

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 701 775**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.02.2012 PCT/CN2012/071648**

87 Fecha y número de publicación internacional: **06.09.2013 WO13127049**

96 Fecha de presentación y número de la solicitud europea: **27.02.2012 E 12870229 (7)**

97 Fecha y número de publicación de la concesión europea: **19.09.2018 EP 2821501**

54 Título: **Procedimiento y dispositivo para detectar microdelección en el área del cromosoma STS**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.02.2019

73 Titular/es:
**BGI GENOMICS CO., LTD. (100.0%)
Floors 7-14, Building No. 7, BGI Park, No. 21
Hongan 3rd Street, Yantian District
Guangdong Province 518083, CN**

72 Inventor/es:
**LIU, XIAO;
ZHANG, JUNJIE;
XU, HUIQIAN;
SU, ZHENG;
ZHANG, RUIFANG;
WANG, JUN;
WANG, JIAN y
YANG, HUANMING**

74 Agente/Representante:
CARPINTERO LÓPEZ, Mario

ES 2 701 775 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento y dispositivo para detectar microdelección en el área del cromosoma STS

Campo técnico

5 Las realizaciones de la presente divulgación se refieren en general a tecnología de ingeniería genética, más particularmente a un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma y un aparato del mismo.

Antecedentes

La delección es un fenómeno de que se produce una pérdida parcial de cromosoma o moléculas de ADN en un genoma, que es la razón más importante que lleva a la mutación genética.

10 En la actualidad, la técnica de PCR (reacción en cadena de la polimerasa) se usa normalmente en la detección de una microdelección en una región de sitio de secuencia marcada (STS) de un cromosoma. La técnica de PCR es un procedimiento de amplificación selectiva de determinada región especial de ADN in vitro mediante la estimulación de un patrón de réplica de ADN in vivo. El uso de la técnica de PCR tiene ventajas de rapidez y conveniencia cuando se detecta una pequeña cantidad de sitios; además de que requiere conocer las secuencias en ambos extremos de una determinada región de ADN de antemano para el diseño de la sonda, mediante la cual requiere que estas microdelecciones en la región de sitio de secuencia marcada del cromosoma deben darse a conocer por adelantado.

15 El documento KR 2009 0112235 A desvela un ensayo de cribado de infertilidad idiopática en el que se proporciona un chip de microdelección de cromosoma Y humano para analizar rápidamente 19 tipos de marcador de STS y cinco genes para detectar la microdelección de sitios de AZFa, AZFb y AZFc. El ensayo de cribado comprendía: una etapa de concepción de un chip en el que se reconocen 25 tipos de sondas; una etapa de aislamiento de ADN genómico mediante perla magnética a partir de sangre o semen; una etapa de amplificación de nueve tipos de AZFa, diez tipos de AZFb, y nueve tipos de AZFc de marcador de STS o gen mediante PCR; una etapa de introducción de hibridación entre las sondas; y una etapa de cribado del chip y medición de la intensidad de fluorescencia.

20 Sin embargo, cuando se enfrenta a una gran cantidad de muestras o microdelecciones no informadas, la detección usando la técnica PCR puede tener una enorme limitación; además, cuando la región de sitio de secuencia marcada a detