximadamente el 99,4 % o superior, aproximadamente el 99,5 % o superior, aproximadamente el 99,6 % o superior, aproximadamente el 99,7 % o superior, aproximadamente el 99,8 % o superior, o aproximadamente el 99,9 % o superior. En algunas implementaciones, se realiza una identificación sin utilizar una zona de identificación. En algunas implementaciones, se realiza una identificación utilizando una zona de identificación y datos o información adicionales. En algunas implementaciones, se realiza una identificación en función de un valor de ploidía sin el uso de una zona de identificación. En algunas implementaciones, se realiza una identificación sin calcular un valor de ploidía. En algunas implementaciones, se realiza una identificación basada en el examen visual de un perfil (por ejemplo, examen visual de niveles de sección genómica o porción). Se puede realizar una identificación mediante cualquier método adecuado basado en todas o en parte en las determinaciones, valores y/o datos obtenidos mediante los métodos descritos en el presente documento, cuyos ejemplos no limitativos incluyen la determinación de la ploidía fetal, la determinación de la fracción fetal, la 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 categorizadas del número de copias), elevaciones o niveles significativamente diferentes, elevaciones o niveles ajustados (por ejemplo, relleno), similares o combinaciones de los mismos.

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

Algunas veces, una variación genética se asocia con una afección médica. Una determinación del resultado de una variación genética es, algunas veces, una determinación del resultado de la presencia o ausencia de una afección (por ejemplo, una afección médica), enfermedad, un síndrome o una anomalía, o incluye la detección de una afección, enfermedad, un síndrome o una anomalía (por ejemplo, los ejemplos no limitativos enumerados en la tabla 1). En determinadas implementaciones, 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, un síndrome o una anomalía mediante los métodos descritos en el presente documento a veces pueden verificarse independientemente mediante pruebas adicionales (por ejemplo, mediante cariotipado y/o amniocentesis). El análisis y procesamiento de los datos puede proporcionar uno o más resultados. El término ‘‘resultado’’, tal 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 del número de copias). En determinadas implementaciones, el término ‘‘resultado’’, tal 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 del número de copias). En determinadas implementaciones, el término ‘‘resultado’’, tal 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 del número de copias) en un sujeto (por ejemplo, un feto). Algunas veces, un diagnóstico comprende el uso de un resultado. Por ejemplo, un profesional sanitario puede analizar un resultado y proporcionar bases de diagnóstico en, o basándose en parte en, el resultado. En algunas implementaciones, la determinación, detección o el diagnóstico de una afección, un síndrome o una anomalía (por ejemplo, enumerado en la tabla 1) comprende el uso de un determinante del resultado de la presencia o ausencia de una variación genética. En algunas implementaciones, un resultado basado en lecturas de secuencia mapeadas contadas o transformaciones de las mismas es determinante de la presencia o ausencia de una variación genética. En algunas implementaciones, 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 determinadas implementaciones, un diagnóstico comprende una determinación de una presencia o ausencia de una afección, un síndrome o una anomalía. A menudo, un diagnóstico comprende una determinación de una variación genética como la naturaleza y/o la causa de una afección, un síndrome o una anomalía. En determinadas implementaciones, un resultado no es un diagnóstico. Un resultado comprende a menudo uno o más valores numéricos generados usando un método de procesamiento descrito en el presente documento en el contexto de una o más consideraciones de probabilidad. Una consideración de riesgo o probabilidad puede incluir, entre otros: una medida 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 Chi, valores Phi, valores de ploidía, fracción fetal ajustada, ratios de área, elevación o nivel medio, similares o combinaciones de los mismos. Una consideración de probabilidad puede facilitar la determinación de si un sujeto está en riesgo de tener, o tiene, una variación genética, y un determinante del resultado de una presencia o ausencia de un trastorno genético incluye a menudo tal consideración.

Un resultado algunas veces es un fenotipo. Un resultado a veces es un fenotipo con un nivel asociado de confianza (por ejemplo, una medida de incertidumbre, por ejemplo, un feto es positivo para la trisomía 21 con un nivel de confianza del 99 %, un sujeto de prueba es negativo para un cáncer asociado con una variación genética a un nivel de confianza del 95 %). Diferentes métodos para generar valores de resultado a veces pueden producir diferentes tipos de resultados. Generalmente, existen cuatro tipos de puntuaciones o identificaciones posibles que pueden realizarse basadas en los valores de resultados generados con el uso de los métodos descritos en el presente documento: verdadero positivo, falso positivo, verdadero negativo y falso negativo. 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 puede usarse para determinar, por ejemplo, una variación, diferencia o razón de lecturas de secuencia mapeadas que pueden corresponder a una variación genética. Por ejemplo, calcular una puntuación positiva para una variación genética o porción seleccionada a partir de un conjunto de datos, con respecto a un genoma de referencia puede conducir 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 algunas implementaciones, un resultado comprende una elevación o un nivel, un perfil y/o una representación gráfica (por ejemplo, una representación gráfica de perfiles). En aquellas implementaciones en las que un resultado comprende un perfil, puede usarse un perfil adecuado o combinación de perfiles para un resultado. Los ejemplos no limitativos 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 incluye, algunas veces, un resultado nulo (por ejemplo, un punto de datos entre dos agrupaciones, un valor numérico con una desviación estándar que abarca valores tanto para la presencia como para la ausencia de una variación genética, un conjunto de datos con un gráfico de perfiles que no es similar a los gráficos de perfiles para sujetos que tienen o están libres de la variación genética que se investiga). En algunas implementaciones, un resultado indicativo de un resultado nulo todavía es un resultado determinante, y la determinación puede incluir la necesidad de información adicional y/o una repetición de la generación y/o análisis de datos para determinar la presencia o ausencia de una variación genética.

Puede generarse un resultado después de realizar una o más etapas de procesamiento descritas en el presente documento, en algunas implementaciones. En determinadas implementaciones, se genera un resultado que resulta de una de las etapas de procesamiento descritas en el presente documento, y en algunas implementaciones, puede generarse un resultado después de realizar cada manipulación estadística y/o matemática de un conjunto de datos. Un resultado que corresponde a la determinación de la presencia o ausencia de una variación genética puede expresarse en una forma adecuada, forma que comprende, sin limitarse a, una probabilidad (por ejemplo, razón de probabilidades, valor de p), probabilidad, valor dentro o fuera de una agrupación, valor de por encima o por debajo de un valor umbral, valor dentro de un rango (por ejemplo, un rango 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 determinadas implementaciones, la comparación entre las 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, mal etiquetadas, combinadas, y similares)).

En algunas implementaciones, un resultado comprende un valor por encima o por debajo de un umbral predeterminado o valor de punto de corte (por ejemplo, mayor de 1, menor de 1), y un nivel de incertidumbre o confianza asociado con el valor. En determinadas implementaciones, un umbral predeterminado o valor de corte es una elevación o nivel esperado o un rango de elevación o nivel esperado. Un resultado puede describir además una suposición usada en el procesamiento de datos. En determinadas implementaciones, un resultado comprende un valor que se encuentra dentro o fuera de un rango predeterminado de valores (por ejemplo, un rango umbral) y el nivel de incertidumbre o de confianza asociado para ese valor que está dentro o fuera del rango. En algunas implementaciones, 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 rango de valores de predeterminado, y su nivel de incertidumbre o de confianza asociado para ese valor es igual o está dentro o fuera de un rango. Algunas veces, un resultado se representa gráficamente como una representación gráfica (por ejemplo, gráfico de perfiles).

Tal como se mencionó anteriormente, un resultado puede caracterizarse como un verdadero positivo, verdadero negativo, falso positivo o falso negativo. La expresión "verdadero positivo", tal como se usa en el presente documento, se refiere a un sujeto diagnosticado correctamente de una variación genética. La expresión "falso positivo", tal como se usa en el presente documento, se refiere a un sujeto mal identificado como que tiene una variación genética. La expresión "verdadero negativo", tal como se usa en el presente documento, se refiere a un sujeto identificado correctamente como que no tiene una variación genética. La expresión "falso negativo", tal como se usa en el presente documento, se refiere a un sujeto mal identificado como que no tiene una variación genética. Dos medidas de rendimiento para cualquier método dado pueden calcularse basándose en las razones de estas apariciones: (i) un valor de sensibilidad, que generalmente es la fracción de positivos previstos que se identifican correctamente como positivos; y (ii) un valor de especificidad, que generalmente es la fracción de negativos previstos correctamente identificados como negativos.

En determinadas implementaciones, uno o más de sensibilidad, especificidad y/o nivel de confianza se expresan como un porcentaje. En algunas implementaciones, el porcentaje, independientemente para cada variable, es mayor de aproximadamente el 90 % (por ejemplo, aproximadamente el 90, 91, 92, 93, 94, 95, 96, 97, 98 o el 99 %, o mayor del 99 % (por ejemplo, aproximadamente el 99,5 %, o mayor, aproximadamente el 99,9 % o mayor, aproximadamente el 99,95 % o mayor, aproximadamente el 99,99 % o mayor)). El coeficiente de variación (CV) en algunas implementaciones se expresa

como un porcentaje, y algunas veces el porcentaje es de aproximadamente el 10 % o menos (por ejemplo, aproximadamente el 10, 9, 8, 7, 6, 5, 4, 3, 2 o el 1 %, o menos del 1 % (por ejemplo, aproximadamente el 0,5 % o menos, aproximadamente el 0,1 % o menos, aproximadamente el 0,05 % o menos, aproximadamente el 0,01 % o menos)). Una probabilidad (por ejemplo, que un resultado particular no se deba al azar) en determinadas implementaciones se expresa como una puntuación Z, un valor de p o los resultados de una prueba de la t. En algunas implementaciones, una varianza medida, intervalo de confianza, sensibilidad, especificidad y similares (por ejemplo, denominados colectivamente parámetros de confianza) para un resultado pueden generarse usando una o más manipulaciones de procesamiento de datos descritas en el presente documento. Los ejemplos específicos de generación de resultados y los niveles de confianza asociados se describen en el apartado de ejemplos y en la solicitud internacional de patente n°. PCT/US12/59123 (WO2013/0522913).

El término "sensibilidad", tal como se usa en el presente documento, se refiere al número de verdaderos positivos dividido entre el número de verdaderos positivos más el número de falsos negativos, donde la sensibilidad (sens) puede estar dentro del rango de $0 \leq sens \leq 1$. El término "especificidad", tal como se usa en el presente documento, se refiere al número de verdaderos negativos dividido entre el número de verdaderos negativos más el número de falsos positivos, donde la sensibilidad (espec) puede estar dentro del rango de $0 \leq espec \leq 1$. En algunas implementaciones, a veces se selecciona un método que tenga una sensibilidad y una especificidad iguales a uno, o del 100 %, o cercanas a uno (por ejemplo, entre aproximadamente el 90 % y aproximadamente el 99 %). En algunas implementaciones, se selecciona un método que tiene una sensibilidad igual a 1 o el 100 % y, en determinadas implementaciones, se selecciona un método que tiene una sensibilidad próxima 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 algunas implementaciones se selecciona un método que tiene una especificidad equivalente a 1 o el 100 % y, en determinadas implementaciones, se selecciona un método que tiene una especificidad próxima 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 %).

En algunas implementaciones, se determina la presencia o ausencia de una variación genética (por ejemplo, aneuploidía cromosómica) en un feto. En dichas implementaciones, se determina la presencia o ausencia de una variación genética fetal (por ejemplo, aneuploidía cromosómica fetal).

En determinadas implementaciones, se determina la presencia o ausencia de una variación genética (por ejemplo, aneuploidía cromosómica) para una muestra. En dichas implementaciones, se determina la presencia o ausencia de una variación genética en el ácido nucleico de la muestra (por ejemplo, aneuploidía cromosómica). En algunas implementaciones, una variación detectada o no detectada reside en la muestra de ácido nucleico de una fuente, pero no en la muestra de ácido nucleico de otra fuente. Los ejemplos no limitativos 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 limitativos, una variación genética particular detectada o no detectada (i) reside en el ácido nucleico placentario pero no en el ácido nucleico fetal y no en el ácido nucleico materno; (ii) reside en el ácido nucleico fetal, pero no en el ácido nucleico materno; o (iii) reside en el ácido nucleico materno, pero no en el ácido nucleico fetal.

Después de generar uno o más resultados, se usa a menudo un resultado para proporcionar una determinación de la presencia o ausencia de una variación genética y/o afección médica asociada. Normalmente, se proporciona un resultado a un profesional de atención sanitaria (por ejemplo, técnico o gerente de laboratorio; médico o asistente). A menudo, un módulo de resultados proporciona un resultado. En determinadas implementaciones un resultado es proporcionado por un módulo de representación gráfica. En determinadas implementaciones, se proporciona un resultado en un periférico o componente de un aparato o una máquina. Por ejemplo, algunas veces una impresora o pantalla de visualización proporciona un resultado. En algunas implementaciones, un determinante del resultado de la presencia o ausencia de una variación genética se proporciona a un profesional sanitario en forma de un informe, y en determinadas implementaciones el informe comprende una visualización de un valor de resultado y un parámetro de confianza asociado. Generalmente, un resultado puede mostrarse en un formato adecuado que facilita la determinación de la presencia o ausencia de una variación genética y/o afección médica. Los ejemplos no limitativos de formatos adecuados para usar para informar y/o visualizar conjuntos de datos o para informar sobre un resultado incluyen datos digitales, un gráfico, un gráfico 2D, un gráfico 3D y un gráfico 4D, una imagen, un pictograma, una tabla, un gráfico de barras, un gráfico circular, un diagrama de flujo, un diagrama de dispersión, un mapa, un histograma, un gráfico de densidad, un gráfico de funciones, un diagrama de circuitos, un diagrama de bloques, un mapa de burbujas, un diagrama de constelaciones, un diagrama de contorno, un cartograma, un diagrama de araña, un diagrama de Venn, un nomograma, y similares, y una combinación de los anteriores. Varios ejemplos de representaciones de resultados se muestran en las figuras y se describen en los ejemplos.

Generar un resultado puede considerarse una transformación de datos leídos de secuencia de ácido nucleico, o similares, en una representación de un ácido nucleico celular de un sujeto, en determinadas implementaciones. Por ejemplo, el análisis de lecturas de secuencia de ácido nucleico de un sujeto y la generación de un resultado y/o perfil cromosómico puede considerarse una transformación de fragmentos de lectura de secuencia relativamente pequeños en una representación de una estructura cromosómica relativamente grande. En algunas implementaciones, un resultado procede de una transformación de lecturas de secuencia de un sujeto (por ejemplo, una mujer embarazada), en una representación de una estructura existente (por ejemplo, un genoma, un cromosoma o segmento del mismo) presente en el sujeto (por ejemplo, un ácido nucleico materno y/o fetal). En algunas implementaciones, un resultado comprende una transformación de lecturas de secuencia de un primer sujeto (por ejemplo, una mujer embarazada), 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 mujer embarazada) y/o un segundo sujeto (por ejemplo, un feto).

En determinadas implementaciones, se puede generar un resultado según el análisis de uno o más eventos de ondículas. En algunas implementaciones, la presencia o ausencia de una variación genética se determina según una ondícula, un evento de ondícula o un evento de ondícula compuesto (por ejemplo, la presencia o ausencia de una ondícula, un evento de ondícula o un evento de ondícula compuesto). En algunas implementaciones, dos eventos de ondícula derivados de dos representaciones de descomposición del mismo perfil son sustancialmente iguales (por ejemplo, según una comparación) y se determina la presencia de una aneuploidía, microduplicación o microdeleción cromosómica. En algunas implementaciones, la presencia de un evento de ondícula compuesto indica la presencia de una aneuploidía, microduplicación o microdeleción cromosómica. En algunas implementaciones, se determina la presencia de una aneuploidía cromosómica completa según la presencia de una ondícula, un evento de ondícula o un evento de ondícula compuesto en un perfil y el perfil es un segmento de un genoma (por ejemplo, un segmento más grande que un cromosoma, por ejemplo, un segmento que representa dos o más cromosomas, un segmento que representa un genoma completo). En algunas implementaciones, se determina la presencia de una aneuploidía cromosómica completa según la presencia de una ondícula, un evento de ondícula o un evento de ondícula compuesto en un perfil y los bordes de las ondículas son sustancialmente los mismos que los bordes de un cromosoma. En determinadas implementaciones, la presencia de una microduplicación o microdeleción se determina cuando al menos un borde de una ondícula, evento de ondícula o evento de ondícula compuesto de un perfil es diferente de un borde de un cromosoma y/o la ondícula está dentro de un cromosoma. En algunas implementaciones, se determina la presencia de una microduplicación, y un nivel o AUC para una ondícula, evento de ondícula o evento de ondícula compuesto es sustancialmente superior a un nivel de referencia (por ejemplo, una región euploide). En algunas implementaciones, se determina la presencia de una microdeleción, y un nivel o AUC para una ondícula, evento de ondícula o evento de ondícula compuesto es sustancialmente inferior a un nivel de referencia (por ejemplo, una región euploide). En algunas implementaciones, los eventos de ondícula identificados en dos o más representaciones de descomposición diferentes no son sustancialmente iguales (por ejemplo, son diferentes) y se determina la ausencia de aneuploidía, microduplicación y/o microdeleción cromosómica. En algunas implementaciones, la ausencia de una ondícula, evento de ondícula o evento de ondícula compuesto en un perfil o representación de descomposición de un perfil indica la ausencia de una aneuploidía cromosómica, microduplicación o microdeleción.

Validación

En algunas implementaciones, un método descrito en el presente documento comprende una validación. En algunas implementaciones, un análisis de decisiones (por ejemplo, un árbol de decisiones), una determinación de la presencia o ausencia de una variación genética (por ejemplo, una variación del número de copias, una microduplicación, una microdeleción, una aneuploidía), realizar una identificación y/o una determinación de un resultado comprende una validación. Puede utilizarse cualquier proceso de validación adecuado para validar un método, identificación o resultado descrito en el presente documento.

En algunas implementaciones, una validación comprende validar o invalidar un evento de ondícula identificado en una representación de descomposición. Un evento de ondícula validado confirma la presencia de un evento de ondícula. Un evento de ondícula invalidado cambia una identificación que indica la presencia de un evento de ondícula a la ausencia de un evento de ondícula. Por ejemplo, en algunas implementaciones, después de la identificación de un evento de ondícula mediante un proceso de segmentación, se puede realizar una validación donde el evento de ondícula se valida o se invalida. Un evento de ondícula que se invalida indica la ausencia de una aneuploidía, microduplicación o microdeleción cromosómica en un perfil. En algunas implementaciones, una validación comprende una determinación de la presencia o ausencia de un evento de ondícula con determinaciones de falsos negativos y/o falsos positivos reducidas. Un evento de ondícula se puede validar mediante un método adecuado, cuyos ejemplos no limitativos incluyen un proceso de “bordes deslizantes”, un proceso de “dejar una fuera”, similares o una combinación de los mismos.

En algunas implementaciones, una validación comprende generar un nivel de significación para un evento de ondícula o un evento de ondícula compuesto. En algunas implementaciones, el nivel de significación es una puntuación Z, un valor de z, un valor de *p* o similares. En algunas implementaciones, una validación comprende generar una medida de incertidumbre. En algunas implementaciones, una medida de incertidumbre está asociada con un nivel de significación. Por ejemplo, a veces se determina un nivel de significación promedio, medio o en mediana, y se determina una medida de incertidumbre para el nivel de significación promedio, medio o en mediana.

En algunas implementaciones, un evento de ondícula se valida o invalida según un nivel de significación y/o una medida de incertidumbre. Una ondícula validada o invalidada puede ser un evento de ondícula compuesto validado o invalidado. En algunas implementaciones, la presencia o ausencia de un evento de ondícula validado se determina según un nivel de significación y/o una medida de incertidumbre para un evento de ondícula. En algunas implementaciones, la ausencia de un evento de ondícula validado indica la ausencia de una aneuploidía, microduplicación o microdeleción cromosómica. En algunas implementaciones, la presencia de un evento de ondícula validado confirma la presencia de un evento de ondícula. En algunas implementaciones, la presencia de dos o más eventos de ondícula validados conduce a la determinación o generación de un evento de ondícula compuesto. En algunas implementaciones, la presencia de uno o más eventos de ondícula validados, en parte, determina la presencia de una aneuploidía, microduplicación o microdeleción cromosómica con un mayor nivel de confianza. En algunas implementaciones, la presencia de un evento de ondícula indica, en parte, la presencia de un síndrome de DiGeorge. En algunas implementaciones, la ausencia de un evento de ondícula validado indica la ausencia de una aneuploidía, microduplicación o microdeleción cromosómica.

Validación de bordes deslizantes

En algunas implementaciones, una validación comprende un proceso de "bordes deslizantes". Un proceso adecuado de "bordes deslizantes" se puede usar directamente o se puede adaptar para validar una ondícula. En algunas implementaciones, un proceso de "bordes deslizantes" comprende segmentar un evento de ondícula (por ejemplo, un evento de ondícula representado por un conjunto de porciones), o un segmento que se sospecha que comprende un evento de ondícula, en múltiples subconjuntos de porciones. En algunas implementaciones, el evento de ondícula es un conjunto de porciones para un cromosoma completo o un segmento de un cromosoma. En algunas implementaciones, el evento de ondícula es un conjunto de porciones que comprenden una región asociada con una variación genética conocida o un trastorno genético conocido. En algunas implementaciones, el evento de ondícula comprende una región de DiGeorge.

En determinadas implementaciones, un proceso de "bordes deslizantes" comprende segmentar un evento de ondícula identificado (un conjunto de porciones) en múltiples subconjuntos de porciones donde cada uno de los subconjuntos de porciones representa un evento de ondícula con bordes similares pero diferentes. En algunas implementaciones, el evento de ondícula identificado originalmente se incluye en el análisis. Por ejemplo, el evento de ondícula identificado originalmente se incluye como uno de los múltiples subconjuntos de porciones. Los subconjuntos de porciones se pueden determinar variando uno o ambos bordes de la ondícula originalmente identificada mediante cualquier método adecuado. En algunas implementaciones, el borde izquierdo se puede cambiar generando de este modo ondículas con diferentes bordes izquierdos. En algunas implementaciones, el borde derecho se puede cambiar generando de este modo ondículas con diferentes bordes derechos. En algunas implementaciones se pueden cambiar los bordes derecho e izquierdo. En algunas implementaciones, los bordes se cambian moviendo el borde una o más porciones adyacentes de un genoma de referencia hacia la izquierda o hacia la derecha de los bordes originales.

En una implementación de un enfoque de bordes deslizantes, la ondícula original se cambia moviendo ambos bordes en 15 porciones de un genoma de referencia, creando de este modo una cuadrícula de ondículas de 15 por 15 (por ejemplo, 225 subconjuntos diferentes de porciones). Por ejemplo, mientras se mantiene estable el borde derecho, el borde izquierdo se puede mover hacia la derecha en 7 porciones de un genoma de referencia y luego hacia la izquierda en 7 porciones de un genoma de referencia, generando de este modo 15 posibles bordes izquierdos. Mientras se mantiene estable cada uno de los 15 bordes izquierdos, el borde derecho se puede mover hacia la derecha en 7 porciones de un genoma de referencia y hacia la izquierda en siete porciones de un genoma de referencia, generando de este modo 15 posibles bordes derechos. Los subconjuntos resultantes comprenden 225 ondículas diferentes (por ejemplo, subconjuntos de porciones de un genoma de referencia).

En algunas implementaciones, uno o ambos bordes se cambian en entre 5 y 30 porciones de un genoma de referencia. En algunas implementaciones, un borde se mueve en 1, 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 o 30 porciones de un genoma de referencia en cualquier dirección. En algunas implementaciones, independientemente del tamaño de la porción, se cambia un borde para generar un rango de borde de aproximadamente 100.000 a aproximadamente 2.000.000 pares de bases, de 250.000 a aproximadamente 1.500.000 pares de bases, o de aproximadamente 500.000 a aproximadamente 1.000.000 pares de bases para uno o ambos bordes. En algunas implementaciones, independientemente del tamaño de la porción, se cambia un borde para generar un rango de borde de aproximadamente 500.000, 600.000, 700.000, 750.000, 800.000, 900.000 o aproximadamente 1.000.000 pares de bases para uno o ambos bordes.

En algunas implementaciones, una ondícula identificada comprende un primer extremo y un segundo extremo, y la segmentación comprende (i) eliminar una o más porciones del primer extremo del conjunto de porciones mediante eliminación recursiva, proporcionando de este modo un subconjunto de porciones con cada eliminación recursiva, (ii) terminar la eliminación recursiva en (i) después de n repeticiones proporcionando de este modo n + 1 subconjuntos de porciones, donde el conjunto de porciones es un subconjunto, y donde cada subconjunto comprende un número diferente de porciones, un primer subconjunto final y un segundo subconjunto final, (iii) eliminar una o más porciones del extremo del segundo subconjunto de cada uno de los n + 1 subconjuntos de porciones proporcionados en (ii) mediante eliminación recursiva; y (iv) terminar la eliminación recursiva en (iii) después de n repeticiones, proporcionando de este modo múltiples subconjuntos de porciones. En algunas implementaciones, los múltiples subconjuntos equivalen a (n + 1)2 subconjuntos.

En algunas implementaciones, n es igual a un número entero entre 5 y 30. En algunas implementaciones n es igual a 1, 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 o 30.

En determinadas implementaciones de un enfoque de bordes deslizantes, se determina un nivel de significación (por ejemplo, una puntuación Z, un valor de *p*) para cada uno de los subconjuntos de porciones de un genoma de referencia y se determina un nivel de significación promedio, medio o en mediana según el nivel de significación determinado para todos los subconjuntos.

En algunas implementaciones, el nivel de significación es una puntuación Z o un valor de *p*. En algunas implementaciones, una puntuación Z se calcula según la siguiente fórmula:

$$Z_i=(E_i\text{-Med. } E(n))/DAM$$

donde E_i es una determinación cuantitativa del nivel de la ondícula i, Med.$E_{(n)}$ es la mediana del nivel para todas las ondículas generadas por un proceso de bordes deslizantes y DAM es la mediana de la desviación absoluta para Med.$E_{(n)}$, y Z_i es la puntuación Z resultante para la ondícula i. En algunas implementaciones, DAM puede reemplazarse mediante cualquier medida de incertidumbre adecuada. En algunas implementaciones, E_i es cualquier medida adecuada de un nivel, cuyos ejemplos no limitativos incluyen la mediana de un nivel, nivel promedio, nivel medio, suma de los recuentos de las porciones, o similares.

En algunas implementaciones, se determina una puntuación Z en mediana, media o promedio para todas las ondículas generadas mediante un proceso de bordes deslizantes, y se genera una medida de incertidumbre (por ejemplo, DAM) a partir de la puntuación Z en mediana, media o promedio. En algunas implementaciones, una ondícula (por ejemplo, la ondícula original identificada) se valida o invalida según la puntuación Z en mediana, media o promedio determinada para todas las ondículas generadas por un proceso de bordes deslizantes y una medida de incertidumbre para la puntuación Z en mediana, media o promedio. En algunas implementaciones, se predetermina un rango predeterminado (por ejemplo, un rango umbral) para el nivel de significación (por ejemplo, una puntuación Z). En algunas implementaciones, el rango predeterminado para una puntuación Z para la ausencia de un evento de ondícula es de aproximadamente 3,5 a aproximadamente -3,5, de aproximadamente 3,25 a aproximadamente -3,25, de aproximadamente 3,0 a aproximadamente -3,0, de aproximadamente 2,75 a aproximadamente - 2,75 o de aproximadamente 2,5 a aproximadamente -2,5. En algunas implementaciones, una puntuación Z en mediana, media o promedio con un valor fuera del rango predeterminado confirma la presencia de una ondícula validada según el método de "bordes deslizantes". En algunas implementaciones, una puntuación Z en mediana, media o promedio con un valor dentro del rango predeterminado invalida un evento de ondícula según el método de "bordes deslizantes" y/o determina la ausencia de un evento de ondícula (por ejemplo, la ausencia de un evento de ondícula validado). En algunas implementaciones, una puntuación Z mediana, media o promedio con un valor absoluto superior a aproximadamente 2, 2,25, 2,5, 2,75, 3,0, 3,25 o 3,5 confirma la presencia de y/o valida una ondícula según el método de "bordes deslizantes". En algunas implementaciones, una puntuación Z en mediana, media o promedio con un valor absoluto inferior a aproximadamente 2, 2,25, 2,5, 2,75, 3,0, 3,25 o 3,5 determina la ausencia y/o invalida un evento de ondícula según el método de "bordes deslizantes". En algunas implementaciones, una medida de incertidumbre asociada con una mediana de la puntuación Z determina, en parte, si una ondícula se valida o invalida. En algunas implementaciones, un evento de ondícula se valida si la puntuación Z en mediana, media o promedio está fuera de un rango umbral, y la medida de incertidumbre (por ejemplo, la DAM) se superpone con el rango umbral en menos del 0 % (por ejemplo, no se superpone), 5 %, 10 %, 20 %, 25 %, 30 %, 35 % o 40 % de la medida de incertidumbre. En algunas implementaciones, un evento de ondícula se invalida si la puntuación Z en mediana, media o promedio está fuera de un rango umbral y la medida de incertidumbre (por ejemplo, la DAM) se superpone con el rango umbral en más de aproximadamente el 25 %, 30 %, 40 %, 50 %, 60 % o más de aproximadamente el 70 % de la medida de incertidumbre.

En algunas implementaciones, se genera una distribución para el nivel de significación (por ejemplo, puntuaciones Z) determinado para todas las ondículas generadas por un proceso de bordes deslizantes. En determinadas implementaciones, una ondícula se valida o invalida según el nivel de significación en mediana, medio o promedio y/o una distribución del nivel de significación. En algunas implementaciones, aproximadamente el 50 %, 60 %, 70 %, 75 %, 80 %, 85 %, 90 % o aproximadamente el 95 % o más de una distribución está fuera de un rango predeterminado para el nivel de significación, y se valida una ondícula. Por ejemplo, para un rango predeterminado de puntuaciones Z de 3,0 a -3,0, un evento de ondícula validado puede tener una mediana de la puntuación Z y el 70 % o más de la distribución de puntuaciones Z con un valor absoluto superior a 3,0.

Validación de dejar una fuera

En algunas implementaciones, una validación comprende un proceso de "dejar una fuera". Se puede utilizar un proceso adecuado de "dejar una fuera". En algunas implementaciones, un proceso de "dejar una fuera" proporciona un nivel de confianza asociado con un conjunto seleccionado de muestras de referencia. En algunas implementaciones, un proceso de "dejar una fuera" proporciona una medida de incertidumbre asociada con un conjunto seleccionado de muestras de referencia. En algunas implementaciones, un proceso de "dejar una fuera" valida o invalida un evento de ondícula según un nivel de confianza y/o medida de incertidumbre determinada según un conjunto seleccionado de muestras de referencia.

En algunas implementaciones, se realiza un proceso de "dejar una fuera" para una muestra de prueba y dos o más muestras de referencia (por ejemplo, un conjunto de muestras de referencia, a veces denominado en el presente documento conjunto original). En algunas implementaciones, la muestra de prueba se incluye como una de las dos o más muestras de referencia. En algunas implementaciones, la muestra de prueba no se incluye como una de las dos o más muestras de referencia. En algunas implementaciones, el proceso de "dejar una fuera" comprende eliminar una de dos o más muestras de referencia del conjunto original de muestras proporcionando de este modo un subconjunto de muestras de referencia. En determinadas implementaciones, el proceso de eliminar una muestra de referencia del conjunto original se repite para cada muestra de referencia del conjunto. A menudo, cuando se elimina una muestra de referencia del conjunto original, la muestra de referencia previamente eliminada, si la hubiera, se devuelve al conjunto original. En algunas implementaciones, solo se elimina una muestra de referencia de cualquier subconjunto. El resultado suele ser múltiples subconjuntos de muestras de referencia (a veces denominados en el presente documento múltiples subconjuntos de muestras) donde a cada subconjunto le falta una de las muestras de referencia del conjunto original.

En determinadas implementaciones, el proceso de "dejar una fuera" comprende determinar un nivel de significación según cada subconjunto de los subconjuntos de muestras de referencia. En determinadas implementaciones, se calcula entonces un nivel de significación medio, promedio o en mediana a partir de los valores de nivel de significación determinados para todos los subconjuntos. En algunas implementaciones, se calcula una medida de incertidumbre (por ejemplo, una DAM) según el nivel de significación medio, promedio o en mediana. En algunas implementaciones, una ondícula se valida o invalida según un nivel de significación en mediana, medio o promedio y/o la medida de incertidumbre generada según el proceso de "dejar una fuera".

En determinadas implementaciones del proceso "dejar una fuera", un nivel de significación es una puntuación Z o un valor de *p*. En algunas implementaciones, una puntuación Z para el proceso de "dejar una fuera" se calcula según la siguiente fórmula:

$$Z_i=(E_i\text{-Med. } E_{(n)})/\text{DAM}$$

donde E_i es una determinación cuantitativa del nivel de la ondícula i, Med.$E_{(n)}$ es la mediana del nivel para la ondícula i para un subconjunto de muestras de referencia y DAM es la mediana de la desviación absoluta para Med.$E_{(n)}$, y Z_i es la puntuación Z resultante para la ondícula i. En algunas implementaciones, una DAM puede reemplazarse mediante cualquier medida de incertidumbre adecuada. En algunas implementaciones, E_i es cualquier medida adecuada de un nivel, cuyos ejemplos no limitativos incluyen la mediana de un nivel, nivel promedio, nivel medio, suma de los recuentos de las porciones, o similares.

En algunas implementaciones, una validación comprende un proceso de "bordes deslizantes" y un proceso de "dejar una fuera". Por ejemplo, en algunas implementaciones, los subconjuntos de muestras de referencia (por ejemplo, generados a partir del proceso de "dejar una fuera") se generan a partir de un conjunto de muestras de referencia generado mediante el "proceso de bordes deslizantes". Por ejemplo, para una muestra de prueba dada, un proceso de "bordes deslizantes" puede producir 225 ondículas para una ondícula identificada a partir de un proceso de segmentación y luego se realiza un proceso de "dejar una fuera" utilizando un conjunto de 10 muestras de referencia. En el ejemplo anterior, se calcula una mediana compuesta, una media o un promedio del nivel de significación (por ejemplo, una mediana compuesta de la puntuación Z) y una medida compuesta de incertidumbre (por ejemplo, una DAM compuesta) a partir de las 2250 puntuaciones Z resultantes. En algunas implementaciones, una ondícula identificada mediante un proceso de segmentación se valida o invalida según una mediana compuesta del nivel de significación (por ejemplo, una mediana compuesta de la puntuación Z) y/o una medida compuesta de incertidumbre (por ejemplo, una DAM compuesta).

En algunas implementaciones, un análisis de decisiones comprende determinar la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica según la puntuación Z o la puntuación Z compuesta para un evento de ondícula (por ejemplo, un evento de ondícula compuesto). En algunas implementaciones, un evento de ondícula es indicativo de una trisomía, y el evento de ondícula es para un conjunto de porciones que representan un cromosoma completo. En determinadas implementaciones, un evento de ondícula es indicativo de una aneuploidía de un cromosoma completo cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a un valor o umbral predeterminado. En determinadas implementaciones, un evento de ondícula es indicativo de una aneuploidía de un cromosoma completo cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a un valor predeterminado de aproximadamente 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,75, 3,8, 3,85, 3,9, 3,95, 4,0, 4,05, 4,1, 4,15, 4,2, 4,3, 4,4 o aproximadamente 4,5. En determinadas implementaciones, un evento de ondícula es indicativo de una trisomía cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a 3,95. En determinadas implementaciones, un evento de ondícula es indicativo de una trisomía cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual al valor absoluto de una puntuación Z determinada para (i) una ondícula identificada según un proceso de descomposición de ondículas de Haar o (ii) una ondícula identificada según un proceso CBS. En determinadas implementaciones, un evento de ondícula es indicativo de una trisomía cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a un múltiplo del valor absoluto de una puntuación Z determinada para (i) una ondícula identificada según un proceso de descomposición de ondículas de Haar o (ii) una ondícula identificada según

un proceso CBS. En algunas implementaciones, un múltiplo del valor absoluto de una puntuación Z es el valor absoluto de una puntuación Z multiplicado por aproximadamente 0,4, 0,5, 0,6, 0,7, 0,8 o aproximadamente 0,9.

En determinadas implementaciones, un evento de ondícula (por ejemplo, un evento de ondícula significativo) es indicativo de una trisomía cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a 3,95 y es superior o igual al valor absoluto de una puntuación Z determinada para (i) una ondícula identificada según un proceso de descomposición de ondículas de Haar o (ii) una ondícula identificada según un proceso CBS. En determinadas implementaciones, un evento de ondícula es indicativo de una trisomía cuando la puntuación Z absoluta para un conjunto de porciones que representan un cromosoma completo es superior o igual a 3,95 y es superior o igual a un múltiplo del valor absoluto de una puntuación Z determinada para (i) una ondícula identificada según un proceso de descomposición de ondículas de Haar o (ii) una ondícula identificada según un proceso CBS. En algunas implementaciones, un múltiplo del valor absoluto de una puntuación Z es el valor absoluto de una puntuación Z multiplicado por aproximadamente 0,4, 0,5, 0,6, 0,7, 0,8 o aproximadamente 0,9.

En algunas implementaciones, un evento de ondícula no es indicativo de una trisomía, y se determina la presencia de una microdeleción o microduplicación cuando el valor absoluto de una puntuación Z determinada para (i) la ondícula identificada según un proceso de descomposición de ondículas de Haar y (ii) la ondícula identificada según un proceso CBS es superior o igual a aproximadamente 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,75, 3,8, 3,85, 3,9, 3,95, 4,0, 4,05, 4,1, 4,15, 4,2, 4,3, 4,4 o aproximadamente 4,5. En algunas implementaciones, un evento de ondícula no es indicativo de una trisomía y se determina la presencia de una microdeleción o microduplicación. En algunas implementaciones, un evento de ondícula no es indicativo de una trisomía, y se determina la presencia de una microdeleción o microduplicación cuando el valor absoluto de una puntuación Z determinada para (i) la ondícula identificada según un proceso de descomposición de ondículas de Haar y (ii) la ondícula identificada según un proceso CBS es superior o igual a 3,95. En algunas implementaciones, un evento de ondícula no es indicativo de una trisomía y se determina la presencia de una microdeleción o microduplicación, y la ondícula identificada según un proceso de descomposición de ondículas de Haar es sustancialmente la misma que la ondícula identificada según un proceso CBS.

En algunas implementaciones, la determinación de un resultado (por ejemplo, la determinación de la presencia o ausencia de una variación genética, por ejemplo, en un feto) comprende un análisis de decisiones. En algunas implementaciones, un método para determinar la presencia o ausencia de una aneuploidía, microduplicación o microdeleción cromosómica en un feto con determinaciones reducidas de falsos negativos y falsos positivos, comprende un análisis de decisiones. En algunas implementaciones, un análisis de decisiones comprende una serie de métodos o etapas de métodos.

Uso de resultados

Un profesional sanitario, u otro individuo cualificado, que recibe un informe que comprende uno o más resultados determinantes de la presencia o ausencia de una variación genética, puede usar los datos visualizados en el informe para realizar una identificación en relación con el estado del paciente o sujeto de prueba. El profesional sanitario puede realizar una recomendación basada en el resultado proporcionado, en algunas implementaciones. Un profesional sanitario o individuo cualificado puede proporcionar a un paciente o sujeto de prueba una identificación o puntuación con respecto a la presencia o ausencia de la variación genética basándose en el valor o valores de resultado y parámetros de confianza asociados proporcionados en un informe, en algunas implementaciones. En determinadas implementaciones, un profesional sanitario o un individuo cualificado realiza una puntuación o identificación manualmente, usando la observación visual del informe proporcionado. En determinadas implementaciones, una puntuación o identificación se realiza mediante una rutina automatizada, algunas veces integrada en software, y revisada por un profesional sanitario o individuo cualificado para obtener precisión antes de proporcionar información a un paciente o sujeto de prueba. La expresión “recibir un informe”, tal como se usa en el presente documento, se refiere a obtener, mediante un medio de comunicación, una representación escrita y/o gráfica que comprende un resultado, que después de la revisión permite a un profesional sanitario u otro individuo cualificado determinar la presencia o ausencia de una variación genética en un paciente o sujeto de prueba. El informe puede generarse mediante un ordenador o mediante la introducción de datos por seres humanos, y puede comunicarse usando medios electrónicos (por ejemplo, a través de Internet, a través de ordenador, a través de fax, desde una ubicación de red a otra ubicación en el mismo sitio físico o en sitios físicos diferentes), o mediante otro método para enviar o recibir datos (por ejemplo, servicio de correo, servicio de mensajería y similares). En algunas implementaciones, el resultado se transmite a un profesional de atención sanitaria en un medio adecuado incluyendo, sin limitación, de modo verbal, en forma de documento o archivo. El archivo puede ser, por ejemplo, pero sin limitarse a, un archivo acústico, un archivo legible por ordenador, un archivo en papel, un archivo de laboratorio o un archivo de historial clínico.

La expresión “proporcionar un resultado” y equivalentes gramaticales de la misma, tal como se usa en el presente documento puede referirse además a un método para obtener tal información incluyendo, sin limitación, obtener la información de un laboratorio (por ejemplo, un archivo de laboratorio). Un archivo de laboratorio puede generarse por un laboratorio que lleva a cabo uno o más ensayos o una o más etapas de procesamiento de datos para determinar la presencia o ausencia de la afección médica. El laboratorio puede estar en la misma ubicación o en una ubicación diferente (por ejemplo, en otro país) que el personal que identifica la presencia o ausencia de la afección medica del archivo de laboratorio. Por ejemplo, el archivo de laboratorio puede generarse en una ubicación y transmitirse a otra ubicación en la

cual la información en la misma se transmitirá al sujeto femenino gestante. El archivo de laboratorio puede estar en forma tangible o en forma electrónica (por ejemplo, forma legible por ordenador), en determinadas implementaciones.

En algunas implementaciones, puede proporcionarse un resultado a un profesional de atención sanitaria, médico o individuo cualificado de un laboratorio y el profesional de atención sanitaria, médico o individuo cualificado puede realizar un diagnóstico basándose en el resultado. En algunas implementaciones, puede proporcionarse un resultado a un profesional de atención sanitaria, médico o individuo cualificado de un laboratorio y el profesional de atención sanitaria, médico o individuo cualificado puede realizar un diagnóstico basado, en parte, en el resultado junto con información y/o datos adicionales y otros resultados.

Un profesional sanitario o individuo cualificado puede proporcionar una recomendación adecuada basándose en los resultados o resultados proporcionados en el informe. Los ejemplos no limitativos de recomendaciones que pueden proporcionarse basándose en el informe de resultados proporcionado incluyen cirugía, radioterapia, quimioterapia, asesoramiento genético, soluciones de tratamiento después del nacimiento (por ejemplo, planificación vital, cuidado asistido a largo plazo, medicamentos, tratamientos sintomáticos), interrupción del embarazo, trasplante de órganos, transfusión de sangre, similares o combinaciones de los anteriores. En algunas implementaciones, la recomendación depende de la clasificación basada en los resultados proporcionada (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 prueba, recuentos de referencia, nivel de desviación) subyacentes a una determinación de la presencia o ausencia de una variación genética (o determinación de euploides o no euploides para una región de prueba). Para las identificaciones referidas a la presencia o ausencia de una variación genética cercana o cuestionable, el personal de laboratorio puede volver a ordenar la misma prueba y/u ordenar una prueba diferente (por ejemplo, cariotipado y/o amniocentesis en el caso de determinaciones de aneuploidía fetal), que usan el mismo ácido nucleico de muestra o diferente de un sujeto de prueba.

Variaciones genéticas y afecciones médicas

La presencia o ausencia de una varianza genética puede determinarse usando un método, una máquina o un aparato descrito en el presente documento. En determinadas implementaciones, la presencia o ausencia de una o más variaciones genéticas se determina según un resultado proporcionado mediante los métodos, máquinas y aparatos descritos en el presente documento. Una variación genética es generalmente un fenotipo genético particular presente en determinados individuos y a menudo una variación genética está presente en una subpoblación estadísticamente significativa de individuos. En algunas implementaciones, una variación genética es una anomalía cromosómica (por ejemplo, aneuploidía), anomalía cromosómica parcial o mosaicismo, cada uno de los cuales se describe con mayor detalle en el presente documento. Los ejemplos no limitativos de variaciones genéticas incluyen una o más deleciones (por ejemplo, microdeleciones), duplicaciones (por ejemplo, microduplicaciones), inserciones, mutaciones, polimorfismos (por ejemplo, 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 ser de cualquier longitud, y en algunas implementaciones, tiene aproximadamente 1 base o par de bases (pb) a aproximadamente 250 megabases (Mb) de longitud. En algunas implementaciones, una inserción, repetición, deleción, duplicación, mutación o polimorfismo tiene de aproximadamente 1 base o par de bases (pb) a aproximadamente 1000 kilobases (kb) de longitud (por ejemplo, aproximadamente 10 pb, 50 pb, 100 pb, 500 pb, 1 kb, 5 kb, 10 kb, 50 kb, 100 kb, 500 kb o 1000 kb de longitud).

Una variación genética es algunas veces una deleción. En determinadas implementaciones, una deleción es una mutación (por ejemplo, una aberración genética) en la que falta una parte de un cromosoma o una secuencia de ADN. Una deleción es a menudo la pérdida de material genético. Puede eliminarse cualquier número de nucleótidos. Una deleción puede comprender la deleción 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 combinación de los mismos. Una deleción puede comprender una microdeleción. Una deleción puede comprender la deleción de una sola base.

Algunas veces, una variación genética es una duplicación genética. En determinadas implementaciones, una duplicación es una mutación (por ejemplo, una aberración genética) en la que una parte de un cromosoma o una secuencia de ADN se copia y se inserta de nuevo en el genoma. En determinadas implementaciones, una duplicación genética (es decir, duplicación) es cualquier duplicación de una región de ADN. En algunas implementaciones, una duplicación es una secuencia de ácido nucleico que se repite a menudo en tándem, dentro de un genoma o cromosoma. En algunas implementaciones, una duplicación puede comprender una copia de uno o más cromosomas enteros, un segmento de un cromosoma, un alelo, un gen, un intrón, un exón, cualquier región no codificante, cualquier región codificante, segmento de los mismos o combinación de los mismos. Una duplicación puede comprender una microduplicación. Una duplicación comprende, algunas veces, 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 variar desde regiones pequeñas (miles de pares de bases) hasta cromosomas enteros en algunos casos. Las duplicaciones se producen a menudo como resultado de un error en la recombinación homóloga o debido a un evento de retrotransposón. Las

duplicaciones se han asociado con determinados tipos de enfermedades proliferativas. Las duplicaciones pueden caracterizarse con el uso de microalineamientos genómicos o hibridación genética comparativa (HGC).

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

Tal como se usa en el presente documento, una "variación del número de copias" es generalmente una clase o un tipo de variación genética o aberración cromosómica. Una variación del número de copias puede ser una deleción (por ejemplo, microdeleción), duplicación (por ejemplo, una microduplicación) o inserción (por ejemplo, una microinserción). A menudo, el prefijo "micro", tal como se usa en el presente documento, algunas veces es un segmento de ácido nucleico con una longitud menor de 5 Mb. Una variación del número de copias puede incluir una o más deleciones (por ejemplo, microdeleción), duplicaciones y/o inserciones (por ejemplo, una microduplicación, microinserción) de un segmento de un cromosoma. En determinadas implementaciones, una duplicación comprende una inserción. En determinadas implementaciones, una inserción es una duplicación. En determinadas implementaciones, una inserción no es una duplicación. Por ejemplo, a menudo una duplicación de una secuencia en una porción aumenta los recuentos de la porción en la que se encuentra la duplicación. A menudo, una duplicación de una secuencia en una porción aumenta la elevación o el nivel. En determinadas implementaciones, una duplicación presente en porciones que conforman una primera elevación o nivel aumenta la elevación o nivel en relación con una segunda elevación o nivel donde la duplicación está ausente. En determinadas implementaciones, una inserción aumenta los recuentos de una porción, y una secuencia que representa la inserción está presente (es decir, duplicada) en otro lugar dentro de la misma porción. En determinadas implementaciones, una inserción no aumenta significativamente los recuentos de una porción o elevación o nivel, y la secuencia que se inserta no es una duplicación de una secuencia dentro de la misma porción. En determinadas implementaciones, 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 algunas implementaciones, una variación del número de copias es una variación del número de copias fetal. A menudo, una variación del número de copias fetal es una variación del número de copias en el genoma de un feto. En algunas implementaciones, una variación del número de copias es una variación del número de copias materno y/o fetal. En algunas implementaciones, una variación del número de copias materno y/o fetal es una variación del número de copias dentro del genoma de una mujer embarazada (por ejemplo, una mujer que porta un feto), una mujer que da a luz o una mujer capaz de portar un feto. Una variación del número de copias puede ser una variación del número de copias heterocigotas en la que la variación (por ejemplo, una duplicación o deleción) está presente en un alelo de un genoma. Una variación del número de copias puede ser una variación del número de copias homocigotas en la que la variación está presente en ambos alelos de un genoma. En algunas implementaciones, una variación del número de copias es una variación del número de copias heterocigotas u homocigotas fetal. En algunas implementaciones, una variación del número de copias es una variación del número de copias heterocigotas u homocigotas materno y/o fetal. Algunas veces, una variación del número de copias está presente en un genoma materno y un genoma fetal, un genoma materno y no en un genoma fetal, o un genoma fetal y no en un genoma materno.

"Ploidía" es una referencia al número de cromosomas presentes en un feto o su madre. En determinadas implementaciones, la "ploidía" es lo mismo que "ploidía cromosómica". En seres humanos, por ejemplo, los cromosomas autosómicos están presentes a menudo en pares. Por ejemplo, en ausencia de una variación genética, la mayoría de los seres humanos tienen dos de cada cromosoma autosómico (por ejemplo, cromosomas 1-22). La presencia del complemento normal de 2 cromosomas autosómicos en un ser humano se denomina a menudo euploide. "Microploidía" tiene un significado similar a ploidía. "Microploidía" se suele referir 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 del número de copias (por ejemplo, una deleción, duplicación y/o una inserción) dentro de un cromosoma (por ejemplo, una deleción, duplicación o inserción homocigota o heterocigota, similares o ausencia de las mismas). La "ploidía" y la "microploidía" a veces se determinan después de la normalización de los recuentos de una elevación o un nivel en un perfil. Por tanto, una elevación o un nivel que represente un par de cromosomas autosómicos (por ejemplo, un euploide) se normaliza a menudo a una ploidía de 1. Del mismo modo, una elevación o un nivel dentro de un segmento de un cromosoma que represente la ausencia de una duplicación, deleción o inserción se normaliza a menudo a una microploidía de 1. La ploidía y la microploidía son a menudo específicas de un bin o de una porción (por ejemplo, específicas de una porción) y específicas de una muestra. La ploidía se define a menudo como múltiplos enteros de ½, representando los valores de 1, ½, 0, 3/2 y 2 euploide (por ejemplo, 2 cromosomas), 1 cromosoma presente (por ejemplo, una deleción cromosómica), ningún cromosoma presente, 3 cromosomas (por ejemplo, una trisomía) y 4 cromosomas, respectivamente. De manera similar, la microploidía se define a menudo como múltiplos enteros de ½, representando los valores de 1, ½, 0, 3/2 y 2 euploide (por ejemplo, sin variación en el número de copias), una deleción

heterocigota, deleción homocigota, duplicación heterocigota y duplicación homocigota, respectivamente. Algunos ejemplos de valores de ploidía para un feto se proporcionan en la tabla 2.

En algunas implementaciones, la microploidía de un feto coincide con la microploidía de la madre del feto (es decir, el sujeto femenino gestante). En determinadas implementaciones, 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 del número de copias heterocigotas, la variación del número de copias homocigotas o ambos son euploides. En determinadas implementaciones, 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 heterocigota para una variación del número de copias, la madre es homocigota para una variación del número de copias y la microploidía del feto no coincide (por ejemplo, no es igual) con la microploidía de la madre para la variación del número de copias especificada.

Una microploidía se asocia a menudo con una elevación o un nivel esperado. Por ejemplo, a veces una elevación o un nivel (por ejemplo, una elevación o un nivel en un perfil, a veces una elevación o un nivel que no incluye sustancialmente ninguna variación del 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 homocigota es 2, una duplicación heterocigota es 1,5, una deleción heterocigota es 0,5 y una deleción homocigota es cero.

Una variación genética para la cual se identifica la presencia o ausencia para un sujeto se asocia con una afección médica en determinadas implementaciones. Por tanto, la tecnología descrita en el presente documento puede usarse para identificar la presencia o ausencia de una o más variaciones genéticas asociadas con una afección médica o un estado médico. Los ejemplos no limitativos de afecciones médicas incluyen las 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, hongo, levadura) y preeclampsia.

Los ejemplos no limitativos de variaciones genéticas, afecciones y estados médicos se describen más adelante.

Sexo del feto

En algunas implementaciones, la predicción de un sexo del feto o trastorno relacionado con el sexo (por ejemplo, aneuploidía de cromosoma sexual) puede determinarse mediante un método, un máquina y/o un aparato descrito en el presente documento. La determinación del sexo se basa generalmente en un cromosoma sexual. En seres humanos, hay dos cromosomas sexuales, los cromosomas X e Y. El cromosoma Y contiene un gen, SRY, que desencadena el desarrollo embrionario como hombre. Los cromosomas Y de seres humanos y otros mamíferos contienen además otros genes necesarios para la producción normal de esperma. Los individuos con XX son de sexo femenino y XY son de sexo masculino y variaciones no limitativas, a menudo denominadas aneuploidías en cromosomas sexuales, incluyen X0, XYY, XXX y XXY. En determinadas implementaciones, los hombres 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 Triple X) o un solo cromosoma X en lugar de dos (X0; Síndrome de Turner). En determinadas implementaciones, solo una porción de las células en un individuo se ve afectada por una aneuploidía en cromosomas sexuales, que puede denominarse mosaicismo (por ejemplo, mosaicismo de Turner). Otros casos incluyen aquellos en los que SRY se daña (lo que conduce a una mujer XY) o se copia en la X (lo que conduce a un hombre XX).

En determinados casos, puede ser beneficioso determinar el sexo de un feto en el útero. Por ejemplo, un paciente (por ejemplo, mujer embarazada) con antecedentes familiares de uno o más trastornos ligados al sexo puede desear determinar el sexo del feto que porta para ayudar a evaluar el riesgo de que el feto herede tal trastorno. Los trastornos ligados al sexo incluyen, sin limitación, trastornos ligados al cromosoma X e Y. Los trastornos ligados al cromosoma X incluyen trastornos recesivos ligados al cromosoma X y dominantes ligados al cromosoma X. Los ejemplos de trastornos recesivos ligados al cromosoma X incluyen, sin limitación, trastornos inmunitarios (por ejemplo, enfermedad granulomatosa crónica (CYBB), síndrome de Wiskott-Aldrich, inmunodeficiencia combinada severa ligada al cromosoma X, agammaglobulinemia ligada al cromosoma X, síndrome de hiper-IgM tipo 1, IPEX, enfermedad linfoproliferativa ligada al cromosoma X, deficiencia de properdina), trastornos hematológicos (por ejemplo, hemofilia A, hemofilia B, anemia sideroblástica ligada al cromosoma X), trastornos endocrinos (por ejemplo, síndrome de insensibilidad a andrógenos/enfermedad de Kennedy, síndrome de Kallmann KAL1, hipoplasia suprarrenal congénita ligada al cromosoma X), trastornos metabólicos (por ejemplo, deficiencia de ornitina transcarbamilasa, síndrome oculocerebrorrenal, adrenoleucodistrofia, deficiencia de glucosa-6-fosfato deshidrogenasa, deficiencia de piruvato deshidrogenasa, enfermedad de Danon/glucogenosis tipo lib, enfermedad de Fabry, síndrome de Hunter, síndrome de Lesch-Nyhan, enfermedad de Menkes/síndrome de asta occipital), trastornos del sistema nervioso (por ejemplo, síndrome de Coffin-Lowry, síndrome MASA, síndrome de alfa talasemia con retraso mental ligado al cromosoma X, síndrome de Siderius con retraso mental ligado al cromosoma 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 (por ejemplo, disqueratosis congénita, displasia ectodérmica hipohidrótica (DEH), ictiosis ligada al cromosoma X, distrofia corneal endotelial ligada al cromosoma X), trastornos neuromusculares (por ejemplo, distrofia muscular de Becker/Duchenne, miopatía centronuclear (MTM1), síndrome de Conradi-Hunermann, distrofia muscular de Emery-Dreifuss 1), trastornos urológicos (por ejemplo, síndrome de Alport, enfermedad de Dent, diabetes insípida nefrógena ligada al cromosoma X), trastornos óseos/dentales (por ejemplo, AMELX amelogénesis 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-Tranebjærg, síndrome nasodigitoacústico). Los ejemplos de trastornos dominantes ligados al cromosoma X incluyen, sin limitación, hipofosfatemia ligada al cromosoma X, hipoplasia dérmica focal, síndrome del cromosoma X frágil, síndrome de Aicardi, incontinencia pigmentaria, síndrome de Rett, síndrome CHILD, síndrome de Lujan-Fryns y síndrome bucofaciodigital 1. Los ejemplos de trastornos ligados al cromosoma Y incluyen, sin limitación, esterilidad masculina, retinitis pigmentosa y azoospermia.

Anomalías cromosómicas

En algunas implementaciones, la presencia o ausencia de una anomalía cromosómica fetal puede determinarse usando un método, una máquina o un aparato 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), incluyendo deleciones y duplicaciones provocadas por translocaciones desequilibradas. La expresión ‘‘anomalía cromosómica’’ o ‘‘aneuploidía’’, 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 patrón de bandas que se encuentra en individuos sanos de una especie particular, por ejemplo, un genoma euploide (en humanos, 46,XX o 46,XY). Dado que los diferentes organismos tienen complementos cromosómicos ampliamente variables, el término ‘‘aneuploidía’’ no se refiere a un número particular de cromosomas, sino más bien a la situación en la que el contenido cromosómico dentro de una célula o células dadas de un organismo es anómalo. En algunas implementaciones, el término ‘‘aneuploidía’’ en el presente documento se refiere a un desequilibrio de material genético provocado por una pérdida o ganancia de un cromosoma completo o parte de un cromosoma. Un ‘‘aneuploidía’’ puede referirse a una o más deleciones y/o inserciones de un segmento de un cromosoma. El término ‘‘euploide’’, en algunas implementaciones, se refiere a un complemento normal de cromosomas.

El término ‘‘monosomía’’, tal como se usa en el presente documento, se refiere a la falta de un cromosoma del complemento normal. La monosomía parcial puede producirse en translocaciones o deleciones desequilibradas, en las cuales solo un segmento del cromosoma está presente en una sola copia. La monosomía de cromosomas sexuales (45, X) provoca el síndrome de Turner, por ejemplo. El término ‘‘disomía’’ se refiere a la presencia de dos copias de un cromosoma. Para organismos tales como seres humanos que tienen dos copias de cada cromosoma (aquellos que son diploides o ‘‘euploides’’), la disomía es la condición normal. Para 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 contribución del otro progenitor).

El término ‘‘trisomía’’, tal 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 extra, que se encuentra en el síndrome de Down humano, se denomina ‘‘trisomía 21’’. La trisomía 18 y la trisomía 13 son otras dos trisomías autosómicas humanas. La trisomía de los cromosomas sexuales puede observarse en mujeres (por ejemplo, 47, XXX en el síndrome triple X) o en hombres (por ejemplo, 47, XXY en el síndrome de Klinefelter; o 47, XYY en el síndrome de Jacobs). En algunas implementaciones, una trisomía es una duplicación de la mayor parte o la totalidad de un autosoma. En determinadas implementaciones, una trisomía es una aneuploidía cromosómica completa que produce tres casos (por ejemplo, tres copias) de un tipo particular de cromosoma (por ejemplo, en lugar de dos instancias (es decir, un par) de un tipo particular de cromosoma para un euploide).

Los términos ‘‘tetrasomía’’ y ‘‘pentasomía’’, tal 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 ven con autosomas, se ha notificado tetrasomía y pentasomía en cromosomas sexuales en seres humanos, incluyendo XXXX, XXXY, XXYY, XYYY, XXXXX, XXXXY, XXXYY, XXYYY y XYYYY.

Las anomalías cromosómicas pueden estar provocadas 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 provocan ausencia de disyunción en múltiples cromosomas, (iii) unión merotélica que se produce cuando un cinetocoro se une a ambos polos del huso mitótico, (iv) una formación de huso multipolar cuando se forman más de dos polos de huso, (v) una formación de huso monopolar cuando se forma solamente un único polo de huso, y (vi) un producto intermedio tetraploide que se produce como resultado final del mecanismo de huso monopolar.

Las expresiones ‘‘monosomía parcial’’ y ‘‘trisomía parcial’’, tal como se usan en el presente documento, se refieren a un desequilibrio de material genético provocado por la pérdida o ganancia de parte de un cromosoma. Una monosomía parcial o trisomía parcial puede resultar de una translocación desequilibrada, en 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 implicado en el cromosoma derivado.

El término ‘‘mosaicismo’’, tal como se usa en el presente documento, se refiere a aneuploidía en algunas células, pero no en todas las células, de un organismo. Determinadas anomalías cromosómicas pueden existir como anomalías cromosómicas con mosaicismo y sin mosaicismo. Por ejemplo, determinados individuos con trisomía 21

tienen síndrome de Down y algunos tienen síndrome de Down sin mosaicismo. 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, uno de los cromosomas 21 se duplicó. El mosaicismo somático se produce probablemente a través de mecanismos distintos de aquellos normalmente asociados con síndromes genéticos que involucran aneuploidía completa o con mosaicismo. El mosaicismo somático se ha identificado en determinados tipos de cánceres y en neuronas, por ejemplo. En determinados casos, la trisomía 12 se ha identificado en la leucemia linfocítica crónica (LLC) y la trisomía 8 se ha identificado en la leucemia mieloide aguda (LMA). Además, los síndromes genéticos en los que un individuo está predispuesto a rotura de cromosomas (síndromes de inestabilidad cromosómica) se asocian frecuentemente con un mayor riesgo de diversos tipos de cáncer, destacando así 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 sin mosaicismo y con mosaicismo.

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

Tabla 1A

Cromosoma	Anomalía	Asociación con la enfermedad
X	XO	Síndrome de Turner
Y	XXY	Síndrome de Klinefelter
Y	XYY	Síndrome doble Y
Y	XXX	Síndrome de trisomía X
Y	XXXX	Síndrome tetra-X
Y	deleción de Xp21	Síndrome de Duchenne/Becker, hipoplasia suprarrenal congénita, enfermedad granulomatosa crónica
Y	deleción de Xp22	deficiencia de esteroide sulfatasa
Y	deleción de YXq26	Enfermedad linfoproliferativa ligada al cromosoma X
1	monosomía-trisomía (somática) 1p	neuroblastoma
2	monosomía-trisomía 2q	retardo del crecimiento, retraso del desarrollo y mental, y anomalías físicas menores
3	monosomía-trisomía (somática)	Linfoma no hodgkiniano
4	monosomía-trisomía (somática)	Leucemia no linfocítica aguda (LNLA)
5	5p	Maullido de gato; síndrome de Lejeune
5	monosomía-trisomía (somática) 5q	síndrome mielodisplásico
6	monosomía-trisomía (somática)	sarcoma de células claras
7	deleción de 7q11.23	Síndrome de William
7	monosomía-trisomía	síndrome de monosomía 7 de la infancia; somáticos: adenomas corticales renales; síndrome mielodisplásico
8	deleción de 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 Rethore
9	trisomía	síndrome de trisomía 9 completa; síndrome de trisomía 9 con mosaicismo
10	Monosomía-trisomía (somática)	LLA o LNLA
11	11p-	Aniridia; tumor de Wilms
11	11q-	Síndrome de Jacobsen
11	monosomía (somática)-trisomía	linajes mieloides afectados (LNLA, SMD)
12	monosomía-trisomía (somática)	LLC, tumor de células de la granulosa juvenil (TCGJ)
13	13q-	síndrome 13q; síndrome de Orbeli
13	deleción de 13q14	retinoblastoma
13	monosomía-trisomía	Síndrome de Patau
14	monosomía-trisomía (somática)	trastornos mieloides (SMD, LNLA, LMC atípica)
15	monosomía con deleción de 15q11-q13	Prader-Willi, síndrome de Angelman

15	trisomía (somática)	linajes mieloides y linfoides afectados (por ejemplo, SMD, LNLA, LLA, LLC)
16	deleción de 16q13.3	Rubenstein-Taybi
3	monosomía-trisomía (somática)	carcinomas de células renales papilares (malignos)
17	17p-(somático)	Síndrome p17 en neoplasias malignas mieloides
17	deleción de 17q11.2	Smith-Magenis
17	17q13.3	Miller-Dieker
17	monosomía-trisomía (somática)	adenomas corticales renales
17	trisomía 17p11.2-12	Síndrome 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	deleción de 20p11.2-12	Alagille
20	20p-	somático: SMD, LNLA, policitemia vera, leucemia neutrofílica crónica
20	monosomía-trisomía (somática)	carcinomas de células renales papilares (malignos)
21	monosomía-trisomía	Síndrome de Down
22	deleción de 22q11.2	Síndrome de DiGeorge, síndrome velocardiofacial, síndrome de anomalías conotruncales y faciales, síndrome de Opitz G/BBB autosómico dominante, síndrome cardiofacial de Caylor
22	monosomía-trisomía	síndrome de trisomía 22 completa

Tabla 1B

Síndrome	**Cromosoma**	**Inicio**	**Fin**	**Intervalo (Mb)**	**Grado**
síndrome de microdeleción de 12q14	12	65.071.919	68.645.525	3,57	
síndrome de microdeleción de 15q13.3	15	30.769.995	32.701.482	1,93	
15q24 síndrome de microdeleción recurrente	15	74.377.174	76.162.277	1,79	
síndrome de sobrecrecimiento de 15q26	15	99.357.970	102.521.392	3,16	
síndrome de microdeleción de 16p11.2	16	29.501.198	30.202.572	0,70	
síndrome de microdeleción de 16p11.2-p12.2	16	21.613.956	29.042.192	7,43	
microdeleción recurrente de 16p13.11 (locus de susceptibilidad a trastornos neurocognitivos)	16	15.504.454	16.284.248	0,78	
microduplicación recurrente de 16p13.11 (locus de susceptibilidad a trastornos neurocognitivos)	16	15.504.454	16.284.248	0,78	
síndrome de microdeleción recurrente de 17q21.3	17	43.632.466	44.210.205	0,58	1
síndrome de microdeleción de 1p36	1	10.001	5.408.761	5,40	1
microdeleción recurrente de 1q21.1 (locus de susceptibilidad a trastornos del desarrollo neurológico)	1	146.512.930	147.737.500	1,22	3
microduplicación recurrente de 1q21.1 (posible locus de susceptibilidad a trastornos del desarrollo neurológico)	1	146.512.930	147.737.500	1,22	3
locus de susceptibilidad 1q21.1 para la trombocitopenia-síndrome de radio ausente (TAR)	1	145.401.253	145.928.123	0,53	3
síndrome de deleción de 22q11 (síndrome velocardiofacial / DiGeorge)	22	18.546.349	22.336.469	3,79	1
síndrome de duplicación de 22q11	22	18.546.349	22.336.469	3,79	3
síndrome de deleción distal de 22q11.2	22	22.115.848	23.696.229	1,58	

síndrome de deleción de 22q13 (síndrome de Phelan-Mcdermid)	22	51.045.516	51.187.844	0,14	1
síndrome de microdeleción de 2p15-16.1	2	57.741.796	61.738.334	4,00	
síndrome					
síndrome de deleción de 2q33.1	2	196.925.089	205.206.940	8,28	1
monosomía de 2q37	2	239.954.693	243.102.476	3,15	1
síndrome de microdeleción de 3q29	3	195.672.229	197.497.869	1,83	
síndrome de microduplicación de 3q29	3	195.672.229	197.497.869	1,83	
síndrome de duplicación de 7q11.23	7	72.332.743	74.616.901	2,28	
síndrome de deleción de 8p23.1	8	8.119.295	11.765.719	3,65	
síndrome de deleción subtelomérica de 9q	9	140.403.363	141.153.431	0,75	
Leucodistrofia autosómica dominante de inicio en la edad adulta (ADLD)	5	126.063.045	126.204.952	0,14	1
Síndrome de Angelman (tipo 1)	15	22.876.632	28.557.186	5,68	
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	1
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 del 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	
Síndrome de maullido de gato (deleción de 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	1
Poliposis adenomatosa familiar	5	112.101.596	112.221.377	0,12	
Predisposición genética a las parálisis por presión (HNPP)	17	13.968.607	15.434.038	1,47	
Leri-Weill	X	751.878	867.875	0,12	1
Discondrostosis (LWD) – SHOX deleción					
Leri-Weill discondrostosis (LWD) – SHOX deleción	X	460.558	753.877	0,29	
Miller-Dieker síndrome (SMD)	17	1	2.545.429	2,55	1
microdeleción de NF1 síndrome	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 de 17p11.2)	17	16.706.021	20.482.061	3,78	
Potocki-Shaffer síndrome	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
Síndrome de 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	
Síndrome de Rubenstein-Taybi	16	3.781.464	3.861.246	0,08	1
Síndrome de 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 dividido (SHFM1)	7	95.533.860	96.779.486	1,25	
Deficiencia de esteroide sulfatasa (STS)	X	6.441.957	8.167.697	1,73	
WAGR 11p13 síndrome de deleción	11	31.803.509	32.510.988	0,71	
Williams Beuren Síndrome (WBS)	7	72.332.743	74.616.901	2,28	1
Síndrome de Wolf-Hirschhorn	4	10.001	2.073.670	2,06	1
Duplicación de Xq28 (MECP2)	X	152.749.900	153.390.999	0,64	

Las afecciones de grado 1 a menudo tienen una o más de las siguientes características; anomalía patógena; fuerte acuerdo entre los genetistas; altamente penetrante; todavía puede tener un fenotipo variable pero algunas características comunes; todos los casos en la bibliografía tienen un fenotipo clínico; ningún caso de individuos sanos con la anomalía; no informado en las bases de datos de la DVG ni encontrado en población sana; datos funcionales que confirman el

efecto de dosificación de un solo gen o de múltiples genes; genes candidatos confirmados o fuertes; implicaciones de gestión clínica definidas; riesgo de cáncer conocido con implicaciones para la vigilancia; múltiples fuentes de información (OMIM, Genereviews, Orphanet, Unique, Wikipedia); y/o disponible para uso diagnóstico (asesoramiento reproductivo).

Las afecciones de grado 2 a menudo tienen una o más de las siguientes características; probable anomalía patógena; altamente penetrante; fenotipo variable sin características consistentes además de DD; pequeño número de casos/informes en la bibliografía; todos los casos informados tienen un fenotipo clínico; sin datos funcionales ni genes patógenos confirmados; múltiples fuentes de información (OMIM, Genereviews, Orphanet, Unique, Wikipedia); y/o puede usarse con fines de diagnóstico y asesoramiento reproductivo.

Las afecciones de grado 3 a menudo tienen una o más de las siguientes características; locus de susceptibilidad; individuos sanos o padres no afectados de un índice descrito; presente en poblaciones de control; no penetrante; fenotipo leve e inespecífico; características menos consistentes; sin 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 presenta un hallazgo clínico novedoso; y/o precaución cuando se usa con fines de diagnóstico y consejos cautelosos para el asesoramiento reproductivo.

Preeclampsia

En algunas implementaciones, la presencia o ausencia de preeclampsia se determina usando un método, una máquina o un aparato descrito en el presente documento. La preeclampsia es una afección en la que surge hipertensión durante el embarazo (es decir, hipertensión inducida por el embarazo) y se asocia con cantidades significativas de proteína en la orina. En determinadas implementaciones, la preeclampsia se asocia además 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 hipermetilado derivado del feto extracelular y la gravedad de la preeclampsia. En determinados ejemplos, se observa una mayor metilación del ADN para el gen H19 en la placenta con preeclampsia en comparación con los controles normales.

La preeclampsia es una de las causas principales de morbimortalidad materna y fetal/neonatal en todo el mundo. Los ácidos nucleicos circulantes, libres de células en plasma y suero son biomarcadores novedosos con aplicaciones clínicas prometedoras en diferentes campos médicos, incluyendo el diagnóstico prenatal. Se han notificados cambios cuantitativos del ADN fetal libre de células (flc) en plasma materno como indicador para la preeclampsia inminente en diferentes estudios, por ejemplo, usando PCR cuantitativa en tiempo real para los loci SRA o DYS 14 específicos del sexo masculino. En los casos de preeclampsia de inicio temprano, pueden observarse niveles elevados en el primer trimestre. El aumento de los niveles de ADNflc antes del inicio de los síntomas puede deberse a hipoxia/reoxigenación dentro del espacio intervelloso que conduce al estrés oxidativo tisular y al aumento de la apoptosis y necrosis placentaria. Además de la evidencia de mayor descarga de ADNflc en la circulación materna, también existe evidencia de menor aclaramiento renal de ADNflc en la preeclampsia. Dado que la cantidad de ADN fetal se determina actualmente mediante la cuantificación de secuencias específicas del cromosoma Y, los enfoques alternativos, tales como la medición del ADN libre de células total o el uso de marcadores epigenéticos fetales independientes del sexo, tales como la metilación del ADN, ofrecen una alternativa. EL ARN libre de células de origen placentario es otro biomarcador alternativo que puede usarse para analizar y diagnosticar preeclampsia en la práctica clínica. EL ARN fetal se asocia con partículas placentarias subcelulares que lo protegen frente a la degradación. Los niveles de ARN fetal a veces son diez veces mayores en mujeres embarazadas con preeclampsia en comparación con los controles y, por tanto, es un biomarcador alternativo que puede usarse para analizar y diagnosticar preeclampsia en la práctica clínica.

Patógenos

En algunas implementaciones, la presencia o ausencia de una afección patógena se determina mediante un método, una máquina o un aparato descrito en el presente documento. Una afección patógena puede estar provocada por la infección de un huésped por un patógeno incluyendo, pero sin limitarse a, una bacteria, un virus u hongo. Dado que los patógenos poseen normalmente ácido nucleico (por ejemplo, ADN genómico, ARN genómico, ARNm) que puede distinguirse del ácido nucleico del huésped, los métodos, las máquinas y los aparatos proporcionados en el presente documento pueden usarse para determinar la presencia o ausencia de un patógeno. A menudo, los patógenos poseen ácido nucleico con características únicas para un patógeno particular, tal como, por ejemplo, estado epigenético y/o una o más variaciones, duplicaciones y/o deleciones de secuencias. Por tanto, los métodos proporcionados en el presente documento pueden usarse para identificar un patógeno particular o variante de patógenos (por ejemplo, cepa).

Cánceres

En algunas implementaciones, la presencia o ausencia de un trastorno de proliferación celular (por ejemplo, un cáncer) se determina usando un método, una máquina o un aparato descrito en el presente documento. Por ejemplo, los niveles de ácido nucleico libre de células en suero pueden elevarse en pacientes con diversos tipos de cáncer en comparación con pacientes sanos. Los pacientes con enfermedades metastásicas, por ejemplo, algunas veces pueden tener niveles de ADN en suero aproximadamente dos veces más altos que los pacientes no metastásicos. Los pacientes con enfermedades metastásicas pueden identificarse además por marcadores específicos de cáncer y/o determinados

polimorfismos de un solo nucleótido o repeticiones cortas en tándem, por ejemplo. Los ejemplos no limitativos de tipos de cáncer que pueden correlacionarse positivamente con niveles elevados de ADN circulante incluyen cáncer de mama, cáncer colorrectal, cáncer gastrointestinal, cáncer hepatocelular, cáncer de pulmón, melanoma, linfoma no hodgkiniano, leucemia, mieloma múltiple, cáncer de vejiga, hepatoma, cáncer cervicouterino, cáncer de esófago, cáncer de páncreas y cáncer de próstata. Diversos cánceres pueden tener, y algunas veces pueden liberarse en el torrente sanguíneo, ácidos nucleicos con características que son distinguibles de los ácidos nucleicos de las células sanas no cancerosas, tales como, por ejemplo, el estado epigenético y/o variaciones, duplicaciones y/o deleciones de secuencia. Tales características pueden ser, por ejemplo, específicas para un tipo particular de cáncer. Por tanto, se contempla además que un método proporcionado en el presente documento puede usarse para identificar un tipo particular de cáncer.

Se puede usar software para realizar una o más etapas en los procesos descritos en el presente documento, incluyendo, pero sin limitación; contar, procesar datos, generar un resultado y/o proporcionar una o más recomendaciones basadas en los resultados generados, como se describe con mayor detalle de aquí en adelante.

Máquinas, software e interfaces

Determinados procedimientos y métodos descritos en el presente documento (por ejemplo, cuantificación, mapeo, normalización, establecimiento de rango, ajuste, categorización, recuento y/o determinación de lecturas de secuencia, recuentos, elevaciones o niveles (por ejemplo, elevaciones o niveles) y/o perfiles) no pueden realizarse a menudo sin un ordenador, procesador, software, módulo u otro aparato. Los métodos descritos en el presente documento normalmente son métodos implementados por ordenador, y una o más porciones de un método a veces las realizan uno o más procesadores (por ejemplo, microprocesadores), ordenadores o aparatos controlados por microprocesadores. Las implementaciones relacionadas con los métodos descritos en este documento son aplicables generalmente al mismo procedimiento o a procedimientos relacionados implementados por instrucciones en los sistemas, aparatos y productos de programa informático descritos en el presente documento. En algunas implementaciones, los procedimientos y métodos descritos en el presente documento (por ejemplo, cuantificación, recuento y/o determinación de lecturas de secuencia, recuentos, elevaciones o niveles y/o perfiles de secuencia) se realizan mediante métodos automatizados. En algunas implementaciones, una o más etapas y un método descrito en el presente documento se llevan a cabo mediante un procesador y/o un ordenador, y/o se llevan a cabo junto con la memoria. En algunas implementaciones, un método automatizado se incorpora en software, módulos, procesadores, periféricos y/o un aparato y/o una máquina que comprende similares, que determinan lecturas de secuencia, recuentos, mapeo, etiquetas de secuencia mapeadas, elevaciones o niveles, perfiles, normalizaciones, comparaciones, establecimiento de rango, categorización, ajustes, representación gráfica, resultados, transformaciones e identificaciones. Tal como se usa en el presente documento, software se refiere a instrucciones de programas legibles por ordenador que, cuando se ejecutan por un procesador, realizan operaciones informáticas, tal como se describe en el presente documento.

Las lecturas, recuentos, elevaciones o niveles y perfiles de secuencias derivados de un sujeto de prueba (por ejemplo, un paciente, una mujer embarazada) y/o de un sujeto de referencia pueden analizarse y procesarse adicionalmente para determinar la presencia o ausencia de una variación genética. Las lecturas, recuentos, elevaciones o niveles y/o perfiles de secuencias se denominan, algunas veces, “datos” o “conjuntos de datos”. En algunas implementaciones, los datos o conjuntos de datos pueden caracterizarse por una o más características o variables (por ejemplo, basadas en secuencia [por ejemplo, contenido de GC, secuencia de nucleótidos específica, similares], específicas de función [por ejemplo, genes expresados, genes de cáncer, similares], basadas en la ubicación [específicas del genoma, específicas del cromosoma, específicas de un bin o de una porción], similares y combinaciones de los mismos). En determinadas implementaciones, los datos o conjuntos de datos pueden organizarse en una matriz que tiene dos o más dimensiones basándose en una o más características o variables. Los datos organizados en matrices pueden organizarse usando cualquier característica o variable adecuada. Un ejemplo no limitativo de datos en una matriz incluye datos organizados por edad materna, ploidía materna y contribución fetal. En determinadas implementaciones, los conjuntos de datos caracterizados por una o más características o variables a veces se procesan después del recuento.

Pueden usarse aparatos, software e interfaces para llevar a cabo los métodos descritos en el presente documento. Con el uso de aparatos, programas e interfaces, un usuario puede introducir, solicitar, consultar o determinar opciones para usar información, programas o procedimientos particulares (por ejemplo, mapear lecturas de secuencia, procesar datos mapeados y/o proporcionar un resultado), que puede implicar implementar algoritmos de análisis estadístico, algoritmos de significación estadística, algoritmos estadísticos, etapas iterativas, algoritmos de validación y representaciones gráficas, por ejemplo. En algunas implementaciones, un usuario puede introducir un conjunto de datos como información de entrada, un usuario puede descargar uno o más conjuntos de datos mediante un medio de *hardware* adecuado (por ejemplo, unidad flash) y/o un usuario puede enviar un conjunto de datos de un sistema a otro para el procesamiento posterior y/o proporcionar un resultado (por ejemplo, enviar datos de lectura de secuencia de 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 producir un resultado y/o informe).

Un sistema comprende normalmente uno o más aparatos. Cada aparato comprende uno o más de memoria, uno o más procesadores e instrucciones. Cuando un sistema incluye dos o más aparatos, algunos o todos los aparatos pueden estar ubicados en la misma ubicación, algunos o todos los aparatos pueden estar ubicados en ubicaciones diferentes, todos los aparatos pueden estar ubicados en una ubicación y/o todos los aparatos pueden estar ubicados

en ubicaciones diferentes. Cuando un sistema incluye dos o más aparatos, algunos o todos los aparatos pueden estar ubicados en la misma ubicación que un usuario, algunos o todos los aparatos pueden estar ubicados en una ubicación distinta a un usuario, todos los aparatos pueden estar ubicados en la misma ubicación que el usuario y/o todos los aparatos pueden estar ubicados en una o más ubicaciones diferentes al usuario.

Algunas veces, un sistema comprende un aparato computarizado y un aparato de secuenciación, en el que el aparato de secuenciación está configurado para recibir ácido nucleico físico y generar lecturas de secuencia, y el aparato computarizado está configurado para procesar las lecturas del aparato de secuenciación. Algunas veces, el aparato computarizado está configurado para determinar la presencia o ausencia de una variación genética (por ejemplo, variación del número de copias; aneuploidía cromosómica fetal) de las lecturas de secuencia.

Por ejemplo, un usuario puede colocar una consulta en un software que, después, puede adquirir un conjunto de datos por medio de acceso a Internet y, en determinadas implementaciones, puede solicitarse a un procesador programable que adquiera un conjunto de datos adecuado basándose en parámetros dados. Un procesador programable también puede solicitar a un usuario que seleccione una o más opciones de conjuntos de datos seleccionadas por el procesador basándose en parámetros dados. Un procesador programable puede solicitar a un usuario que seleccione una o más opciones de conjuntos de datos seleccionadas por el procesador basándose en la información que se encuentra a través de Internet, otra información interna o externa o similares. Las opciones pueden elegirse 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, etapas iterativas, uno o más algoritmos de validación y una o más representaciones gráficas de métodos, aparatos o programas informáticos.

Los sistemas abordados en el presente documento pueden comprender componentes generales de sistemas informáticos tales como, por ejemplo, servidores de red, sistemas portátiles, sistemas de escritorio, sistemas de mano, asistentes digitales personales, quioscos informáticos y similares. Un sistema informático puede comprender uno o más medios de entrada tales como un teclado, una pantalla táctil, un ratón, reconocimiento de voz u otros medios para permitir que el usuario introduzca datos en el sistema. Un sistema puede comprender además una o más salidas, incluyendo, aunque no de forma limitativa, una pantalla de visualización (por ejemplo, CRT o LCD), un altavoz, una máquina de fax, impresora (por ejemplo, impresora láser, de chorro de tinta, impacto, en blanco y negro o a color) u otra salida útil para proporcionar una salida visual, auditiva y/o impresa de información (por ejemplo, resultado y/o informe).

En un sistema, los medios de entrada y salida pueden conectarse a una unidad central de procesamiento que puede comprender entre otros componentes, un microprocesador para ejecutar instrucciones de programa y memoria para almacenar código de programa y datos. En algunas implementaciones, los procedimientos pueden implementarse como un solo sistema de usuario ubicado en un solo lugar geográfico. En determinadas implementaciones, los procedimientos pueden implementarse como un sistema multiusuario. En el caso de una implementación multiusuario, múltiples unidades centrales de procesamiento pueden conectarse por medio de una red. La red puede ser local, que abarca un único departamento en una parte de un edificio, todo un edificio, abarcar múltiples edificios, abarcar una región, abarcar todo un país o ser mundial. La red puede ser privada, ser propiedad de y estar controlada por un proveedor, o puede implementarse como un servicio basado en Internet en donde el usuario accede a una página web para introducir y recuperar información. En consecuencia, en determinadas implementaciones, 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 ubicación o en múltiples ubicaciones, y los datos pueden mapearse y/o procesarse en serie y/o en paralelo. Por tanto, puede usarse una configuración y un control adecuados para mapear y/o procesar datos usando múltiples máquinas, tales como en redes locales, redes remotas y/o plataformas informáticas de tipo “nube”.

Un sistema puede incluir una interfaz de comunicaciones en algunas implementaciones. Una interfaz de comunicaciones permite la transferencia de software y datos entre un sistema informático y uno o más dispositivos externos. Los ejemplos no limitativos de interfaces de comunicaciones incluyen un módem, una interfaz de red (tal como una tarjeta Ethernet), un puerto de comunicaciones, una ranura y tarjeta PCMCIA, y similares. El software y los datos transferidos por medio de una interfaz de comunicaciones generalmente están en forma de señales, que pueden ser señales electrónicas, electromagnéticas, ópticas y/u otras señales capaces de recibirse por una interfaz de comunicaciones. A menudo, se proporcionan señales a una interfaz de comunicaciones mediante un canal. Un canal transporta a menudo señales y puede implementarse usando hilo o cable, fibra óptica, una línea telefónica, un enlace de teléfono celular, un enlace de RF y/u otros canales de comunicaciones. Por tanto, en un ejemplo, puede usarse una interfaz de comunicaciones para recibir información de señal que puede detectarse por un módulo de detección de señal.

Los datos pueden introducirse mediante un dispositivo y/o método adecuado incluyendo, pero sin limitarse a, dispositivos de entrada manual o dispositivos de entrada de datos directa (DDE). Los ejemplos no limitativos de dispositivos manuales incluyen teclados, teclados de concepto, pantallas táctiles, lápices ópticos, ratón, bolas de rastreo, palancas de mando, tabletas gráficas, escáneres, cámaras digitales, digitalizadores de vídeo y dispositivos de reconocimiento de voz. Los ejemplos no limitativos de DDE incluyen lectores de código de barras, códigos de tira magnética, tarjetas inteligentes, reconocimiento de caracteres de tinta magnética, reconocimiento de caracteres ópticos, reconocimiento de marcas ópticas y documentos de respuesta.

En algunas implementaciones, la salida de un aparato de secuenciación puede servir como datos que pueden introducirse a través de un dispositivo de entrada. En determinadas implementaciones, las lecturas de secuencia mapeadas pueden servir como datos que pueden introducirse a través de un dispositivo de entrada. En determinadas implementaciones, se generan datos simulados mediante un procedimiento *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 la investigación y los experimentos realizados con el uso de un ordenador. Los procedimientos *in silico* incluyen, pero no se limitan a, mapeo de lecturas de secuencia y procesamiento de lecturas de secuencia mapeadas según los procedimientos descritos en la presente invención.

Un sistema puede incluir un software útil para realizar un procedimiento descrito en el presente documento, y el software puede incluir uno o más módulos para realizar tales procedimientos (por ejemplo, módulo de secuenciación, módulo de procesamiento lógico, módulo de organización de visualización de datos). El término "software" se refiere a instrucciones de programas legibles por ordenador que, cuando se ejecutan por un ordenador, realizan operaciones informáticas. Las instrucciones ejecutables por el uno o más procesadores a veces se proporcionan como código ejecutable, que cuando se ejecutan, pueden 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, e instrucciones (por ejemplo, procesos, rutinas, subrutinas) incorporadas en el software pueden implementarse o realizarse por un procesador. Por ejemplo, un módulo (por ejemplo, un módulo de software) puede formar parte de un programa que realiza un procedimiento o tarea particular. El término "módulo" se refiere a una unidad funcional autónoma que puede usarse en un aparato o sistema de software más grande. Un módulo puede comprender un conjunto de instrucciones para llevar a cabo una función del módulo. Un módulo puede transformar información y/o datos. La información y/o los datos pueden estar en una forma adecuada. Por ejemplo, la información y/o los datos pueden ser digitales o analógicos. En determinadas implementaciones, la información y/o los datos pueden ser paquetes, bytes, caracteres o bits. En algunas implementaciones, la información y/o los datos pueden ser cualquier información o dato recopilado, ensamblado o utilizable. Los ejemplos no limitativos de información y/o datos incluyen un medio adecuado, imágenes, vídeo, sonido (por ejemplo, frecuencias, audibles o no audibles), números, constantes, un valor, objetos, tiempo, funciones, instrucciones, mapas, referencias, secuencias, lecturas, lecturas mapeadas, elevaciones o niveles, rangos, umbrales, señales, visualizaciones, representaciones o transformaciones de los mismos. Un módulo puede aceptar o recibir información y/o datos, transformar la información y/o los datos en una segunda forma y proporcionar o transferir la segunda forma a un aparato o una máquina, periférico, componente u otro módulo. Un módulo puede realizar una o más de las siguientes funciones no limitativas: mapear lecturas de secuencia, proporcionar recuentos, ensamblar porciones, proporcionar o determinar una elevación o un nivel, proporcionar un perfil de recuento, normalizar (por ejemplo, normalizar lecturas, normalizar recuentos, y similares), proporcionar un perfil de recuento normalizado o elevaciones o niveles de recuentos normalizados, comparar dos o más elevaciones o niveles, proporcionar una medida de incertidumbre, proporcionar o determinar elevaciones o niveles esperados y rangos esperados (por ejemplo, rangos de elevación o nivel esperados, rangos umbral y elevaciones umbral), proporcionar ajustes a las elevaciones o los niveles (por ejemplo, ajustar una primera elevación o nivel, ajustar una segunda elevación o nivel, ajustar un perfil de un cromosoma o un segmento del mismo y/o relleno), proporcionar identificación (por ejemplo, identificar una variación del número de copias, variación genética o aneuploidía), categorizar, representar gráficamente y/o determinar un resultado, por ejemplo. Un procesador puede, en determinadas implementaciones, llevar a cabo las instrucciones en un módulo. En algunas implementaciones, se requiere que uno o más procesadores lleven a cabo instrucciones en un módulo o grupo de módulos. Un módulo puede proporcionar información y/o datos a otro módulo, aparato o fuente y puede recibir información y/o datos de otro módulo, aparato o fuente.

Un medio de almacenamiento legible por ordenador no transitorio a veces comprende un programa ejecutable almacenado en el mismo y, a veces, el programa instruye a un microprocesador para que realice una función (por ejemplo, un método descrito en el presente documento). Algunas veces, un producto de programa informático se incorpora en un medio legible por ordenador tangible y, algunas veces, se incorpora en un medio legible por ordenador no transitorio. Algunas veces, un módulo se almacena en un medio legible por ordenador (por ejemplo, disco, unidad) o en memoria (por ejemplo, memoria de acceso aleatorio). Un módulo y procesador capaces de implementar instrucciones de un módulo pueden estar ubicados en un aparato o una máquina o en diferentes aparatos. Un módulo y/o procesador capaces de implementar una instrucción para un módulo pueden estar ubicados en la misma ubicación que un usuario (por ejemplo, red local) o en una ubicación diferente de un usuario (por ejemplo, red remota, sistema de tipo nube). En implementaciones en las que se lleva a cabo un método junto con dos o más módulos, los módulos pueden estar ubicados en el mismo aparato, uno o más módulos pueden estar ubicados en diferentes aparatos en la misma ubicación física, y uno o más módulos pueden estar ubicados en diferentes aparatos en ubicaciones físicas diferentes.

Un aparato o una máquina, en algunas implementaciones, comprende al menos un procesador para llevar a cabo las instrucciones en un módulo. A veces, se accede a los recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia por un procesador que ejecuta instrucciones configuradas para llevar a cabo un método descrito en el presente documento. Los recuentos a los que se accede por un procesador pueden estar dentro de la memoria de un sistema, y puede accederse a los recuentos y colocarse en la memoria del sistema después de que se obtienen. En algunas implementaciones, un aparato o una máquina incluye un procesador (por ejemplo, uno o más procesadores), procesador que puede realizar y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) de un módulo. En algunas implementaciones, un aparato o una máquina incluye múltiples procesadores, tales como procesadores coordinados y que funcionan en paralelo. En algunas implementaciones, un aparato o una máquina funciona con uno o más

procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)). En algunas implementaciones, un aparato o una máquina comprende un módulo. En determinadas implementaciones, un aparato o una máquina comprende uno o varios módulos. Un aparato o una máquina que comprende un módulo puede recibir y transferir a menudo uno o más de información y/o datos a y desde otros módulos. En determinadas implementaciones, un aparato o una máquina comprende periféricos y/o componentes. En determinadas implementaciones, un aparato o una máquina puede comprender uno o más periféricos o componentes que pueden transferir datos y/o información a y desde otros módulos, periféricos y/o componentes. En determinadas implementaciones un aparato o una máquina interactúa con un periférico y/o componente que proporciona datos y/o información. En determinadas implementaciones, los periféricos y componentes ayudan a un aparato o una máquina a llevar a cabo una función o interactúan directamente con un módulo. Los ejemplos no limitativos de periféricos y/o componentes incluyen un periférico de ordenador, método o dispositivo de almacenamiento o E/S adecuados incluyendo, pero sin limitarse a, escáneres, impresoras, pantallas de visualización (por ejemplo, monitores, LED, LCT o CRT), cámaras, micrófonos, paneles (por ejemplo, ipad, tabletas), pantallas táctiles, teléfonos inteligentes, teléfonos móviles, dispositivos de E/S USB, dispositivos de almacenamiento masivo USB, teclados, un ratón para ordenador, lápices digitales, módems, discos duros, memorias USB, unidades flash, un procesador, un servidor, CD, DVD, tarjetas gráficas, dispositivos de E/S especializados (por ejemplo, secuenciadores, fotoceldas, tubos fotomultiplicadores, lectores ópticos, sensores, etc.), una o más celdas de flujo, componentes de manipulación 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.

Uno o más de un módulo de secuenciación, un módulo de procesamiento lógico y un módulo de organización de visualización de datos pueden usarse en un método descrito en el presente documento. Los módulos a veces son controlados por un microprocesador. En determinadas implementaciones, un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos, o un aparato que comprende uno o más de tales módulos, recopilan, ensamblan, reciben, proporcionan y/o transfieren información y/o datos a o desde otro módulo, aparato, componente, periférico u operador de un aparato. Por ejemplo, algunas veces un operador de un aparato proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos. Un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos pueden recibir información y/o datos de otro módulo, los ejemplos no limitativos de los mismos incluyen un módulo de procesamiento lógico, módulo de secuenciación, módulo de organización de visualización de datos, módulo de secuenciación, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de normalización, módulo de comparación, módulo de establecimiento de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos y/o módulo de procesamiento lógico, similares o combinaciones de los mismos. La información y/o los datos derivados de o transformados por un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos pueden transferirse de un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos a un módulo de secuenciación, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de normalización, módulo de comparación, módulo de establecimiento de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico u otro aparato y/o módulo adecuado. Un módulo de secuenciación puede recibir información y/o datos de un módulo de procesamiento lógico y/o módulo de secuenciación y transferir información y/o datos a un módulo de procesamiento lógico y/o un módulo de mapeo, por ejemplo. En determinadas implementaciones, un módulo de procesamiento lógico orquesta, controla, limita, organiza, ordena, distribuye, divide, transforma y/o regula información y/o datos o la transferencia de información y/o datos a y desde uno o más de otros módulos, periféricos o dispositivos. Un módulo de organización de visualización de datos puede recibir información y/o datos de un módulo de procesamiento lógico y/o módulo de representación gráfica y transferir información y/o datos a un módulo de procesamiento lógico, módulo de representación gráfica, pantalla de visualización, periférico o dispositivo. Un aparato que comprende un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos puede comprender al menos un procesador. En algunas implementaciones, se proporcionan información y/o datos por un aparato que incluye un procesador (por ejemplo, uno o más procesadores), procesador que puede realizar y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) desde el módulo de procesamiento lógico, módulo de secuenciación y/o módulo de organización de visualización de datos. En algunas implementaciones, un módulo de procesamiento lógico, módulo de secuenciación o módulo de organización de visualización de datos funciona con uno o más procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)).

El *software* a menudo se proporciona en un producto de programa que contiene instrucciones de programa grabadas en un medio legible por ordenador, incluidos, aunque no de forma limitativa, 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, disquetes, similares y otros medios similares en los que se pueden grabar las instrucciones del programa. En la implementación en línea, un servidor y un sitio web mantenidos por una organización pueden estar configurados 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 obtiene o recibe de manera específica datos (por ejemplo, un módulo receptor de datos que recibe datos leídos

de secuencias y/o datos leídos mapeados) y puede incluir un módulo que procesa de manera específica los datos (por ejemplo, un módulo de procesamiento que procesa los 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 secuencia, lecturas mapeadas) mediante medios de comunicación por ordenador desde 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 la que se recibe, o puede generarse en una ubicación diferente y transmitirse a la ubicación de recepción. En algunas implementaciones, la información de entrada se modifica antes de procesarse (por ejemplo, se coloca en un formato susceptible al procesamiento (por ejemplo, tabulado)). En algunas implementaciones, en el presente documento, se proporcionan medios de almacenamiento legibles por ordenador no transitorios tales como, por ejemplo, un medio de almacenamiento legible por ordenador no transitorio que comprende un programa ejecutable almacenado en el mismo, donde el programa está configurado para (a) obtener lecturas de secuencia de ácido nucleico de muestra de un sujeto de prueba; (b) mapear las lecturas de secuencia obtenidas en (a) en un genoma conocido, cuyo genoma conocido se ha dividido en porciones; (c) contar las lecturas de secuencia mapeadas dentro de las porciones; (d) generar un perfil de recuento normalizado de muestra normalizando los recuentos de las porciones obtenidas en (c); y (e) determinar la presencia o ausencia de una variación genética a partir del perfil de recuento normalizado de la muestra en (d). En algunas implementaciones, en el presente documento, se proporcionan medios de almacenamiento no transitorios legibles por ordenador tales como, por ejemplo, un medio de almacenamiento no transitorio legible por ordenador que comprende un programa ejecutable almacenado en el mismo, en donde el programa instruye a un microprocesador para que realice lo siguiente: (a) acceder a lecturas de secuencia de nucleótidos mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba de una mujer embarazada, (b) determinar una o más estimaciones de la curvatura para la muestra de prueba a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) una característica de mapeo para las porciones del genoma de referencia y (c) calcular un nivel de sección genómica normalizado de cada una de las porciones del genoma de referencia para la muestra de prueba según (1) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la muestra de prueba, (2) la una o más estimaciones de la curvatura determinadas en (b) para la muestra de prueba, y (3) una o más estimaciones de la curvatura específicas de la porción de cada una de las múltiples porciones del genoma de referencia a partir de una relación ajustada entre (i) una o más estimaciones de la curvatura específicas de la muestra para una pluralidad de muestras, y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la pluralidad de muestras, configuradas de ese modo para proporcionar niveles de sección genómica calculados, donde el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de sección genómica calculados.

En algunas implementaciones, se proporcionan productos de programa informático, tales como, por ejemplo, un producto de programa informático que comprende un medio utilizable por ordenador que tiene un código de programa legible por ordenador incorporado en el mismo, estando el código de programa legible por ordenador adaptado para ejecutarse para implementar un método que comprende: (a) obtener lecturas de secuencia de ácido nucleico de muestra de un sujeto de prueba; (b) mapear las lecturas de secuencia obtenidas en (a) en un genoma conocido, cuyo genoma conocido se ha dividido en porciones; (c) contar las lecturas de secuencia mapeadas dentro de las porciones; (d) generar un perfil de recuento normalizado de muestras normalizando los recuentos de las porciones obtenidas en (c); y (e) determinar la presencia o ausencia de una variación genética a partir del perfil de recuento normalizado de la muestra en (d).

El software puede incluir uno o más algoritmos en determinadas implementaciones. Un algoritmo puede usarse para procesar datos y/o proporcionar un resultado o informe según una secuencia finita de instrucciones. Un algoritmo es a menudo una lista de instrucciones definidas para completar una tarea. Partiendo de un estado inicial, las instrucciones pueden describir un cálculo que avanza a través de una serie definida de estados sucesivos, terminando eventualmente en un estado 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 de búsqueda, algoritmo de clasificación, algoritmo de fusión, algoritmo numérico, algoritmo gráfico, algoritmo de cadena, algoritmo de modelado, algoritmo de genometría 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 funcionan en combinación. Un algoritmo puede ser de cualquier clase de complejidad adecuada y/o complejidad parametrizada. Puede usarse un algoritmo para el cálculo y/o procesamiento de datos y, en algunas implementaciones, puede usarse en un enfoque determinista o probabilístico/predictivo. Puede implementarse un algoritmo en un entorno informático usando un lenguaje de programación adecuado, son ejemplos no limitativos de los mismos C, C++, Java, Perl, Python, Fortran, y similares. En algunas implementaciones, un algoritmo puede configurarse o modificarse para incluir margen de errores, análisis estadístico, significación estadística y/o comparación con otra información o conjuntos de datos (por ejemplo, aplicable cuando se usa una red neural o algoritmo de agrupamiento).

En determinadas implementaciones, pueden implementarse varios algoritmos para su uso en software. Estos algoritmos pueden entrenarse con datos sin procesar en algunas implementaciones. Para cada nueva muestra de datos sin procesar, los algoritmos entrenados pueden producir un conjunto o resultado de datos procesados representativos. Un conjunto de datos procesados algunas veces tiene una complejidad reducida en comparación con el conjunto de datos original que se procesó. Basándose en un conjunto procesado, el rendimiento de un algoritmo

entrenado puede evaluarse basándose en la sensibilidad y especificidad, en algunas implementaciones. Un algoritmo con la mayor sensibilidad y/o especificidad puede identificarse y usarse, en determinadas implementaciones.

En determinadas implementaciones, los datos simulados (o simulación) pueden ayudar al procesamiento de datos, por ejemplo, entrenando un algoritmo o someter a prueba un algoritmo. En algunas implementaciones, los datos simulados incluyen varios muestreos hipotéticos de agrupamientos diferentes de lecturas de secuencia. Los datos simulados pueden basarse en lo que podría esperarse de una población real o pueden sesgarse para someter a prueba un algoritmo y/o asignar una clasificación correcta. Los datos simulados también se denominan en el presente documento datos "virtuales". Las simulaciones pueden realizarse mediante un programa informático en determinadas implementaciones. Una etapa posible en el uso de un conjunto de datos simulados es evaluar la confianza de un resultado identificado, por ejemplo, cuán bien coincide un muestreo aleatorio o representa mejor los datos originales. Un enfoque consiste en calcular un valor de probabilidad (valor de p) que estima la probabilidad de una muestra aleatoria con mejor puntuación que las muestras seleccionadas. En algunas implementaciones, puede evaluarse un modelo empírico, en el cual se supone que al menos una muestra coincide con una muestra de referencia (con o sin variaciones resueltas). En algunas implementaciones, otra distribución, tal como una distribución de Poisson, por ejemplo, puede usarse para definir la distribución de probabilidad.

Un sistema puede incluir uno o más procesadores en determinadas implementaciones. Un procesador puede conectarse a un bus de comunicaciones. Un sistema informático puede incluir una memoria principal, a menudo memoria de acceso aleatorio (RAM), y puede incluir además una memoria secundaria. La memoria en algunas implementaciones comprende un medio de almacenamiento no transitorio legible por ordenador. 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, tarjeta de memoria y similares. Una unidad de almacenamiento extraíble a menudo lee y/o escribe en una unidad de almacenamiento extraíble. Los ejemplos no limitativos de unidades de almacenamiento extraíbles incluyen un disquete, una cinta magnética, un disco óptico y similares, que pueden leerse y escribirse, por ejemplo, en una unidad de almacenamiento extraíble. Una unidad de almacenamiento extraíble puede incluir un medio de almacenamiento utilizable por ordenador que tiene almacenados un software y/o datos informáticos.

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

En algunas implementaciones, 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. Los ejemplos no limitativos de tales sistemas incluyen un cartucho de programa e interfaz de cartucho (tales como los que se encuentran en dispositivos de videojuegos), un chip de memoria extraíble (tal como una EPROM o PROM) y un enchufe asociado, y otras unidades de almacenamiento extraíbles e interfaces que permiten que el software y los datos se transfieran desde la unidad de almacenamiento extraíble a un sistema informático.

Una entidad puede generar recuentos de lecturas de secuencia, mapear las lecturas de secuencia en porciones, contar las lecturas mapeadas y utilizar las lecturas mapeadas contadas en un método, sistema, aparato, máquina o producto de programa informático descrito en el presente documento, en algunas implementaciones. Los recuentos de lecturas de secuencia mapeadas en porciones a veces se transfieren por una entidad a una segunda entidad para su uso por la segunda entidad en un método, sistema, aparato, máquina o producto de programa informático descrito en el presente documento, en determinadas implementaciones.

En algunas implementaciones, una entidad genera lecturas de secuencia y una segunda entidad mapea esas lecturas de secuencia en porciones en un genoma de referencia en algunas implementaciones. La segunda entidad cuenta, a veces, las lecturas mapeadas y utiliza las lecturas mapeadas contadas en un método, sistema, aparato, máquina o producto de programa informático descrito en el presente documento. En determinadas implementaciones, la segunda entidad transfiere las lecturas mapeadas a una tercera entidad, y la tercera entidad cuenta las lecturas mapeadas y utiliza las lecturas mapeadas en un método, sistema, aparato, máquina o producto de programa informático descrito en el presente documento. En determinadas implementaciones, 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, aparato, máquina o producto de programa informático descrito en el presente documento. En implementaciones que involucran una tercera entidad, la tercera entidad a veces es la misma que la primera entidad. Es decir, la primera entidad transfiere, a veces, lecturas de secuencia a una segunda entidad, segunda entidad que puede mapear lecturas de secuencia en 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, aparato, máquina o producto informático descrito en el presente documento, en donde la tercera entidad a veces es la misma que la primera entidad y, a veces, la tercera entidad es diferente de la primera o segunda entidad.

En algunas implementaciones, una entidad obtiene sangre de una mujer embarazada, opcionalmente, aísla ácido nucleico de la sangre (por ejemplo, del plasma o suero) y transfiere la sangre o ácido nucleico a una segunda entidad que genera lecturas de secuencia del ácido nucleico.

La FIG. 17 ilustra un ejemplo no limitativo de un entorno informático 510 en el que pueden implementarse 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. Tampoco debe interpretarse que el entorno informático 510 tiene alguna dependencia o requisito relacionado con uno o una combinación de componentes ilustrados en el entorno informático 510. En determinadas implementaciones, se puede utilizar un subconjunto de LOS sistemas, métodos y estructuras de datos que se muestran en la FIG. 17. 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 especial. Los ejemplos de sistemas, entornos y/o configuraciones informáticos conocidos que pueden ser adecuados incluyen, aunque no de forma limitativa, ordenadores personales, servidores, clientes ligeros, clientes pesados, dispositivos portátiles o manuales, sistemas multiprocesador, sistemas basados en microprocesador, decodificadores, electrónica de consumo programable, PC de red, miniordenadores, ordenadores centrales, entornos informáticos distribuidos que incluyen cualquiera de los sistemas o dispositivos anteriores, y similares.

El entorno operativo 510 de la FIG. 17 incluye un dispositivo informático de uso general en forma de ordenador 520, que incluye una unidad de procesamiento 521, una memoria de sistema 522 y un bus 523 del sistema 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 modo que el procesador del ordenador 520 incluya una única unidad de procesamiento central (CPU) o una pluralidad de unidades de procesamiento, comúnmente denominadas un entorno de procesamiento en paralelo. La ordenador 520 puede ser un ordenador convencional, un ordenador distribuido o cualquier otro tipo de ordenador.

El bus 523 del sistema puede ser cualquiera de 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 usa 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 de dentro del ordenador 520, tal como durante el arranque, se almacena en la ROM 524. El ordenador 520 puede incluir además una interfaz 527 de unidad de disco duro para leer y escribir en un disco duro, que no se muestra, una unidad de disco magnético 528 para leer o escribir en un disco magnético extraíble 529, y una unidad de disco óptico 530 para leer desde o escribir en un disco óptico extraíble 531 tal como un CD ROM u otro medio óptico.

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

Se pueden almacenar varios módulos de programa 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 la ordenador personal 520 a través de dispositivos de entrada como un teclado 540 y un dispositivo de puntero 542. Otros dispositivos de entrada (no mostrados) pueden incluir un micrófono, una palanca universal, un controlador de juegos, una antena parabólica, un escáner o similares. Estos y otros dispositivos de entrada a menudo se conectan a la unidad de procesamiento 521 a través de una interfaz 546 de puerto de serie que está acoplada al bus del sistema, pero se pueden conectar mediante otras interfaces, tales como un puerto paralelo, un puerto de juegos o un bus en serie universal (USB). Un monitor 547 u otro tipo de dispositivo de visualización también está conectado al bus 523 del sistema a través de una interfaz, como un adaptador de vídeo 548. Además del monitor, los ordenadores suelen incluir otros dispositivos de salida periféricos (no se muestran), como altavoces e impresoras.

El ordenador 520 puede funcionar en un entorno de red utilizando conexiones lógicas a uno o más ordenadores remotes, como el ordenador remoto 549. Estas conexiones lógicas pueden lograrse mediante un dispositivo de comunicación acoplado a o una parte del ordenador 520, o de otras maneras. El ordenador remoto 549 puede ser otro ordenador, un servidor, un enrutador, un PC de red, un cliente, un dispositivo 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 550 de almacenamiento de memoria en la FIG. 17. Las conexiones lógicas representadas en la FIG. 17 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 oficina, las redes informáticas de toda la empresa, las intranets e Internet, que son todos tipos de redes.

Cuando se usa en un entorno de red LAN, el ordenador 520 se conecta a la red local 551 a través de una interfaz 553 de red o adaptador, que es un tipo de dispositivo de comunicaciones. Cuando se usa en un entorno de red WAN, el ordenador 520 suele incluir 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, está conectado al bus 523 del sistema a través de la interfaz 546 del puerto de serie. En un entorno de red, los módulos de programa representados en relación con el ordenador personal 520, o partes del mismo, pueden almacenarse en el dispositivo de almacenamiento de memoria remota. Se aprecia que las conexiones de red mostradas son ejemplos no limitativos, y que se pueden usar otros dispositivos de comunicaciones para establecer un enlace de comunicaciones entre ordenadores.

En algunas implementaciones, un sistema que comprende uno o más microprocesadores y una memoria, memoria que comprende instrucciones ejecutables por uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba y cuyas instrucciones ejecutables por uno o más microprocesadores están configuradas para (a) determinar una o más estimaciones de la curvatura para la muestra de prueba a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) una característica de mapeo para las porciones del genoma de referencia y (b) calcular un nivel de sección genómica normalizado de cada una de las porciones del genoma de referencia para la muestra de prueba según (1) los recuentos las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la muestra de prueba, (2) la una o más estimaciones de la curvatura determinadas en (b) para la muestra de prueba, y (3) una o más estimaciones de la curvatura específicas de la porción de cada una de las múltiples porciones del genoma de referencia a partir de una relación ajustada entre (i) una o más estimaciones de la curvatura específicas de la muestra para una pluralidad de muestras, y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia para la pluralidad de muestras, configuradas de ese modo para proporcionar niveles de sección genómica calculados, donde el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del el genoma de referencia se reduce en los niveles de sección genómica calculados. En algunas implementaciones del sistema descrito anteriormente, una o más estimaciones de la curvatura específicas de la muestra de (b)(3) se obtienen a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en las porciones del genoma de referencia, y (ii) la característica de mapeo para cada una de las porciones del genoma de referencia, para cada una de la pluralidad de muestras.

Módulos

Se pueden utilizar uno o más módulos en un método descrito en el presente documento, cuyos ejemplos no limitativos incluyen un módulo de procesamiento lógico, 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 establecimiento de rango, módulo de categorización, módulo de representación gráfica, módulo de representación, módulo de relaciones, módulo de resultados y/o módulo de organización de visualización de datos, similares o una combinación de los mismos. Los módulos a veces son controlados por un microprocesador. En determinadas implementaciones, 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 hacia o desde otro módulo, máquina, componente, periférico u operador de un máquina. En algunas implementaciones, los datos y/o la información (por ejemplo, lecturas de secuencia) 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, un fotocélula, 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, temperatura, volumen, flujo, peso), un dispositivo de manejo de fluidos, una impresora, una pantalla (por ejemplo, un LED, LCT o CRT), similares o combinaciones de los mismos. Por ejemplo, algunas veces un operador de una máquina proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo. Un módulo a menudo se configura para transferir datos y/o información hacia o desde otro módulo o máquina. Un módulo puede recibir datos y/o información de otro módulo, cuyos ejemplos no limitativos incluyen un módulo de procesamiento lógico, 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 establecimiento de rango, módulo de categorización, módulo de representación gráfica, módulo de representación, módulo de relaciones, módulo de resultados y/o módulo de organización de visualización de datos, similares o una combinación de los mismos. Un módulo puede manipular y/o transformar información y/o datos. Los datos y/o la información derivados o transformados por un módulo pueden transferirse a otra máquina y/o módulo adecuado, cuyos ejemplos no limitativos incluyen un módulo de procesamiento lógico, 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 rango, módulo de categorización, módulo de representación gráfica, módulo de representación, módulo de relaciones, módulo de resultado y/o módulo de organización de visualización de datos, similares o una combinación de los mismos. Una máquina que comprende un módulo puede comprender al menos un procesador. En algunas implementaciones, los datos y/o la información son recibidos y/o proporcionados por una máquina que comprende un módulo. Una máquina que comprende un módulo puede incluir 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) de un módulo. En algunas implementaciones, un módulo funciona con uno o más procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)).

Módulo de procesamiento lógico

En determinadas implementaciones, un módulo de procesamiento lógico orquesta, controla, limita, organiza, ordena, distribuye, divide, transforma y/o regula información y/o datos o la transferencia de información y/o datos a y desde uno o más de otros módulos, periféricos o dispositivos.

Módulo de organización de visualización de datos

En determinadas implementaciones, un módulo de organización de visualización de datos procesa y/o transforma datos y/o información en un medio visual adecuado, cuyos ejemplos no limitativos incluyen imágenes, vídeo y/o texto (por ejemplo, números, letras y símbolos). En algunas implementaciones, un módulo de organización de visualización de datos procesa, transforma y/o transfiere datos y/o información para su presentación en un sistema de representación adecuado (por ejemplo, un monitor, LED, LCD, CRT, similares o combinaciones de los mismos), una impresora, un periférico o dispositivo. En algunas implementaciones, un módulo de organización de presentación de datos procesa, transforma datos y/o información en una representación visual de un genoma, cromosoma o parte del mismo fetal o materno.

Módulo de secuenciación

En algunas implementaciones, un módulo de secuencia obtiene, genera, reúne, ensambla, manipula, transforma, procesa, transforma y/o transfiere lecturas de secuencia. Un "módulo de recepción de secuencia", tal 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 cualquier máquina que determine la secuencia de un ácido nucleico utilizando una tecnología de secuenciación conocida en la técnica. En algunas implementaciones, un módulo de secuenciación puede alinear, ensamblar, fragmentar, complementar, complementar de manera inversa, comprobar errores o corregir errores en las lecturas de secuencias.

Módulo de mapeo

Las lecturas de secuencia pueden mapearse por un módulo de mapeo o por una máquina que comprende un módulo de mapeo, módulo de mapeo que mapea generalmente lecturas en un genoma de referencia o segmento del mismo. Un módulo de mapeo puede mapear lecturas de secuenciación mediante un método adecuado conocido en la técnica. En algunas implementaciones, se requiere un módulo de mapeo o una máquina que comprende un módulo de mapeo para proporcionar lecturas de secuencia mapeadas.

Módulo de recuento

Los recuentos pueden proporcionarse por un módulo de recuento o por una máquina que comprende un módulo de recuento. En algunas implementaciones, un módulo de recuento cuenta las lecturas de secuencia mapeadas en un genoma de referencia. En algunas implementaciones, un módulo de recuento genera, ensambla y/o proporciona recuentos según un método de recuento conocido en la técnica. En algunas implementaciones, se requiere un módulo de recuento o una máquina que incluya un módulo de recuento para proporcionar recuentos.

Módulo de filtrado

Las porciones de filtrado (por ejemplo, porciones de un genoma de referencia) pueden ser proporcionadas por un módulo de filtrado (por ejemplo, por una máquina que comprenda un módulo de filtrado). En algunas implementaciones, se requiere un módulo de filtrado para proporcionar datos de porciones filtradas (por ejemplo, porciones filtradas) y/o para eliminar porciones de la consideración. En determinadas implementaciones, un módulo de filtrado elimina los recuentos mapeados en una porción de la consideración. En determinadas implementaciones, un módulo de filtrado elimina los recuentos mapeados en una porción de la 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 varios métodos de filtrado conocidos en la técnica o descritos en el presente documento.

Módulo de ponderación

Las porciones ponderadas (por ejemplo, porciones de un genoma de referencia) pueden ser proporcionadas por un módulo de ponderación (por ejemplo, por una máquina que comprenda un módulo de ponderación). En algunas implementaciones, se requiere un módulo de ponderación para ponderar secciones genómicas y/o proporcionar valores de porción ponderados. Un módulo de ponderación puede ponderar las porciones mediante uno o varios métodos de ponderación conocidos en la técnica o descritos en el presente documento.

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 algunas implementaciones, se requiere un módulo de normalización para proporcionar datos normalizados (por ejemplo, recuentos normalizados) obtenidos a partir de lecturas de secuenciación. Un módulo de normalización puede normalizar 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.

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 ser proporcionada por un módulo de sesgo de GC (por ejemplo, por una máquina que comprenda un módulo de sesgo de GC). En algunas implementaciones, se requiere un módulo de sesgo de GC para proporcionar una determinación de sesgo de GC. En algunas implementaciones, 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 secuencias mapeadas en cada una de las porciones de un genoma de referencia y el contenido de GC de cada porción. A veces, un módulo de sesgo de GC forma parte de un módulo de normalización (por ejemplo, módulo de normalización PERUN).

Módulo de nivel

La determinación de niveles (por ejemplo, niveles) y/o el cálculo de niveles de porciones o secciones genómicas para porciones de un genoma de referencia pueden ser proporcionados por un módulo de nivel (por ejemplo, por una máquina que comprenda un módulo de nivel). En algunas implementaciones, se requiere un módulo de nivel para proporcionar un nivel o un nivel de porción o sección genómica calculado (por ejemplo, según la ecuación A, B, L, M, N, O y/o Q). En algunas implementaciones, 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 secuencias mapeadas en cada una de las porciones de un genoma de referencia. En algunas implementaciones, un módulo de nivel calcula un nivel de porción o sección genómica como parte del PERUN. En algunas implementaciones, un módulo de nivel proporciona una porción o un nivel de sección genómica (es decir, L_i) según la ecuación $L_i = (m_i - G_iS)I^{-1}$ en donde G_i es el sesgo de GC, m_i son recuentos medidos mapeados en cada porción de un genoma de referencia, i es una muestra, e I es la intersección y S es la pendiente de la relación a ajustada (por ejemplo, una relación lineal ajustada) entre un sesgo de GC y recuentos de lecturas de secuencia mapeadas en cada una de las porciones de un genoma de referencia.

Módulo de comparación

Un primer nivel puede ser identificado como significativamente diferente de un segundo nivel por un módulo de comparación o por una máquina que comprenda un módulo de comparación. En algunas implementaciones, se requiere un módulo de comparación o una máquina que comprenda un módulo de comparación para proporcionar una comparación entre dos niveles.

Módulo de establecimiento de rango

Un módulo de ajuste de rangos o una máquina que comprenda un módulo de ajuste de rangos pueden proporcionar rangos esperados (por ejemplo, rangos de niveles esperados) para diversas variaciones del número de copias (por ejemplo, duplicaciones, inserciones y/o deleciones) o rangos para la ausencia de una variación del número de copias. En determinadas implementaciones, los niveles esperados son proporcionados por un módulo de ajuste de rangos o por una máquina que comprende un módulo de ajuste de rangos. En algunas implementaciones, se requiere un módulo de ajuste de rango o una máquina que comprenda un módulo de ajuste de rango para proporcionar los niveles y/o rangos esperados.

Módulo de categorización

Una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal, una variación del número de copias fetal, una duplicación, inserción, deleción) puede categorizarse por un módulo de categorización o por una máquina que comprende un módulo de categorización. En determinadas implementaciones, una variación del número de copias (por ejemplo, una variación del número de copias materno y/o fetal) se categoriza por un módulo de categorización. En determinadas implementaciones, un nivel (por ejemplo, un primer nivel) determinado como significativamente diferente de otro nivel (por ejemplo, un segundo nivel) es identificado como representativo de una variación del número de copias por un módulo de categorización. En determinadas implementaciones, la ausencia de una variación del número de copias se determina por un módulo de categorización. En algunas implementaciones, una determinación de una variación del número de copias puede determinarse por una máquina que comprende un módulo de categorización. Un módulo de categorización puede especializarse para categorizar una variación del número de copias materno y/o fetal, una variación del número de copias fetal, una duplicación, deleción o inserción o la falta de los mismos o una combinación de los anteriores. Por ejemplo, un módulo de categorización que identifica una deleción materna puede ser diferente y/o distinto de un

módulo de categorización que identifica una duplicación fetal. En algunas implementaciones, se requiere un módulo de categorización o una máquina que comprende un módulo de categorización para identificar una variación del número de copias o un determinante del resultado de una variación del número de copias.

Módulo de ajuste

En algunas implementaciones, los ajustes (por ejemplo, ajustes para elevaciones o perfiles) los realiza un módulo de ajuste o un aparato que comprende un módulo de ajuste. En algunas implementaciones, se requiere un módulo de ajuste o un aparato que comprende un módulo de ajuste para ajustar una elevación. Una elevación ajustada mediante 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 y/o fetal).

Módulo de representación gráfica

En algunas implementaciones, un módulo de representación gráfica procesa y/o transforma datos y/o información en un medio visual adecuado, cuyos ejemplos no limitativos incluyen un cuadro, diagrama, gráfica, similares o combinaciones de los mismos. En algunas implementaciones, un módulo de representación gráfica procesa, transforma y/o transfiere datos y/o información para su presentación en un sistema de representación adecuado (por ejemplo, un monitor, LED, LCD, CRT, similares o combinaciones de los mismos), una impresora, un periférico o dispositivo. En determinadas implementaciones, un módulo de representación gráfica proporciona una visualización de un recuento, un nivel y/o un perfil. En algunas implementaciones, un módulo de organización de presentación de datos procesa, transforma datos y/o información en una representación visual de un genoma, cromosoma o parte del mismo fetal o materno. En algunas implementaciones, se requiere un módulo de representación gráfica o una máquina que incluya un módulo de representación gráfica para representar un recuento, un nivel o un perfil.

Módulo de representación

En determinadas implementaciones, una representación cromosómica está determinada por un módulo de representación. En determinadas implementaciones, un ECR está determinado por un módulo de representación esperado. En determinadas implementaciones, un MCR está determinado por un módulo de representación. Un módulo de representación puede ser un módulo de representación o un módulo de representación esperado. En algunas implementaciones, un módulo de representación determina una o más relaciones. Como se usa en el presente documento, el término “relación” se refiere a un valor numérico (por ejemplo, un número al que se llega) dividiendo un primer valor numérico entre un segundo valor numérico. Por ejemplo, una relación entre A y B se puede expresar matemáticamente como A/B o B/A y se puede obtener un valor numérico para la relación dividiendo A entre B o dividiendo B entre A. En determinadas implementaciones, un módulo de representación (por ejemplo, un módulo de representación) determina un MCR generando una relación de recuentos. En determinadas implementaciones, un módulo de representación determina un MCR para un autosoma afectado (por ejemplo, el cromosoma 13 en el caso de una trisomía 13, el cromosoma 18 en el caso de una trisomía 18 o el cromosoma 21 en el caso de una trisomía 21). Por ejemplo, a veces un módulo de representación (por ejemplo, un módulo de representación) determina un MCR mediante la generación de una relación de recuentos mapeados en porciones del cromosoma n con respecto al número total de recuentos mapeados en porciones de todos los cromosomas autosómicos representados en un perfil. En determinadas implementaciones, un módulo de representación (por ejemplo, un módulo de representación) determina un MCR mediante la generación de una relación de recuentos mapeados en porciones de un cromosoma sexual (por ejemplo, cromosoma X o Y) con respecto al número total de recuentos mapeados en porciones de todos los cromosomas autosómicos representados en un perfil. En determinadas implementaciones, un módulo de representación (por ejemplo, un módulo de representación esperado) determina un ECR mediante la generación de una relación de porciones. En determinadas implementaciones, un módulo de representación esperado determina un ECR para un autosoma afectado (por ejemplo, el cromosoma 13 en el caso de una trisomía 13, el cromosoma 18 en el caso de una trisomía 18 o el cromosoma 21 en el caso de una trisomía 21). Por ejemplo, a veces un módulo de representación (por ejemplo, un módulo de representación esperado) determina un ECR mediante la generación de una relación de porciones para el cromosoma n con todas las porciones autosómicas en un perfil. En algunas implementaciones, un módulo de representación puede proporcionar una relación de un MCR con respecto a un ECR. En determinadas implementaciones, un módulo de representación o un aparato que comprende un módulo de representación reúne, ensambla, recibe, proporciona y/o transfiere datos y/o información hacia o desde otro módulo, aparato, componente, periférico u operador de un aparato. Por ejemplo, algunas veces un operador de un aparato proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo de representación. Un módulo de representación puede recibir datos y/o información de un módulo de secuenciación, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos y/o módulo de procesamiento lógico. En determinadas implementaciones, los recuentos mapeados normalizados se transfieren a un módulo de representación desde un módulo de normalización. En determinadas implementaciones, los recuentos mapeados normalizados se transfieren a un módulo de representación esperado desde un módulo de normalización. Los datos y/o la información derivados o transformados por un módulo de representación pueden transferirse de un módulo de representación a un módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de

resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico, módulo de fracción fetal u otro aparato y/o módulo adecuado. En determinadas implementaciones, un MCR para los cromosomas 21, 18, 15, un cromosoma X y/o Y se transfiere a un módulo de fracción fetal desde un módulo de representación (por ejemplo, un módulo de representación). En determinadas implementaciones, un ECR para los cromosomas 21, 18, 15, un cromosoma X y/o Y se transfiere a un módulo de fracción fetal desde un módulo de representación (por ejemplo, un módulo de representación esperado). Un aparato que comprende un módulo de representación puede comprender al menos un procesador. En algunas implementaciones, se proporciona una representación por un aparato que incluye un procesador (por ejemplo, uno o más procesadores), procesador que puede realizar y/o implementar una o más instrucciones (por ejemplo, procesos, rutinas y/o subrutinas) del módulo de representación. En algunas implementaciones, un módulo de representación funciona con uno o más procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)).

Módulo de correspondencias

En determinadas implementaciones, una correspondencia está determinada por un módulo de correspondencias. En algunas implementaciones, se genera una correspondencia para la determinación de una fracción fetal y un MCR de un cromosoma X o Y mediante un módulo de correspondencias. En algunas implementaciones se genera una correspondencia para (i) una fracción fetal determinada mediante un primer método y (ii) una fracción fetal determinada mediante un segundo método por un módulo de correspondencias. En determinadas implementaciones, un módulo de correspondencias o un aparato que comprende un módulo de correspondencias reúne, ensambla, recibe, proporciona y/o transfiere datos y/o información hacia o desde otro módulo, aparato, componente, periférico u operador de un aparato. Por ejemplo, algunas veces un operador de un aparato proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo de correspondencias. Un módulo de correspondencias puede recibir datos y/o información de un módulo de secuenciación, módulo de secuenciación, módulo de mapeo, módulo de recuento, módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico y/o un módulo de representación. Los datos y/o la información derivados o transformados por un módulo de correspondencias pueden transferirse de un módulo de correspondencias a un módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico, módulo de representación, módulo de fracción fetal u otro aparato y/o módulo adecuado. Un aparato que comprende un módulo de correspondencias puede comprender al menos un procesador. En algunas implementaciones, los datos y/o la información son proporcionados por un aparato que 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) del módulo de correspondencias. En algunas implementaciones, un módulo de correspondencias funciona con uno o más procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)).

Módulo de fracción fetal

En determinadas implementaciones, una fracción fetal se determina mediante un módulo de fracción fetal. En determinadas implementaciones, un módulo de fracción fetal o un aparato que comprende un módulo de fracción fetal reúne, ensambla, recibe, proporciona y/o transfiere datos y/o información a o desde otro módulo, aparato, componente, periférico u operador de un aparato. Por ejemplo, a veces un operador de un aparato proporciona una constante, un valor umbral, una fórmula o un valor predeterminado a un módulo de fracción fetal. Un módulo de fracción fetal puede recibir datos y/o información de un módulo de secuenciación, módulo de secuenciación, módulo de mapeo, módulo de ponderación, módulo de filtrado, módulo de recuento, módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico, un módulo de representación y/o un módulo de correspondencias. Los datos y/o la información derivados o transformados por un módulo de fracción fetal pueden transferirse de un módulo de fracción fetal a un módulo de normalización, módulo de comparación, módulo de configuración de rango, módulo de categorización, módulo de ajuste, módulo de representación gráfica, módulo de resultados, módulo de organización de visualización de datos, módulo de procesamiento lógico, módulo de representación, módulo de correspondencias, módulo de fracción fetal u otro aparato y/o módulo adecuado. Un aparato que incluye un módulo de fracción fetal puede comprender al menos un procesador. En algunas implementaciones, los datos y/o la información son proporcionados por un aparato que 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) del módulo de fracción fetal. En algunas implementaciones, un módulo de fracción fetal funciona con uno o más procesadores externos (por ejemplo, una red interna o externa, un servidor, dispositivo de almacenamiento y/o una red de almacenamiento (por ejemplo, una nube)).

En algunas implementaciones, el módulo de secuenciación y el módulo de mapeo están configurados para transferir lecturas de secuencia 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 secuencia mapeadas desde el módulo de mapeo al módulo de recuento. A veces, el módulo de recuento y el módulo de filtrado están configurados para transferir recuentos desde el módulo de recuento al módulo de filtrado. A veces, el módulo de recuento y el módulo de

ponderación están configurados para transferir recuentos desde el módulo de recuento al módulo de ponderación. El módulo de mapeo y el módulo de filtrado a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de filtrado. El módulo de mapeo y el módulo de ponderación a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de ponderación. En determinadas implementaciones, el módulo de ponderación, el módulo de filtrado y el módulo de recuento están configurados para transferir porciones filtradas y/o ponderadas desde el módulo de ponderación y el módulo de filtrado al módulo de recuento. El módulo de ponderación y el módulo de normalización a veces están configurados para transferir porciones ponderadas desde el módulo de ponderación al módulo de normalización. El módulo de filtrado y el módulo de normalización a veces están configurados para transferir porciones filtradas desde el módulo de filtrado al módulo de normalización. En algunas implementaciones, el módulo de normalización y/o módulo de comparación están configurados para transferir recuentos normalizados al módulo de comparación y/o módulo de establecimiento de rango. El módulo de comparación, el módulo de establecimiento de rango y/o el módulo de categorización están configurados independientemente para transferir (i) una identificación de una primera elevación que es significativamente diferente de una segunda elevación y/o (ii) un rango de nivel esperado desde el módulo de comparación y/o el módulo de establecimiento de rango al módulo de categorización, en algunas implementaciones. En determinadas implementaciones, el módulo de categorización y el módulo de ajuste están configurados para transferir una elevación categorizada como variación del número de copias desde el módulo de categorización al módulo de ajuste y/o al módulo de fracción fetal. En algunas implementaciones, el módulo de ajuste, el módulo de representación gráfica y el módulo de resultados están configurados para transferir uno o más niveles ajustados desde el módulo de ajuste al módulo de representación gráfica, al módulo de resultados o al módulo de fracción fetal. En ocasiones, el módulo de normalización está configurado para transferir recuentos de lecturas de secuencias normalizadas mapeadas a uno o más de los módulos de comparación, módulo de ajuste de rangos, módulo de categorización, módulo de ajuste, módulo de resultados, módulo de representación gráfica, módulo de fracción fetal o módulo de representación. En algunas implementaciones, un módulo de correspondencias está configurado para recibir información del módulo de representación y está configurado para transferir información al módulo de fracción fetal.

En algunas implementaciones, un aparato (por ejemplo, un primer aparato) comprende un módulo de normalización, un módulo de representación, un módulo de representación esperada, un módulo de fracción fetal y un módulo de correspondencias. En algunas implementaciones, un aparato (por ejemplo, un segundo aparato) comprende un módulo de mapeo y un módulo de recuento. En determinadas implementaciones, un aparato (por ejemplo, un tercer aparato) comprende un módulo de secuenciación.

Módulo de relaciones

En determinadas implementaciones, un módulo de relaciones procesa y/o transforma datos y/o información en una relación. En determinadas implementaciones, una relación es generada y/o transferida desde un módulo de relaciones.

Módulo de resultados

La presencia o ausencia de una variación genética (una aneuploidía, una aneuploidía fetal, una variación del número de copias) es, en algunas implementaciones, identificada por un módulo de resultados o por una máquina que comprende un módulo de resultados. En determinadas implementaciones, una variación genética es identificada por un módulo de resultados. A menudo, un módulo de resultados identifica una determinación de la presencia o ausencia de una aneuploidía. En algunas implementaciones, un determinante del resultado de una variación genética (una aneuploidía, una variación del número de copias) puede identificarse por un módulo de resultados o por una máquina que comprende un módulo de resultados. Un módulo de resultados puede especializarse para determinar una variación genética específica (por ejemplo, una trisomía, una trisomía 21, una trisomía 18). Por ejemplo, un módulo de resultados que identifica una trisomía 21 puede ser diferente y/o diferente de un módulo de resultados que identifica una trisomía 18. En algunas implementaciones, se requiere un módulo de resultados o una máquina que comprende un módulo de resultados para identificar una variación genética o un determinante del resultado de una variación genética (por ejemplo, una aneuploidía, una variación del número de copias). Una variación genética o un determinante del resultado de una variación genética identificada mediante los métodos descritos en el presente documento pueden verificarse independientemente mediante pruebas adicionales (por ejemplo, por secuenciación dirigida de ácido nucleico materno y/o fetal).

Transformaciones

Tal como se mencionó anteriormente, los datos a veces se transforman de una forma a otra. Los términos "transformado/a", "transformación" y derivaciones gramaticales o equivalentes de los mismos, tal 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 de muestra de sujeto de referencia y/o sujeto de prueba) en una representación digital del material de partida físico (por ejemplo, datos de lectura de secuencia), y en algunas implementaciones, incluye una transformación adicional en uno o más valores numéricos o representaciones gráficas de la representación digital que pueden usarse para proporcionar un resultado. En determinadas implementaciones, el uno o más valores numéricos y/o representaciones gráficas de datos representados digitalmente pueden usarse para representar el

aspecto del genoma físico de un sujeto de prueba (por ejemplo, representar virtualmente o representar visualmente la presencia o ausencia de una inserción genómica, duplicación o deleción; representar la presencia o ausencia de una variación en la cantidad física de una secuencia asociada con afecciones médicas). A veces, una representación virtual se transforma además en uno o más valores numéricos o representaciones gráficas de la representación digital del material de partida. Estos métodos pueden transformar material de partida físico en un valor numérico o representación gráfica, o una representación del aspecto físico del genoma de un sujeto de prueba.

En algunas implementaciones, 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 procedimiento de transformación de un material de partida físico en una representación virtual del material de partida (por ejemplo, lecturas de secuencia representativas del material de partida físico). Una característica o variable adecuada puede usarse para reducir la complejidad y/o dimensionalidad del conjunto de datos. Los ejemplos no limitativos de características que pueden seleccionarse para su uso como característica objetivo para el procesamiento de datos incluyen contenido de GC, predicción del sexo del feto, 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 génica, una ontología de proteínas, genes corregulados, genes de señalización celular, genes del ciclo celular, proteínas que pertenecen a los genes anteriores, variantes génicas, variantes de proteínas, genes corregulados, proteínas correguladas, secuencia de aminoácidos, secuencia de nucleótidos, datos de estructura proteica y similares, y combinaciones de los anteriores. Los ejemplos no limitativos de complejidad de conjuntos de datos y/o reducción de dimensionalidad incluyen; reducción de una pluralidad de lecturas de secuencia a gráficos de perfiles, reducción de una pluralidad de lecturas de secuencia a valores numéricos (por ejemplo, valores normalizados, puntuaciones Z, valores de p); reducción de múltiples métodos de análisis a gráficos de probabilidad o puntos únicos; análisis de componentes principales de cantidades derivadas; y similares, o combinaciones de los mismos.

Sistemas de normalización de porciones, aparatos y productos de programas informáticos

En determinados aspectos, se proporciona un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más procesadores y memoria que comprende recuentos de lecturas de secuencia de ácido nucleico de muestra circulante, libre de células de un sujeto de prueba mapeado en porciones de un genoma de referencia; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) generar un perfil de recuento normalizado de muestra normalizando los recuentos de las lecturas de secuencia para cada una de las porciones; y (b) determinar la presencia o ausencia de una aberración cromosómica segmentaria o una aneuploidía fetal o ambas a partir del perfil de recuento normalizado de la muestra de (a).

En determinados aspectos, también se proporciona un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más procesadores y memoria que comprende recuentos de lecturas de secuencia de ácido nucleico de muestra circulante libre de células de un sujeto de prueba mapeado en porciones de un genoma de referencia; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) generar un perfil de recuento normalizado de muestra normalizando los recuentos de las lecturas de secuencia para cada una de las porciones; y (b) determinar la presencia o ausencia de una aberración cromosómica segmentaria o una aneuploidía fetal o ambas a partir del perfil de recuento normalizado de la muestra de (a).

También se proporciona en determinados aspectos un producto de programa informático incorporado de manera 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 secuencia de ácido nucleico de muestra circulante libre de células de un sujeto de prueba mapeado en porciones de un genoma de referencia; (b) generar un perfil de recuento normalizado de muestra normalizando los recuentos de las lecturas de secuencia para cada una de las porciones; y (c) determinar la presencia o ausencia de una aberración cromosómica segmentaria o una aneuploidía fetal o ambas a partir del perfil de recuento normalizado de la muestra de (b).

En algunas implementaciones, los recuentos de las lecturas de secuencia de cada una de las porciones de un segmento del genoma de referencia (por ejemplo, el segmento es un cromosoma) se normalizan individualmente según los recuentos totales de lecturas de secuencia de las porciones del segmento. Algunas veces, se eliminan determinadas porciones del segmento (por ejemplo, se filtran) y las porciones restantes del segmento se normalizan.

En determinadas realizaciones, el sistema, aparato y/o producto de programa informático comprende: (i) un módulo de secuenciación configurado para obtener lecturas de secuencia de ácido nucleico; (ii) un módulo de mapeo configurado para mapear lecturas de secuencia de ácido nucleico en porciones de un genoma de referencia; (iii) un módulo de ponderación configurado para ponderar porciones, (iv) un módulo de filtrado configurado para filtrar porciones o recuentos mapeados en una porción, (v) un módulo de recuento configurado para proporcionar recuentos de lecturas de secuencias de ácido nucleico mapeadas en porciones de un genoma de referencia; (vi) un módulo de normalización configurado para proporcionar recuentos normalizados; (vii) un módulo de comparación configurado para proporcionar una identificación de una primera elevación que es significativamente diferente de una segunda elevación; (viii) un módulo de ajuste de rango configurado para proporcionar uno o más rangos de niveles esperados; (ix) un módulo de categorización configurado para identificar una elevación representativa de una variación del número de copias; (x) un módulo de ajuste

configurado para ajustar un nivel identificado como una variación del número de copias; (xi) un módulo de representación gráfica configurado para representar y mostrar un nivel y/o un perfil; (xii) un módulo de resultados configurado para determinar un resultado (por ejemplo, un resultado determinante de la presencia o ausencia de una aneuploidía fetal); (xiii) un módulo de organización de visualización de datos configurado para indicar la presencia o ausencia de una anomalía cromosómica segmentaria o una aneuploidía fetal o ambas; (xiv) un módulo de procesamiento lógico configurado para realizar uno o más mapeo de lecturas de secuencia, contar lecturas de secuencia mapeadas, normalizar recuentos y generar un resultado; o (xv) una combinación de dos o más de los anteriores.

En algunas implementaciones, el módulo de secuenciación y el módulo de mapeo están configurados para transferir lecturas de secuencia 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 secuencia mapeadas desde el módulo de mapeo al módulo de recuento. A veces, el módulo de recuento y el módulo de filtrado están configurados para transferir recuentos desde el módulo de recuento al módulo de filtrado. A veces, el módulo de recuento y el módulo de ponderación están configurados para transferir recuentos desde el módulo de recuento al módulo de ponderación. El módulo de mapeo y el módulo de filtrado a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de filtrado. El módulo de mapeo y el módulo de ponderación a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de ponderación. En algunas implementaciones, el módulo de ponderación, el módulo de filtrado y el módulo de recuento están configurados para transferir porciones filtradas y/o ponderadas desde el módulo de ponderación y el módulo de filtrado al módulo de recuento. El módulo de ponderación y el módulo de normalización a veces están configurados para transferir porciones ponderadas desde el módulo de ponderación al módulo de normalización. El módulo de filtrado y el módulo de normalización a veces están configurados para transferir porciones filtradas desde el módulo de filtrado al módulo de normalización. En algunas implementaciones, el módulo de normalización y/o módulo de comparación están configurados para transferir recuentos normalizados al módulo de comparación y/o módulo de establecimiento de rango. El módulo de comparación, el módulo de establecimiento de rango y/o el módulo de categorización están configurados independientemente para transferir (i) una identificación de una primera elevación que es significativamente diferente de una segunda elevación y/o (ii) un rango de nivel esperado desde el módulo de comparación y/o el módulo de establecimiento de rango al módulo de categorización, en algunas implementaciones. En determinadas implementaciones, el módulo de categorización y el módulo de ajuste están configurados para transferir una elevación categorizada como una variación del número de copias desde el módulo de categorización hasta el módulo de ajuste. En algunas implementaciones, el módulo de ajuste, el módulo de representación gráfica y el módulo de resultados están configurados para transferir uno o más niveles ajustados desde el módulo de ajuste al módulo de representación gráfica o el módulo de resultados. El módulo de normalización a veces está configurado para transferir recuentos de lectura de secuencia normalizada mapeados en uno o más del módulo de comparación, el módulo de establecimiento de rango, el módulo de categorización, el módulo de ajuste, el módulo de resultados o el módulo de representación gráfica.

Sistemas parametrizados de eliminación de errores y de normalización no sesgada, aparatos y productos de programas informáticos

Se proporcionan en determinados aspectos un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; y (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre (i) el sesgo de GC y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

También se proporciona en algunos aspectos un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una muestra de prueba; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; y (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre (i) el sesgo de GC y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

También se proporciona en determinados aspectos un producto de programa informático incorporado de manera 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 secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante extracelular a partir de una muestra de prueba; (d) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; y (c) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre (i) el sesgo de GC y (ii) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

Se proporcionan en determinados aspectos un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada portadora de un feto; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo de GC y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo los niveles de porción calculados; y (c) identificar la presencia o ausencia de una aneuploidía para el feto según los niveles de porción calculados con una sensibilidad del 95 % o más y una especificidad del 95 % o más.

También se proporciona en determinados aspectos un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada portadora de un feto; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo de GC y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo los niveles de porción calculados; y (c) identificar la presencia o ausencia de una aneuploidía para el feto según los niveles de porción calculados con una sensibilidad del 95 % o más y una especificidad del 95 % o más.

También en determinados aspectos se proporciona un producto de programa informático incorporado de manera 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 secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada portadora de un feto; (d) determinar un sesgo de guanina y citosina (GC) para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada para cada muestra entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) el contenido de GC para cada una de las porciones; (c) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo de GC y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo los niveles de porción calculados; y (d) identificar la presencia o ausencia de una aneuploidía para el feto según los niveles de porción calculados con una sensibilidad del 95 % o más y una especificidad del 95 % o más.

También se proporciona en determinados aspectos un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada portadora de un feto; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar el sesgo experimental para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) una característica de mapeo para cada una de las porciones; y (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo experimental y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

También se proporciona en determinados aspectos un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son

lecturas de ácido nucleico circulante libre de células de una mujer embarazada portadora de un feto; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) determinar el sesgo experimental para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) una característica de mapeo para cada una de las porciones; y (b) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo experimental y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

También se proporciona en determinados aspectos un producto de programa informático incorporado de manera 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 secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante extracelulcar a partir de una muestra de prueba; (b) determinar el sesgo experimental para cada una de las porciones del genoma de referencia para múltiples muestras a partir de una relación ajustada entre (i) los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, y (ii) una característica de mapeo para cada una de las porciones; y (c) calcular un nivel de porción para cada una de las porciones del genoma de referencia a partir de una relación ajustada entre el sesgo experimental y los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia, proporcionando de este modo niveles de porción calculados, por lo que el sesgo en los recuentos de las lecturas de secuencia mapeadas en cada una de las porciones del genoma de referencia se reduce en los niveles de porción calculados.

En determinadas realizaciones, el sistema, aparato y/o producto de programa informático comprende: (i) un módulo de secuenciación configurado para obtener lecturas de secuencia de ácido nucleico; (ii) un módulo de mapeo configurado para mapear lecturas de secuencia de ácido nucleico en porciones de un genoma de referencia; (iii) un módulo de ponderación configurado para ponderar las porciones; (iv) un módulo de filtrado configurado para filtrar porciones o recuentos mapeados en una porción; (v) un módulo de recuento configurado para proporcionar recuentos de lecturas de secuencias de ácidos nucleicos mapeadas en porciones de un genoma de referencia; (vi) un módulo de normalización configurado para proporcionar recuentos normalizados; (vii) un módulo de comparación configurado para proporcionar una identificación de una primera elevación que es significativamente diferente de una segunda elevación; (viii) un módulo de ajuste de rango configurado para proporcionar uno o más rangos de niveles esperados; (ix) un módulo de categorización configurado para identificar una elevación representativa de una variación del número de copias; (x) un módulo de ajuste configurado para ajustar un nivel identificado como una variación del número de copias; (xi) un módulo de representación gráfica configurado para representar y mostrar un nivel y/o un perfil; (xii) un módulo de resultados configurado para determinar un resultado (por ejemplo, un resultado determinante de la presencia o ausencia de una aneuploidía fetal); (xiii) un módulo de organización de visualización de datos configurado para indicar la presencia o ausencia de una anomalía cromosómica segmentaria o una aneuploidía fetal o ambas; (xiv) un módulo de procesamiento lógico configurado para realizar uno o más mapeo de lecturas de secuencia, contar lecturas de secuencia mapeadas, normalizar recuentos y generar un resultado; o (xv) una combinación de dos o más de los anteriores.

En algunas implementaciones, el módulo de secuenciación y el módulo de mapeo están configurados para transferir lecturas de secuencia 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 secuencia mapeadas desde el módulo de mapeo al módulo de recuento. A veces, el módulo de recuento y el módulo de filtrado están configurados para transferir recuentos desde el módulo de recuento al módulo de filtrado. A veces, el módulo de recuento y el módulo de ponderación están configurados para transferir recuentos desde el módulo de recuento al módulo de ponderación. El módulo de mapeo y el módulo de filtrado a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de filtrado. El módulo de mapeo y el módulo de ponderación a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de ponderación. En algunas implementaciones, el módulo de ponderación, el módulo de filtrado y el módulo de recuento están configurados para transferir porciones filtradas y/o ponderadas desde el módulo de ponderación y el módulo de filtrado al módulo de recuento. El módulo de ponderación y el módulo de normalización a veces están configurados para transferir porciones ponderadas desde el módulo de ponderación al módulo de normalización. El módulo de filtrado y el módulo de normalización a veces están configurados para transferir porciones filtradas desde el módulo de filtrado al módulo de normalización. En algunas implementaciones, el módulo de normalización y/o módulo de comparación están configurados para transferir recuentos normalizados al módulo de comparación y/o módulo de establecimiento de rango. El módulo de comparación, el módulo de establecimiento de rango y/o el módulo de categorización están configurados independientemente para transferir (i) una identificación de una primera elevación que es significativamente diferente de una segunda elevación y/o (ii) un rango de nivel esperado desde el módulo de comparación y/o el módulo de establecimiento de rango al módulo de categorización, en algunas implementaciones. En determinadas implementaciones, el módulo de categorización y el módulo de ajuste están configurados para transferir una elevación categorizada como una variación del número de copias desde el módulo de categorización hasta el módulo de ajuste. En algunas implementaciones, el módulo de ajuste, el módulo de representación gráfica y el módulo de resultados están configurados para transferir uno o más niveles ajustados desde el módulo de ajuste al módulo de representación gráfica o el módulo de resultados. El módulo de normalización a veces está configurado para transferir recuentos de lectura de secuencia normalizada mapeados en uno o más del módulo de comparación, el módulo de establecimiento de rango, el módulo de categorización, el módulo de ajuste, el módulo de resultados o el módulo de representación gráfica.

Sistemas, aparatos y productos de programa informático de ajuste

En determinados aspectos, se proporciona un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (d) ajustar la primera elevación mediante un valor predeterminado cuando la primera elevación está dentro de uno de los rangos de elevación esperados, proporcionando de este modo un ajuste de la primera elevación; y (e) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

También se proporciona en determinados aspectos un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (d) ajustar la primera elevación mediante un valor predeterminado cuando la primera elevación está dentro de uno de los rangos de elevación esperados, proporcionando de este modo un ajuste de la primera elevación; y (e) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

También en determinados aspectos se proporciona un producto de programa informático incorporado de manera 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 secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; (b) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (c) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (d) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (e) ajustar la primera elevación mediante un valor predeterminado cuando la primera elevación está dentro de uno de los rangos de elevación esperados, proporcionando de este modo un ajuste de la primera elevación; y (f) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

También se proporciona en determinados aspectos un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; y (d) identificar una variación del número de copias materno y/o fetal dentro de la porción basándose en uno de los rangos de elevación esperados, por lo que la variación del número de copias materno y/o fetal se identifica a partir de las lecturas de la secuencia de ácido nucleico.

También se proporciona, en determinados aspectos, un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas

de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; y (d) identificar una variación del número de copias materno y/o fetal dentro de la porción basándose en uno de los rangos de elevación esperados, por lo que la variación del número de copias materno y/o fetal se identifica a partir de las lecturas de la secuencia de ácido nucleico.

También se proporciona en determinados aspectos un producto de programa informático incorporado de manera 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 secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; (b) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (c) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (d) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; y (e) identificar una variación del número de copias materno y/o fetal dentro de la porción basándose en uno de los rangos de elevación esperados, por lo que la variación del número de copias materno y/o fetal se identifica a partir de las lecturas de la secuencia de ácido nucleico.

También se proporciona en algunos aspectos un sistema que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (d) ajustar la primera elevación según la segunda elevación, proporcionando de este modo un ajuste de la primera elevación; y (e) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

En determinados aspectos, se proporciona un aparato que comprende uno o más procesadores y memoria, memoria que comprende instrucciones ejecutables por el uno o más microprocesadores y memoria que comprende recuentos de lecturas de secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; e instrucciones ejecutables por uno o más procesadores que están configuradas para: (a) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (b) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (c) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (d) ajustar la primera elevación según la segunda elevación, proporcionando de este modo un ajuste de la primera elevación; y (e) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

En algunos aspectos, se proporciona un producto de programa informático incorporado de manera 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 secuencia de ácido nucleico mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada; (b) normalizar los recuentos mapeados en las porciones del genoma de referencia, proporcionando de este modo un perfil de recuentos normalizados para las porciones; (c) identificar una primera elevación de los recuentos normalizados significativamente diferente de una segunda elevación de los recuentos normalizados en el perfil, cuya primera elevación es para un primer conjunto de porciones y cuya segunda elevación es para un segundo conjunto de porciones; (d) determinar un rango de elevación esperado para una variación del número de copias homocigotas y heterocigotas según un valor de incertidumbre para un segmento del genoma; (e) ajustar la primera elevación según la segunda elevación, proporcionando de este modo un ajuste de la primera elevación; y (f) determinar la presencia o ausencia de una aneuploidía cromosómica en el feto según las elevaciones de las porciones que

comprenden el ajuste de (d), mediante lo que el resultado determinante de la presencia o ausencia de la aneuploidía cromosómica se genera a partir de las lecturas de secuencias de ácido nucleico.

En determinadas realizaciones, el sistema, aparato y/o producto de programa informático comprende: (i) un módulo de secuenciación configurado para obtener lecturas de secuencia de ácido nucleico; (ii) un módulo de mapeo configurado para mapear lecturas de secuencia de ácido nucleico en porciones de un genoma de referencia; (iii) un módulo de ponderación configurado para ponderar las porciones; (iv) un módulo de filtrado configurado para filtrar porciones o recuentos mapeados en una porción; (v) un módulo de recuento configurado para proporcionar recuentos de lecturas de secuencias de ácidos nucleicos mapeadas en porciones de un genoma de referencia; (vi) un módulo de normalización configurado para proporcionar recuentos normalizados; (vii) un módulo de comparación configurado para proporcionar una identificación de una primera elevación que es significativamente diferente de una segunda elevación; (viii) un módulo de ajuste de rango configurado para proporcionar uno o más rangos de niveles esperados; (ix) un módulo de categorización configurado para identificar una elevación representativa de una variación del número de copias; (x) un módulo de ajuste configurado para ajustar un nivel identificado como una variación del número de copias; (xi) un módulo de representación gráfica configurado para representar y mostrar un nivel y/o un perfil; (xii) un módulo de resultados configurado para determinar un resultado (por ejemplo, un resultado determinante de la presencia o ausencia de una aneuploidía fetal); (xiii) un módulo de organización de visualización de datos configurado para indicar la presencia o ausencia de una anomalía cromosómica segmentaria o una aneuploidía fetal o ambas; (xiv) un módulo de procesamiento lógico configurado para realizar uno o más mapeo de lecturas de secuencia, contar lecturas de secuencia mapeadas, normalizar recuentos y generar un resultado; o (xv) una combinación de dos o más de los anteriores.

En algunas implementaciones, el módulo de secuenciación y el módulo de mapeo están configurados para transferir lecturas de secuencia 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 secuencia mapeadas desde el módulo de mapeo al módulo de recuento. A veces, el módulo de recuento y el módulo de filtrado están configurados para transferir recuentos desde el módulo de recuento al módulo de filtrado. A veces, el módulo de recuento y el módulo de ponderación están configurados para transferir recuentos desde el módulo de recuento al módulo de ponderación. El módulo de mapeo y el módulo de filtrado a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de filtrado. El módulo de mapeo y el módulo de ponderación a veces están configurados para transferir lecturas de secuencia mapeadas desde el módulo de mapeo al módulo de ponderación. En algunas implementaciones, el módulo de ponderación, el módulo de filtrado y el módulo de recuento están configurados para transferir porciones filtradas y/o ponderadas desde el módulo de ponderación y el módulo de filtrado al módulo de recuento. El módulo de ponderación y el módulo de normalización a veces están configurados para transferir porciones ponderadas desde el módulo de ponderación al módulo de normalización. El módulo de filtrado y el módulo de normalización a veces están configurados para transferir porciones filtradas desde el módulo de filtrado al módulo de normalización. En algunas implementaciones, el módulo de normalización y/o módulo de comparación están configurados para transferir recuentos normalizados al módulo de comparación y/o módulo de establecimiento de rango. El módulo de comparación, el módulo de establecimiento de rango y/o el módulo de categorización están configurados independientemente para transferir (i) una identificación de una primera elevación que es significativamente diferente de una segunda elevación y/o (ii) un rango de nivel esperado desde el módulo de comparación y/o el módulo de establecimiento de rango al módulo de categorización, en algunas implementaciones. En determinadas implementaciones, el módulo de categorización y el módulo de ajuste están configurados para transferir una elevación categorizada como una variación del número de copias desde el módulo de categorización hasta el módulo de ajuste. En algunas implementaciones, el módulo de ajuste, el módulo de representación gráfica y el módulo de resultados están configurados para transferir uno o más niveles ajustados desde el módulo de ajuste al módulo de representación gráfica o el módulo de resultados. El módulo de normalización a veces está configurado para transferir recuentos de lectura de secuencia normalizada mapeados en uno o más del módulo de comparación, el módulo de establecimiento de rango, el módulo de categorización, el módulo de ajuste, el módulo de resultados o el módulo de representación gráfica.

En determinados aspectos, se proporciona un sistema que comprende uno o más procesadores y una memoria, memoria que comprende instrucciones ejecutables por uno o más procesadores y memoria que comprende recuentos de lecturas de secuencia de ácidos nucleicos mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, (b) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, generando de esta manera una evaluación, (c) normalizar los recuentos mediante un proceso seleccionado según la evaluación, generando de esta manera recuentos normalizados con sesgo reducido y (d) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados.

En determinados aspectos, se proporciona un aparato que comprende uno o más procesadores y una memoria, memoria que comprende instrucciones ejecutables por uno o más procesadores y memoria que comprende recuentos de lecturas de secuencia de ácidos nucleicos mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células circulantes de una mujer embarazada, y cuyas instrucciones ejecutables por uno o más procesadores están configuradas para (a) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, (b) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal

o una regresión lineal, generando de esta manera una evaluación, (c) normalizar los recuentos mediante un proceso seleccionado según la evaluación, generando de esta manera recuentos normalizados con sesgo reducido y (d) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados.

En determinados aspectos, se proporciona un producto de programa informático materializado tangiblemente en un medio legible por ordenador, que comprende instrucciones que, cuando son ejecutadas por uno o más procesadores, están configuradas para (a) acceder a recuentos de lecturas de secuencia de ácidos nucleicos mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante libre de células de una mujer embarazada, (b) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba, (c) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, generando de este modo una evaluación, (d) normalizar los recuentos mediante un proceso seleccionado según la evaluación, generando de este modo recuentos normalizados con sesgo reducido y (e) analizar el ácido nucleico de la mujer embarazada según los recuentos normalizados.

Ejemplos

Los siguientes ejemplos se proporcionan únicamente a modo de ilustración y no a modo de limitación. Por lo tanto, los ejemplos expuestos a continuación ilustran determinadas implementaciones y no limitan la tecnología. Los expertos en la técnica reconocerán fácilmente una variedad de parámetros no críticos que podrían cambiarse o modificarse para producir resultados esencialmente iguales o similares.

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

Los métodos y la teoría subyacente descritos en el presente documento pueden usarse para detectar diversas 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

Múltiples intentos de eliminar porciones poco informativas de un genoma de referencia han indicado que la selección de porciones tiene el potencial de mejorar la clasificación.

Ecuación A:

$$M = LI + GS \qquad \textbf{(A)}$$

Los diversos términos de la ec. A tienen los siguientes significados:

M: recuentos medidos, que representan la información primaria contaminada por 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 euploide. Esta es la cantidad enmascarada tanto por errores estocásticos como por sesgos sistemáticos. El nivel cromosómico L es tanto específico de la muestra como específico de la porción.

G: Coeficiente de sesgo de GC medido usando un modelo lineal, LOESS o cualquier enfoque equivalente. G representa la información secundaria, extraída de M y de un conjunto de valores de contenido de GC específicos de porciones, derivados habitualmente del genoma de referencia (pero también pueden derivarse del contenido de GC realmente observado). G es específica de la muestra y no varía a lo largo de la posición genómica. Encapsula 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 la porción.

S: pendiente del modelo lineal. Este parámetro del modelo es fijo para una configuración experimental dada, independiente de la muestra, y específico de la porción.

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

Extracción del nivel cromosómico a partir de recuentos medidos

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

L= (M-GS)/I **(B)**

Como en la ec. A, el coeficiente G de sesgo de GC se evalúa como la pendiente de la regresión entre los recuentos sin procesar M medidos por porciones y el contenido de GC del genoma de referencia. El nivel cromosómico L se usa después para análisis adicionales (valores de Z, deleciones/duplicaciones maternas, microdeleciones/microduplicaciones fetales, sexo del feto, aneuploidías sexuales, etcétera). El procedimiento encapsulado por la Ec. B se denomina eliminación del error parametrizado y normalización no sesgada (PERUN).

Ejemplo 2: Ejemplos de fórmulas

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

Las puntuaciones Z y los valores de *p* calculados a partir de las puntuaciones Z asociadas a las desviaciones del nivel esperado de 1 pueden evaluarse entonces a la luz de la estimación de la incertidumbre en el nivel promedio. Los valores de *p* se basan en una distribución t cuyo orden viene determinado por el número de porciones de un genoma de referencia en un pico. Dependiendo del nivel de confianza deseado, un punto 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 puede utilizarse para comparar directamente el nivel de pico 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 de la *t* que arrojará un valor de *p* que mida la similitud entre dos muestras viene determinado por el número de porciones de un genoma de referencia en el más corto de los dos tramos desviados.

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

Ecuación 8:

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

donde Y_i representa los recuentos medidos para una porción en la muestra de prueba correspondiente a la porción en la mediana del perfil de recuento, F representa la fracción fetal, X representa la ploidía fetal y M_i representa la 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 está afectado y el otro no. 5/4 se usa en el caso de gemelos donde un feto se ve afectado y el otro no, porque el término F en la ecuación (8) representa el ADN fetal total, por tanto, debe tenerse en cuenta todo el ADN fetal. En algunas implementaciones, las grandes deleciones y/o duplicaciones en el genoma materno pueden contabilizarse asignando la ploidía materna, M_i, a cada porción o porción. La ploidía materna se asigna a menudo como un múltiplo de 1/2, y puede estimarse usando normalización basada en porciones, en algunas implementaciones. Debido a que la ploidía materna es a menudo un múltiplo de 1/2, la ploidía materna puede tenerse en cuenta fácilmente y, por tanto, no se incluirá en ecuaciones adicionales para simplificar las derivaciones.

Cuando se evalúa la ecuación (8) en X = 1, (por ejemplo, suposición euploide), la fracción fetal se cancela y resulta la siguiente ecuación para la suma de residuos al cuadrado.

Ecuación 9:

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

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

Ecuación 10:

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

Ecuación 11:

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

Ecuación 12:

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

Cuando se evalúa la ecuación (8) en X = 3/2 (por ejemplo, suposición triploide), resulta la siguiente ecuación para la suma de los residuos al cuadrado.

Ecuación 13:

$$\varphi_T = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left(y_i - f_i - \frac{1}{2}Ff_i\right)^2 = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} + F\left(\Xi_{ff} - \Xi_{fy}\right) + \frac{1}{4}F^2\Xi_{ff} \quad (13)$$

La diferencia entre las ecuaciones (9) y (13) forma el resultado funcional (por ejemplo, phi) que puede usarse para someter a prueba la hipótesis nula (por ejemplo, euploide, X = 1) contra la hipótesis alternativa (por ejemplo, trisomía única, X = 3/2):

Ecuación 14:

$$\varphi = \varphi_E - \varphi_T = F\left(\Xi_{fy} - \Xi_{ff}\right) - \frac{1}{4}F^2\Xi_{ff} \quad (14)$$

Ecuación 18:

$$\varphi = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}[y_i - (1-F)M_i f_i - FXf_i]^2$$

$$= \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[y_i^2 - 2(1-F)M_i f_i y_i - 2FXf_i y_i + (1-F)^2 M_i^2 f_i^2 + 2F(1-F)XM_i f_i^2 + F^2X^2 f_i^2\right]$$

(18)

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

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

El término para la ploidía materna, M_i, puede omitirse en algunas derivaciones matemáticas. La expresión resultante para X corresponde al caso especial relativamente simple, y a menudo que se produce más frecuentemente, cuando la madre no tiene deleciones o duplicaciones en el cromosoma o cromosomas que se evalúan.

Ecuación 21:

$$X = \frac{\Xi_{fy} - (1-F)\,\Xi_{ff}}{F\,\Xi_{ff}} = \frac{\Xi_{fy}}{F\,\Xi_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F}\left(\frac{\Xi_{fy}}{\Xi_{ff}} - 1\right) \quad (21)$$

Xiff y Xify vienen dados por las ecuaciones (11) y (12), respectivamente. En implementaciones donde todos los errores experimentales son despreciables, la resolución de la ecuación (21) da como resultado un valor de 1 para los euploides donde Xiff = Xify. En determinadas implementaciones donde todos los errores experimentales son despreciables, la resolución de la ecuación (21) da como resultado un valor de 3/2 para los triploides (véase la ecuación (15) para la relación triploide entre Xiff y Xify.

Tabla 2

Estado de gestación	**Cr21 fetal**	**Cr18 fetal**	**Cr13 fetal**	**CrX fetal**	**Cr fetal**
Mujer T21	P^F_{ij} = 3/2	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 0
Mujer T18	P^F_{ij} = 1	P^F_{ij} = 3/2	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 0
Mujer T13	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 3/2	P^F_{ij} = 1	P^F_{ij} = 0
Hombre T21	P^F_{ij} = 3/2	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1/2	P^F_{ij} = 1/2
Hombre T18	P^F_{ij} = 1	P^F_{ij} = 3/2	P^F_{ij} = 1	P^F_{ij} = 1/2	P^F_{ij} = 1/2
Hombre T13	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 3/2	P^F_{ij} = 1/2	P^F_{ij} = 1/2
Hombre euploide	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1/2	P^F_{ij} = 1/2
Turner	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1/2	P^F_{ij} = 0
Jacobs	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1/2	P^F_{ij} = 1
Klinefelter	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1/2
TripleX	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 1	P^F_{ij} = 3/2	P^F_{ij} = 0

Ejemplo 3: Enfoque híbrido

En este ejemplo, se utiliza un enfoque híbrido al combinar la normalización de GC aditiva y la selección de bins en función de la variabilidad del recuento. Para 1093 muestras euploides del estudio LDTv2CE, se calculó una DAM para cada uno de los 61927 bins. Se seleccionaron bins con DAM > 0 y DAM < 67,725 (cuantil del 99 %), lo que resultó en la identificación de un conjunto de bins estables. En este ejemplo, se seleccionó un conjunto de 53333 bins autosómicos (FIG. 1). No se realizó escalado según una escala a los recuentos totales.

Una vez identificados los bins estables, cada muestra se normalizó por separado para el sesgo de GC, independientemente de todas las demás muestras. Cuando un coeficiente R2 entre el contenido de GC por bin y los recuentos sin procesar medidos fue superior a 0,6 para una muestra, se aplicó una regresión lineal a los 53333 bins seleccionados para esa muestra. Para una muestra con un coeficiente R2 por debajo del valor de corte de 0,6, se aplicó un suavizado de LOESS a los 53333 bins seleccionados. En cualquier caso, la línea de regresión resultante (ya sea lineal o LOESS) se restó de los recuentos medidos para producir recuentos normalizados. La FIG. 12 y la FIG. 3 ilustran la aplicación del suavizado de LOESS a una muestra con un R2 < 0,6. En la FIG. 2 la variación de los recuentos con el contenido de GC fue no lineal, con un R2 < 0,6. Se aplicó el suavizado de LOESS a los datos de muestra y la FIG. 3 ilustra una representación gráfica de recuentos en función del contenido de GC por bin después de aplicar la corrección de LOESS aditiva. La FIG. 4 y la FIG. 5 ilustran la aplicación de la regresión lineal a una muestra con un R2 > 0,6. En la FIG. 4 la variación de recuentos con contenido de GC fue predominantemente lineal, con un R2 > 0,6. Se aplicó una regresión lineal a los datos de la muestra, y la FIG. 5 ilustra una representación gráfica de recuentos en función del contenido de GC por bin después de aplicar la corrección lineal aditiva.

Se compararon las representaciones de puntuaciones Z para los cromosomas 21, 18 y 13 determinadas usando un enfoque de corrección aditiva híbrido como se ha descrito anteriormente con las representaciones de las puntuaciones Z obtenidas mediante un enfoque de PERUN. Las representaciones de las puntuaciones Z para el cromosoma 21 se derivaron de un conjunto de datos de LDTv2CE (FIG. 6-2). Las representaciones de las puntuaciones Z para el cromosoma 18 se derivaron de un conjunto de datos de precisión clínica (FIG. 8-9) y las representaciones de las puntuaciones Z para el cromosoma 13 se derivaron de un conjunto de datos de validación Clia (FIG. 10-11). Las representaciones de las puntuaciones Z determinadas mediante PERUN se muestran en los paneles superiores y las representaciones de las puntuaciones Z determinadas usando una corrección aditiva híbrida se muestran en los paneles inferiores. Encima de cada panel, se muestran las distancias entre las puntuaciones Z. El enfoque aditivo híbrido utilizó un conjunto preseleccionado de 53333 bins.

Para comparar la selección de bins híbridos aditivos con la selección de bins de PERUN, se realizó la normalización de GC de LOESS en un conjunto de 50034 bins de PERUN de validación cruzada. Los resultados se muestran en la FIG. 12 (PERUN) y la FIG. 13 (corrección de GC híbrida aditiva).

Los enfoques de la puntuación Z de PERUN e híbrido aditivo se compararon más directamente en la FIG. 14 (Cr. 21, datos de LDTv2CE), la FIG. 15 (Cr. 21, Datos de precisión clínica) y la FIG. 16 (Cr. 21, Datos de validación CLIA).

Ejemplo 4

El PERUN lineal, en algunas implementaciones, puede ser unidimensional en el sentido de que utilizó un solo descriptor específico de la muestra, en concreto, un coeficiente de sesgo de GC lineal, para corregir los sesgos sistemáticos en los datos de la muestra. En algunas muestras, el perfil de recuentos en función del contenido de GC era extremadamente curvado. En dichos casos, fue beneficioso modelar la dependencia de los recuentos en función del contenido de GC utilizando una ecuación curvilínea. Los ejemplos de ecuaciones curvilíneas incluyeron expresiones polinómicas, racionales o más generales, como ecuaciones trascendentales. En todos estos casos, se necesitaron coeficientes múltiples para describir la dependencia de los recuentos en función del contenido de GC. En consecuencia, el tratamiento de PERUN correspondiente necesitaba expandirse a múltiples dimensiones, asignándose a cada coeficiente una dimensión separada. Este ejemplo describe una versión multidimensional de PERUN e ilustra los principios que subyacen al PERUN multidimensional al presentar una versión bidimensional de PERUN. Esta implementación particular utiliza polinomios de segundo grado. Un PERUN bidimensional puede extenderse a otras formas funcionales y a un mayor número de dimensiones mediante un método adecuado.

Los recuentos sin procesar se representan en la ecuación (30) como una función cuadrática del contenido de GC:

$$c_i = G_0 + G_1 g_i + G_2 g_i^2 \qquad (30)$$

El término c_i representaba los recuentos sin procesar observados en el bin i, divididos entre los recuentos autosómicos totales. El término g_i era el contenido de GC del bin i. G_0, G_1 y G_2 fueron los coeficientes de regresión de cero, primer y segundo grado, respectivamente. La ec. 30 se generalizó a la siguiente expresión:

$$c_i = \sum_{n=0}^{N} G_n g_i^n \qquad (31)$$

N de la ec. 31 representaba el nivel de truncamiento. La ec. 30 describía adecuadamente incluso las muestras con la curvatura más pronunciada.

Se utilizaron procedimientos de regresión estándar para evaluar G_0, G_1 y G_2 para muestras utilizadas para entrenar los parámetros de PERUN, así como para una muestra que necesitaba ser normalizada.

Cuando se obtuvieron los coeficientes de regresión lineal y cuadrática de la dependencia de los recuentos en función del contenido de GC en un gran número de muestras, se hizo evidente que sus valores estaban correlacionados.

Parametrización de PERUN cuadrática y selección de bins

Para un solo bin, una gran cantidad de muestras de referencia proporcionaron los valores de los coeficientes de regresión G_0, G_1 y G_2. Las mismas muestras también proporcionaron recuentos de bins sin procesar (divididos entre los recuentos autosómicos totales) para el bin seleccionado. Dentro del bin seleccionado, el PERUN cuadrático en su forma más simple adoptó la siguiente relación entre los recuentos de bins específicos de la muestra y los coeficientes de regresión específicos de la muestra G_0, G_1 y G_2:

$$c_i = G_0 + m_0 + G_1 m_1 + G_2 m_2 \qquad (32)$$

Los coeficientes m_0, m_1 y m_2 eran específicos del bin e independientes de la muestra. Estos parámetros de PERUN se extrajeron para cada bin de un conjunto de referencia mediante regresión lineal.

El PERUN cuadrático, formulado mediante la ec. 32, ignoró la fuerte correlación que se sabe que existe entre los coeficientes de regresión G_1 y G_2. Por esta razón, se construyó una versión alternativa del PERUN cuadrático. Este PERUN cuasicuadrático aprovechó una relación conocida entre los coeficientes lineales y cuadráticos G_1 y G_2:

$$G_2 = K_0 + G_1 K_2 \qquad (33)$$

Los coeficientes K_0 y K_2 se obtuvieron mediante regresión lineal de los coeficientes G_1 y G_2 para un gran conjunto de muestras de referencia. La combinación de las ec. 32-33 produjo la siguiente versión casi-cuadrática de PERUN:

$$c_i = G_0 + m_0 + G_1 m_1 + (K_0 + G_1 K_2) m_2 = G_0 + a_0 + G_1 a_1 \quad (38)$$

A continuación, se proporciona la relación entre el conjunto cuasicuadrático de parámetros de PERUN (a_0, a_1) y el conjunto cuadrático de parámetros de PERUN (m_0, m_1, m_2).

$$a_0 = m_0 + K_0 m_2 \quad (36)$$

$$a_1 = m_1 + K_2 m_2 \quad (37)$$

Una tercera alternativa transformó los coeficientes de regresión G_1 y G_2 en un nuevo conjunto de coordenadas generalizadas X_1 y X_2 utilizando una transformación canónica. En primer lugar, se sustrajo un centroide de los puntos de datos del plano G_1/G_2 de los valores de G_1 y G_2 individuales. A continuación, se evaluó la matriz de covarianza para los valores de G_1 y G_2 medidos en las muestras de referencia. A continuación, se diagonalizó la matriz de covarianza y se registraron sus vectores propios y valores propios. Las nuevas coordenadas canónicas X_1 y X_2 se obtuvieron como elementos de los vectores propios de la matriz de covarianza, divididos entre las raíces cuadradas de los valores propios correspondientes. Estas nuevas coordenadas eran ortogonales porque la matriz de covarianza era real y simétrica (un caso especial de matrices hermitianas). Por la misma razón, los valores propios de la matriz de covarianza eran reales. Dado que la matriz de covarianza fue positivamente definida, los valores propios fueron positivos. Por lo tanto, la división entre la raíz cuadrada de valores propios produjo números reales. Dado que la expansión de la dependencia de G_1 en función de G_2 era finita, los valores propios eran distintos de cero. A continuación, se utilizaron las coordenadas canónicas X_1 y X_2 para definir la versión canónica de PERUN cuadrático:

$$c_i = G_0 + \mu_0 + X_1 \mu_1 + X_2 \mu_2 \quad (39)$$

Los parámetros de PERUN canónicos independientes de la muestra y específicos del bin μ_0, μ_1, y μ_2 se relacionaron con el conjunto cuadrático de parámetros de PERUN (m_0, m_1, m_2) por la inversa de la transformación de coordenadas lineales utilizada para generar las coordenadas canónicas X_1 y X_2. μ_0, μ_1 y μ_2 se evaluaron aplicando regresión lineal a un gran conjunto de muestras de referencia.

En las tres versiones de PERUN ampliado (PERUN cuadrático, PERUN cuasicuadrático y PERUN canónico), se utilizaron los parámetros de los bin, una vez evaluados, para seleccionar bien fiables mediante validación cruzada. Opcionalmente, los bins que sobrevivieron a la validación cruzada se pueden filtrar adicionalmente utilizando medidas de capacidad de mapeo y repetibilidad específicas del bin.

Normalización PERUN cuadrática

Se utilizó la siguiente expresión de PERUN cuadrático para normalizar un conjunto de datos recién medido:

$$l_i = \frac{c_i}{G_0 + m_0 + G_1 m_1 + G_2 m_2} \quad (40)$$

El término l_i fue el recuento de bins normalizado, el resultado final de PERUN. La versión cuasicuadrática de PERUN utilizó el siguiente procedimiento de normalización:

$$l_i = \frac{c_i}{G_0 + a_0 + G_1 a_1} \quad (41)$$

Finalmente, el PERUN canónico usó la siguiente expresión para normalizar los recuentos:

$$l_i = \frac{c_i}{G_0 + \mu_0 + X_1 \mu_1 + X_2 \mu_2} \quad (42)$$

Reescalado de perfiles de PERUN

Para eliminar cualquier variabilidad residual específica de la muestra causada por diferencias biológicas (ploidía, duplicaciones/deleciones), se reescalaron los perfiles normalizados adicionalmente. Un procedimiento de reescalado evaluó la mediana y la DAM de la porción autosómica de un perfil normalizado. Los bins de valores atípicos fueron identificados y marcados. El criterio para marcar un bin como un valor atípico fue su desviación de la mediana de los recuentos normalizados. Si esa desviación era superior a tres DAM, el bin se marcaba como un valor atípico. A continuación, se evaluó la mediana de los bins restantes y el perfil se dividió entre ese segundo

valor de la mediana. El reescalado normalizó todos los bins euploides a un nivel de aproximadamente uno. El procedimiento de reescalado minimizó el efecto de cualquier aneuploidía en el nivel de porciones euploides del genoma. El reescalado fue igualmente aplicable a todas las versiones de PERUN, incluida la versión lineal.

El término "un(o)" o "una" puede referirse a uno o a una pluralidad de los elementos que modifica (por ejemplo, "un reactivo" puede significar uno o más reactivos) a menos que se aclare contextualmente que se describe o bien uno de los elementos o bien más de uno de los elementos. El término "aproximadamente", tal como se usa en el presente documento, se refiere a un valor dentro de 10 % del parámetro subyacente (es decir, más o menos el 10 %), y el uso del término "aproximadamente" al comienzo de una cadena de valores modifica cada uno de los valores (es decir, "aproximadamente 1, 2 y 3" se refiere a aproximadamente 1, aproximadamente 2 y aproximadamente 3). Por ejemplo, un peso de "aproximadamente 100 gramos" puede incluir pesos entre 90 gramos y 110 gramos.

Determinadas realizaciones de la tecnología se exponen en la(s) reivindicación/reivindicaciones que sigue(n).

REIVINDICACIONES

1. Un método para determinar la presencia o ausencia de una trisomía cromosómica, que comprende:

(a) obtener recuentos de las lecturas de secuencia mapeadas en porciones de un genoma de referencia, lecturas de secuencia que son lecturas de ácido nucleico circulante extracelular de una muestra de prueba de una mujer embarazada;
(b) generar una regresión para (i) los recuentos y (ii) el contenido de guanina y citosina (GC), para cada una de las porciones del genoma de referencia de la muestra de prueba;
(c) evaluar la bondad de ajuste de los recuentos y del contenido de GC a una regresión no lineal o una regresión lineal, que comprende determinar un coeficiente de correlación de la regresión generada en (b), generando de esta manera una evaluación, en donde la evaluación se determina según el coeficiente de correlación y un valor de corte del coeficiente de correlación, en donde:

(i) el coeficiente de correlación es igual o superior al valor de corte del coeficiente de correlación y la evaluación es indicativa de una regresión lineal, o
(ii) el coeficiente de correlación es inferior al valor de corte del coeficiente de correlación y la evaluación es indicativa de una regresión no lineal;

d) normalizar los recuentos mediante un proceso seleccionado según la evaluación, que comprende:

(i) en los casos en los que la evaluación sea indicativa de una regresión lineal, restar la regresión lineal a los recuentos, o
(ii) en los casos en los que la evaluación sea indicativa de una regresión no lineal, restar la regresión no lineal a los recuentos,

generando de esta manera recuentos normalizados con sesgo de GC reducido; y
(e) determinar la presencia o ausencia de una trisomía cromosómica según los recuentos normalizados.

2. El método de la reivindicación 1, en donde la regresión no lineal se realiza mediante un proceso de LOESS.

3. El método de las reivindicaciones 1 o 2, en donde el valor de corte del coeficiente de correlación es de 0,5 a 0,7.

4. El método de la reivindicación 3, en donde el valor de corte del coeficiente de correlación es 0,6.

5. El método de una cualquiera de las reivindicaciones 1 a 4, que comprende, antes de (a):

(i) determinar un valor de incertidumbre para los recuentos mapeados de cada una de las porciones para múltiples muestras de prueba; y
(ii) seleccionar un subconjunto de porciones que tengan un valor de incertidumbre dentro de un intervalo predeterminado de valores de incertidumbre, reteniendo de esta manera las porciones seleccionadas; donde (a) a (c) se realizan utilizando las porciones seleccionadas.

6. El método de la reivindicación 5, en donde el valor de incertidumbre es una desviación estándar, un error estándar, un error absoluto medio (EAM), una desviación absoluta promedio o una mediana de la desviación absoluta (MDA).

7. El método de la reivindicación 5, en donde el valor de incertidumbre es una mediana de la desviación absoluta (MDA).

8. El método de la reivindicación 5, en donde las porciones seleccionadas están en al menos el cuantil del 99 % de la variabilidad del recuento.

9. El método de una cualquiera de las reivindicaciones 1 a 8, en donde cada porción del genoma de referencia comprende una secuencia de nucleótidos de una longitud predeterminada de aproximadamente 50 kilobases.

10. El método de una cualquiera de las reivindicaciones 1 a 9, en donde la trisomía es la trisomía 21, la trisomía 18 o la trisomía 13.

11. El método de una cualquiera de las reivindicaciones 1 a 10, que comprende, antes de (a), secuenciar el ácido nucleico de la muestra de prueba, proporcionando de esta manera lecturas de secuenciación, mapear las lecturas de secuenciación en las porciones del genoma de referencia y realizar el recuento de las lecturas mapeadas en las porciones.

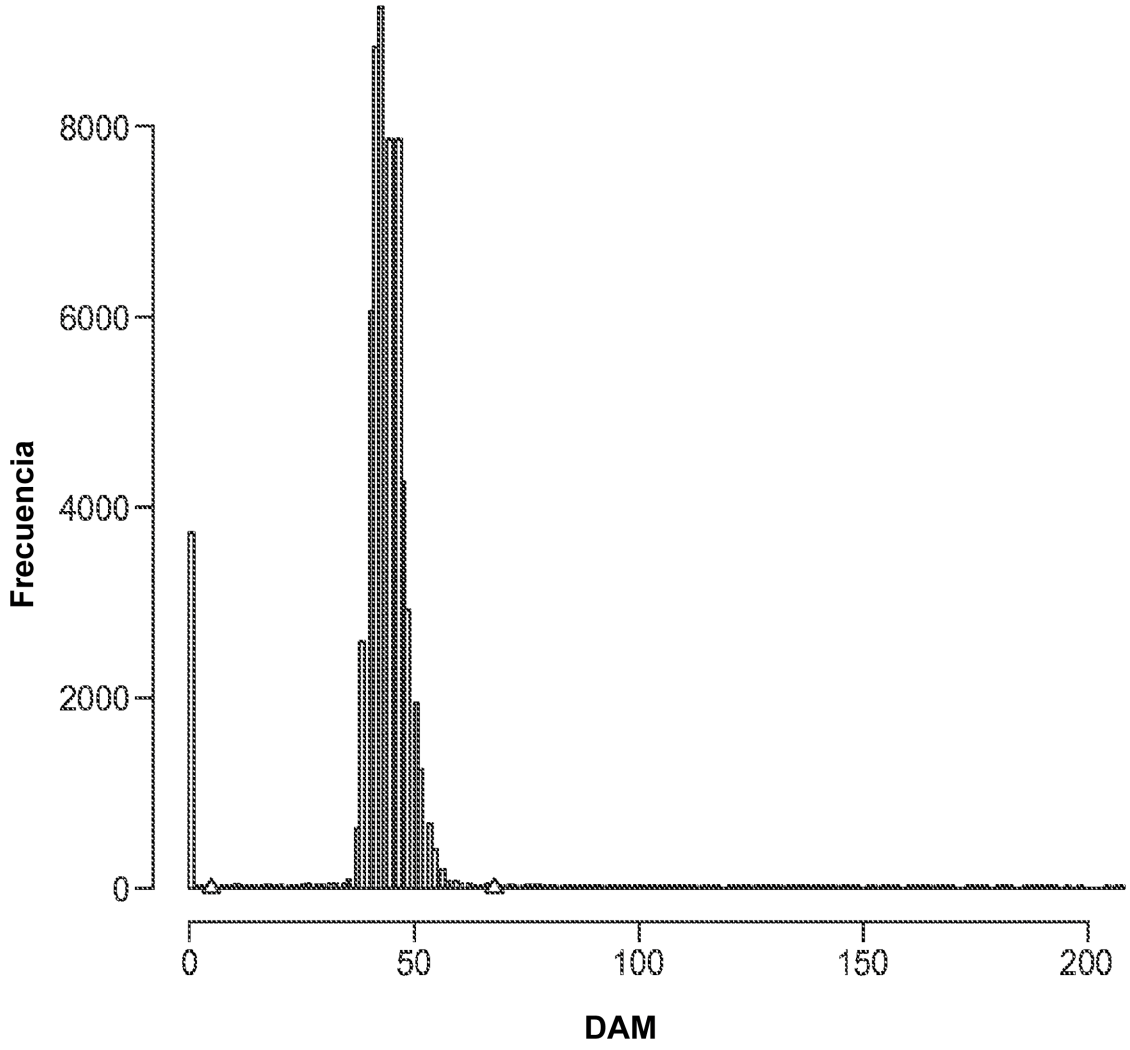


FIG. 1

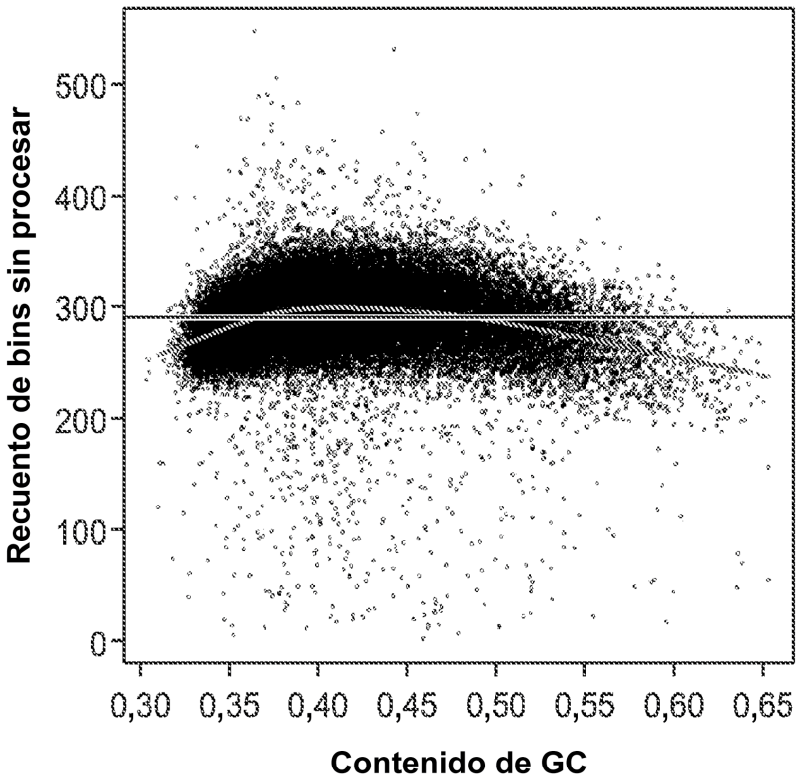

Recuento de bins sin procesar
500
400
300
200
100
0
0,30
0,35
0,40
0,45
0,50
0,55
0,60
0,65
Contenido de GC

FIG. 2

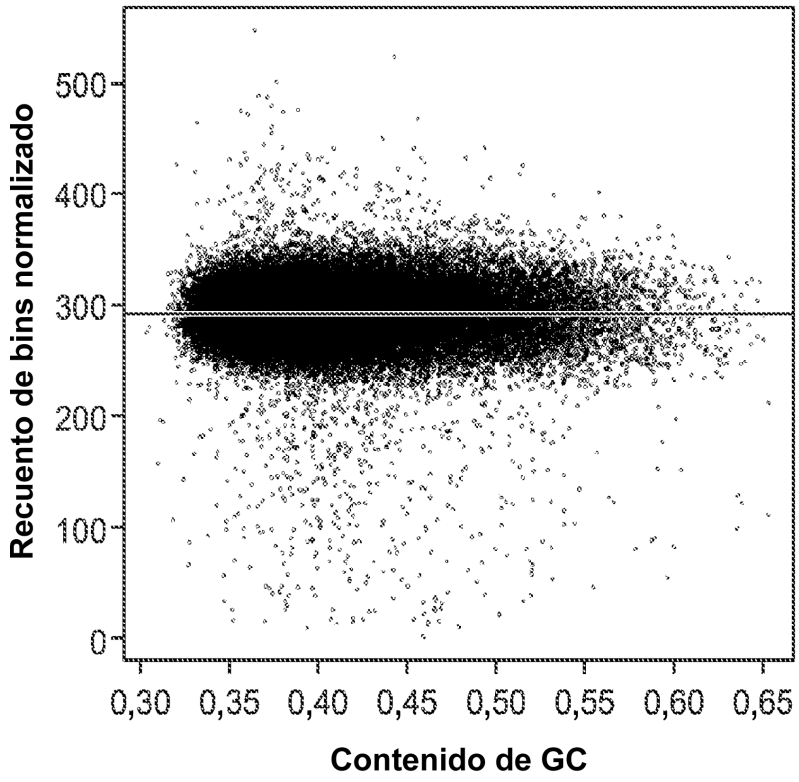

Recuento de bins normalizado
500
400
300
200
100
0
0,30
0,35
0,40
0,45
0,50
0,55
0,60
0,65
Contenido de GC

FIG. 3

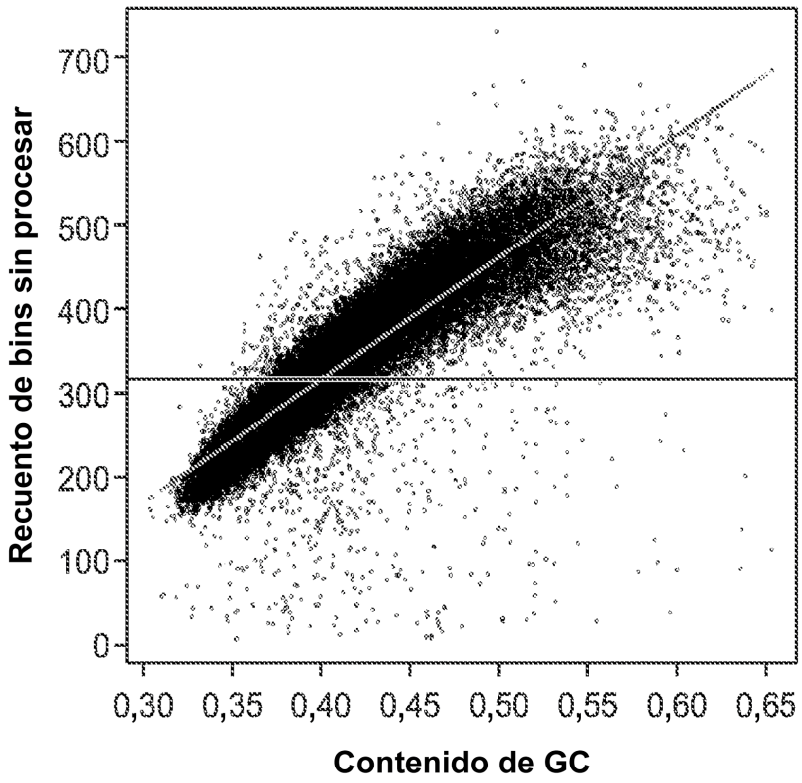

Recuento de bins sin procesar
700
600
500
400
300
200
100
0
0,30
0,35
0,40
0,45
0,50
0,55
0,60
0,65
Contenido de GC

FIG. 4

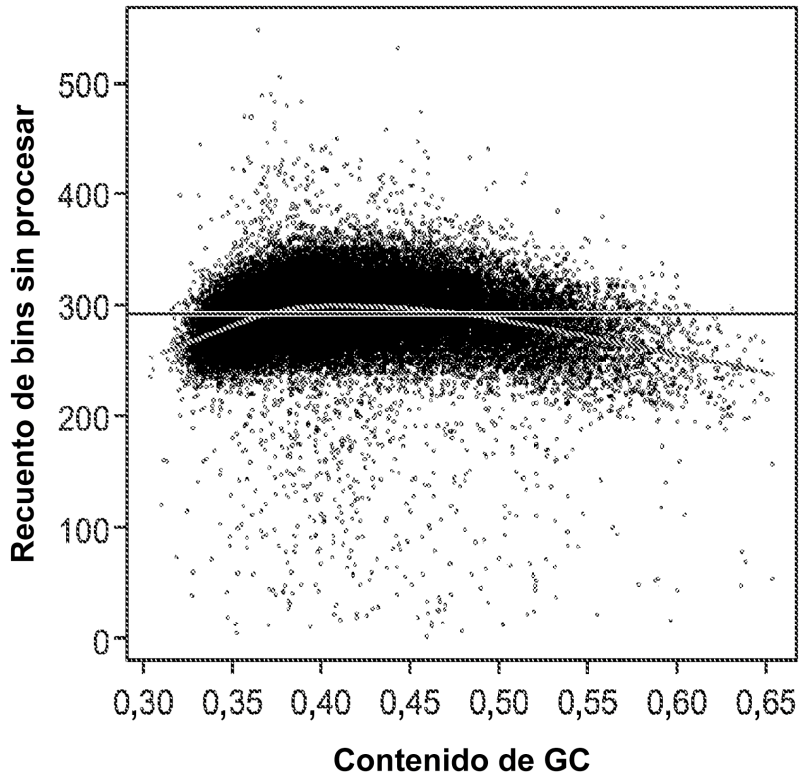

Recuento de bins sin procesar
500
400
300
200
100
0
0,30
0,35
0,40
0,45
0,50
0,55
0,60
0,65
Contenido de GC

FIG. 5

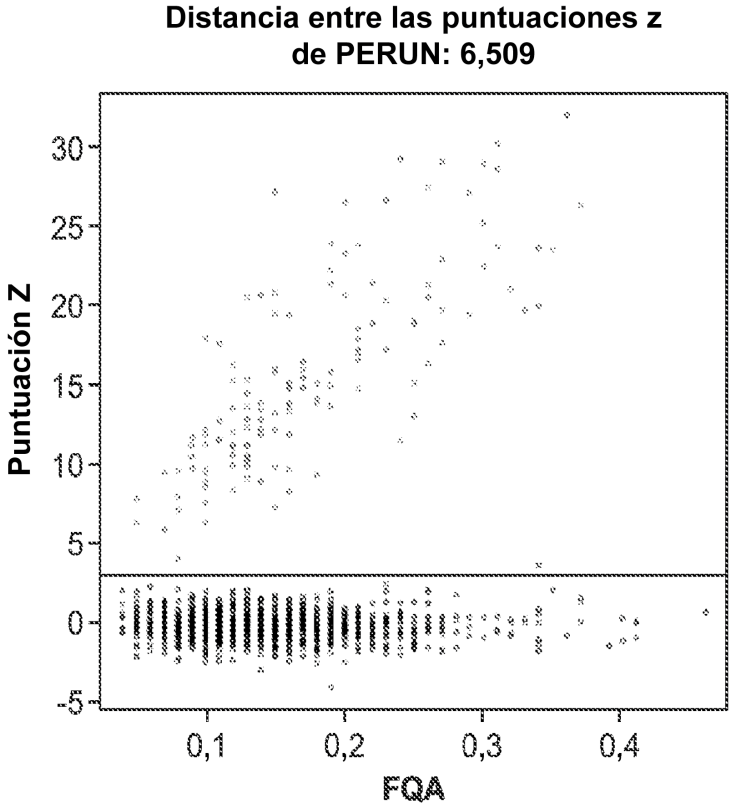

Distancia entre las puntuaciones z
de PERUN: 6,509
Puntuación Z
30
25
20
15
10
5
0
-5
0,1
0,2
0,3
0,4
FQA

FIG. 6

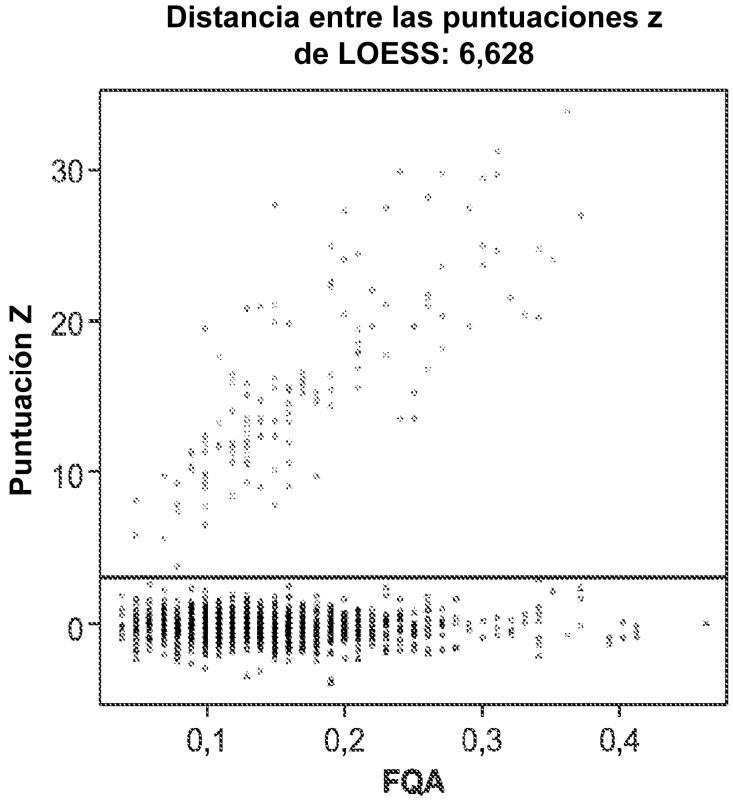

Distancia entre las puntuaciones z
de LOESS: 6,628
Puntuación Z
30
20
10
0
0,1
0,2
0,3
0,4
FQA

FIG. 7

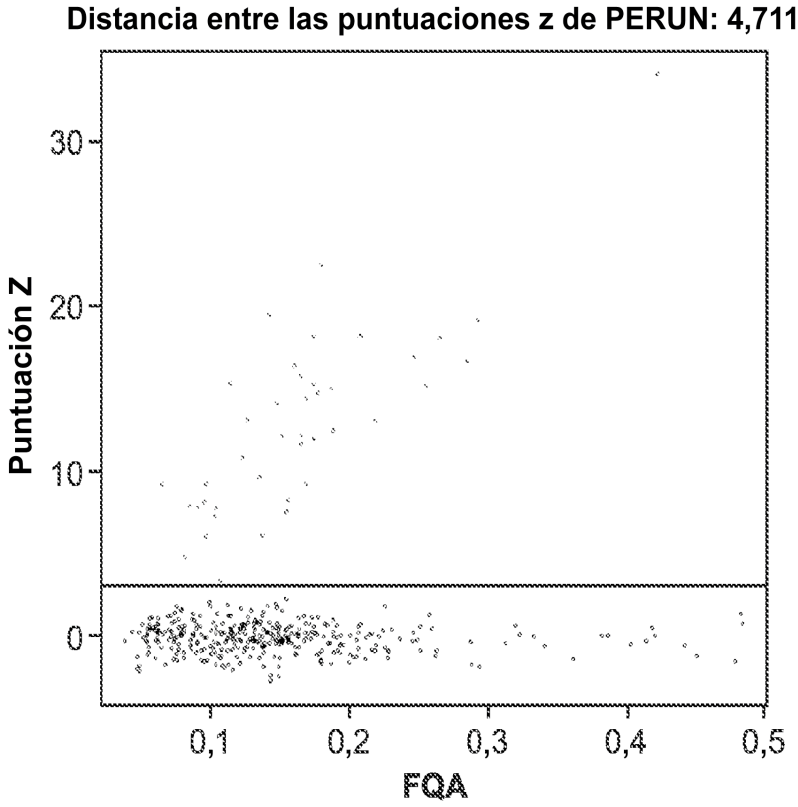

Distancia entre las puntuaciones z de PERUN: 4,711
Puntuación Z
30
20
10
0
0,1
0,2
0,3
0,4
0,5
FQA

FIG. 8

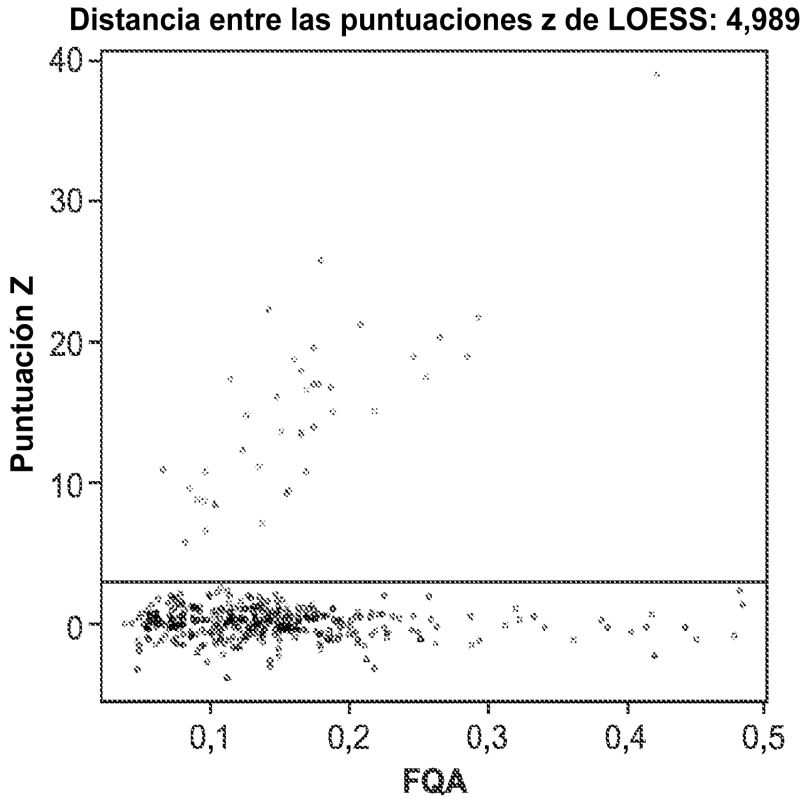

Distancia entre las puntuaciones z de LOESS: 4,989
Puntuación Z
40
30
20
10
0
0,1
0,2
0,3
0,4
0,5
FQA

FIG. 9

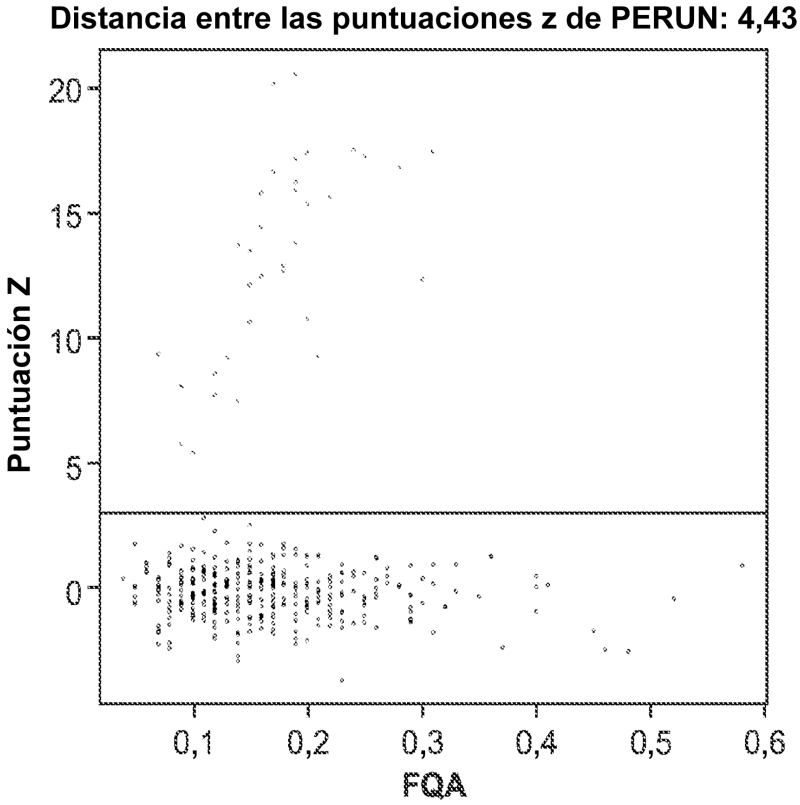

Distancia entre las puntuaciones z de PERUN: 4,431
Puntuación Z
20
15
10
5
0
0,1
0,2
0,3
0,4
0,5
0,6
FQA

FIG. 10

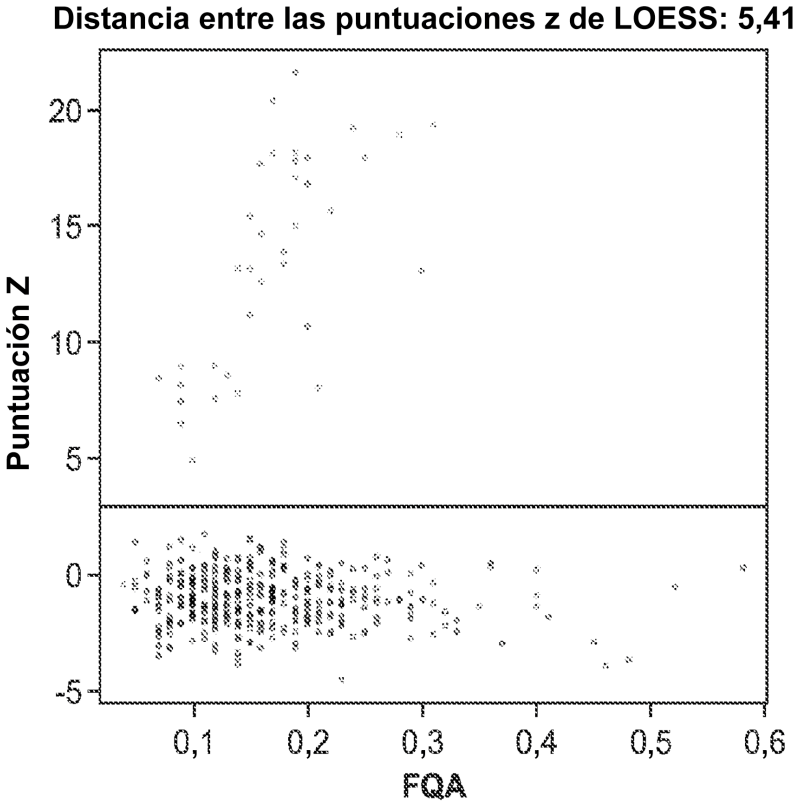

Distancia entre las puntuaciones z de LOESS: 5,419
Puntuación Z
20
15
10
5
0
-5
0,1
0,2
0,3
0,4
0,5
0,6
FQA

FIG. 11

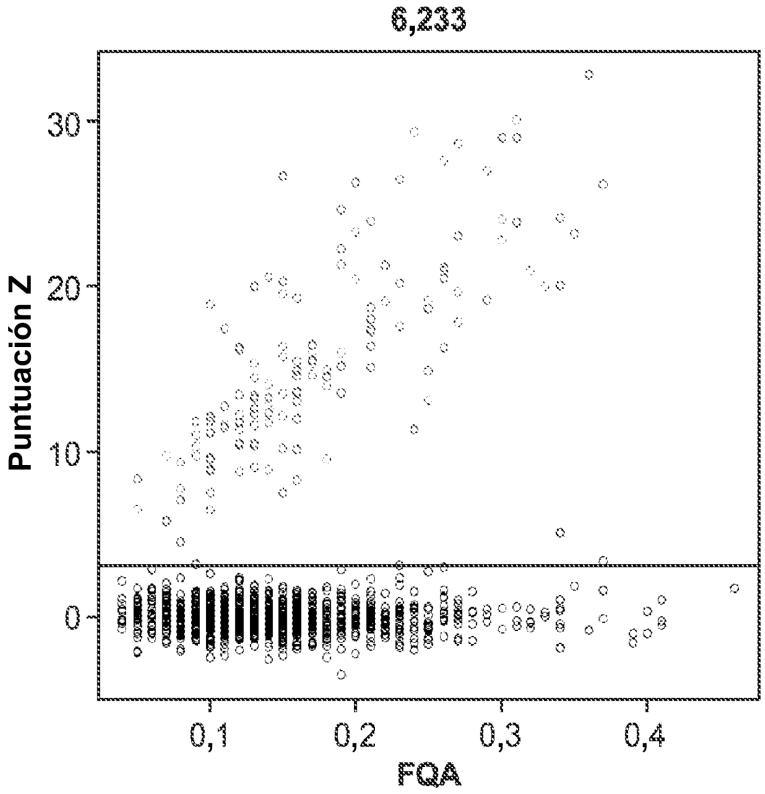

6,233
Puntuación Z
30
20
10
0
0,1
0,2
0,3
0,4
FQA

FIG. 12

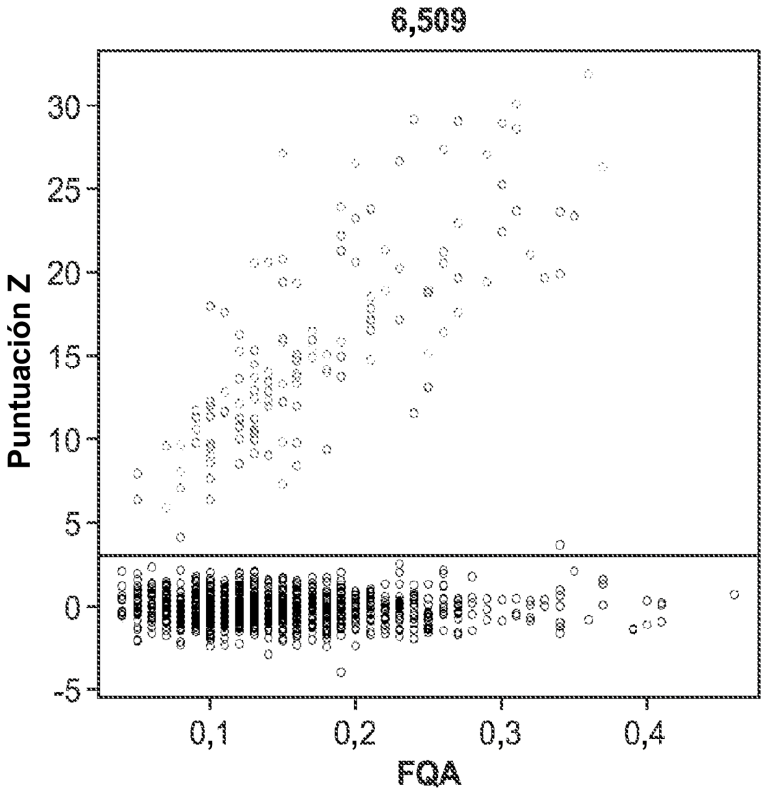

6,509
Puntuación Z
30
25
20
15
10
5
0
-5
0,1
0,2
0,3
0,4
FQA

FIG. 13

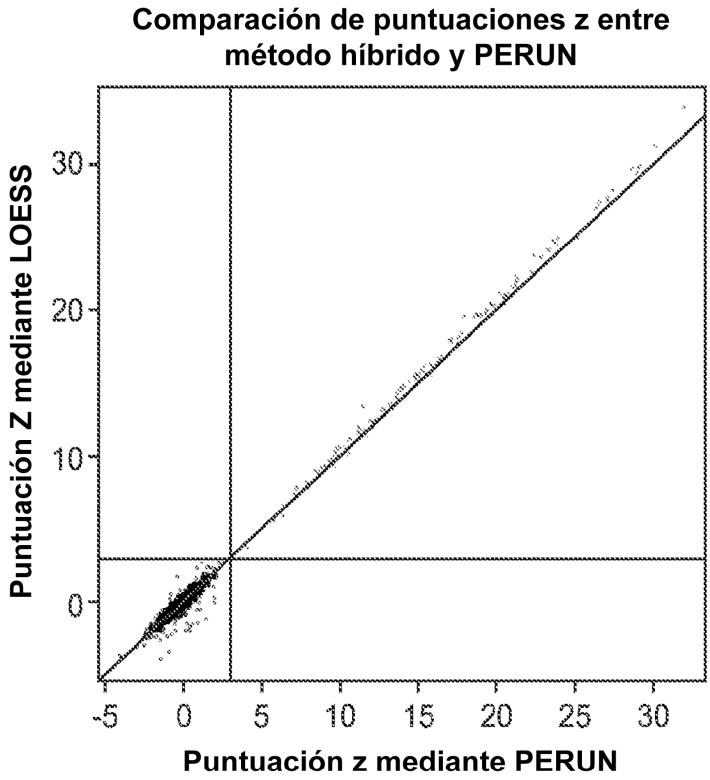

Comparación de puntuaciones z entre método híbrido y PERUN
Puntuación Z mediante LOESS
30
20
10
0
-5
0
5
10
15
20
25
30
Puntuación z mediante PERUN

FIG. 14

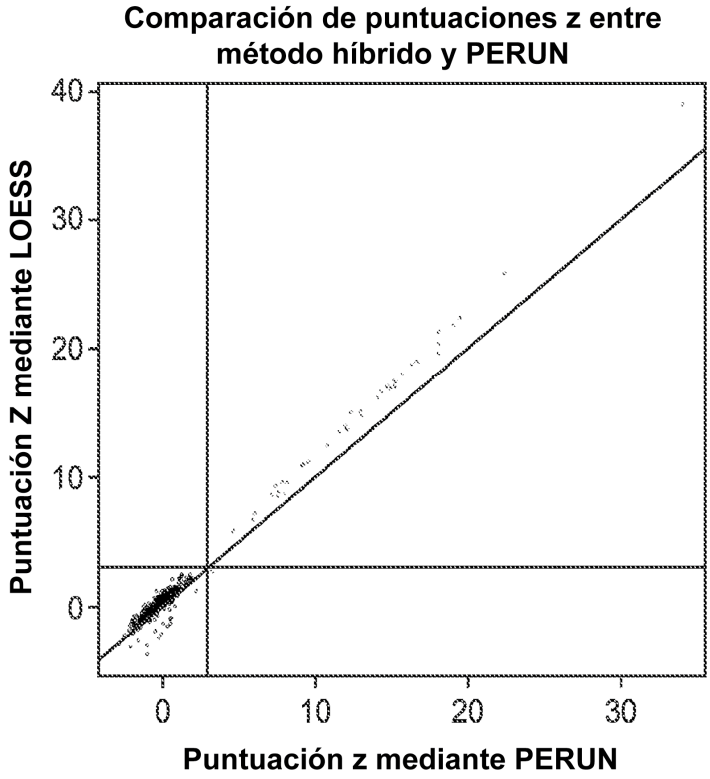

Comparación de puntuaciones z entre método híbrido y PERUN
Puntuación Z mediante LOESS
40
30
20
10
0
0
10
20
30
Puntuación z mediante PERUN

FIG. 15

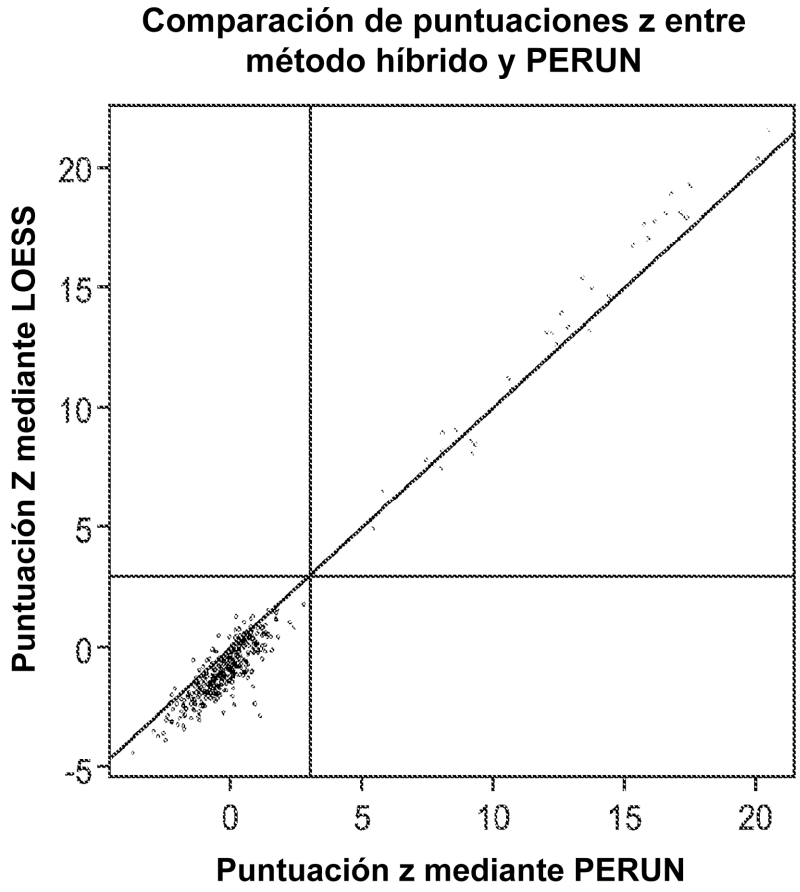
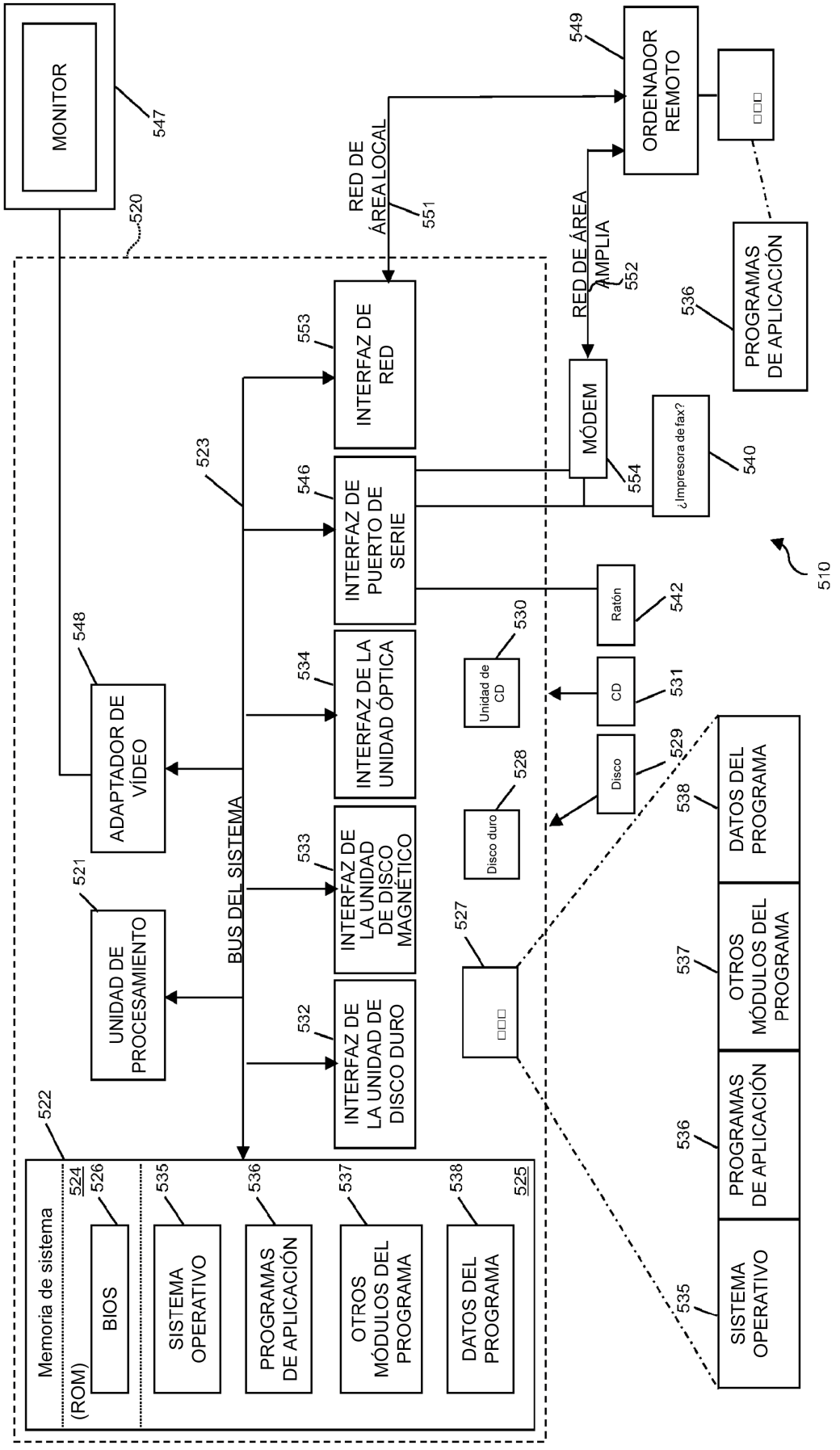

Comparación de puntuaciones z entre método híbrido y PERUN
Puntuación Z mediante LOESS
20
15
10
5
0
-5
0
5
10
15
20
Puntuación z mediante PERUN

FIG. 16

FIG. 17


MONITOR
547
520
549
ORDENADOR REMOTO
RED DE ÁREA LOCAL
551
RED DE ÁREA AMPLIA
552
INTERFAZ DE RED
553
INTERFAZ DE PUERTO DE SERIE
546
523
MÓDEM
554
¿Impresora de fax?
540
PROGRAMAS DE APLICACIÓN
536
548
ADAPTADOR DE VÍDEO
534
INTERFAZ DE LA UNIDAD ÓPTICA
533
INTERFAZ DE LA UNIDAD DE DISCO MAGNÉTICO
532
INTERFAZ DE LA UNIDAD DE DISCO DURO
521
UNIDAD DE PROCESAMIENTO
BUS DEL SISTEMA
Unidad de CD
530
CD
531
Disco
529
Disco duro
528
Ratón
542
527
510
522
Memoria de sistema
(ROM)
524
526
BIOS
535
SISTEMA OPERATIVO
536
PROGRAMAS DE APLICACIÓN
537
OTROS MÓDULOS DEL PROGRAMA
538
DATOS DEL PROGRAMA
525
DATOS DEL PROGRAMA
538
OTROS MÓDULOS DEL PROGRAMA
537
PROGRAMAS DE APLICACIÓN
536
SISTEMA OPERATIVO
535

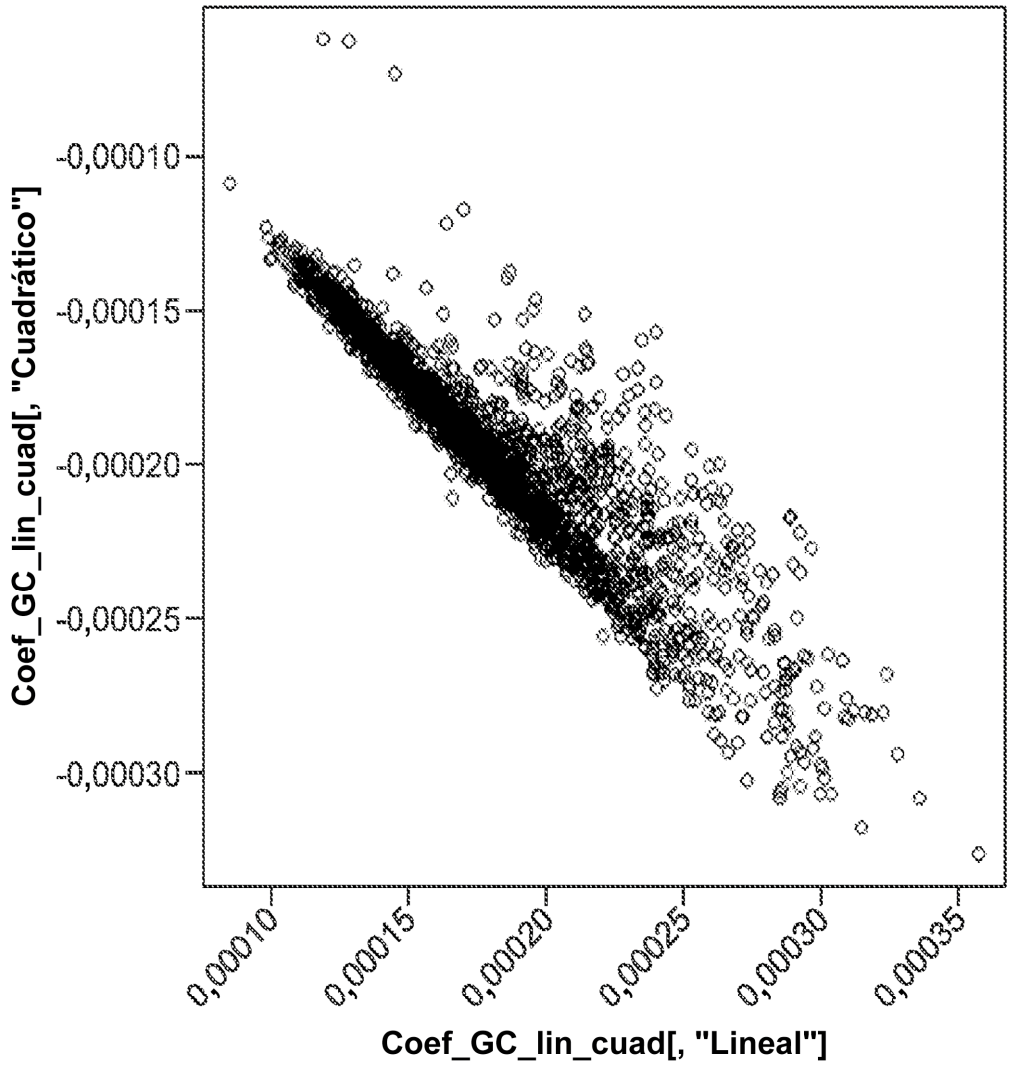

-0,00010
-0,00015
-0,00020
-0,00025
-0,00030
Coef_GC_lin_cuad[, "Cuadrático"]
0,00010
0,00015
0,00020
0,00025
0,00030
0,00035
Coef_GC_lin_cuad[, "Lineal"]

FIG. 18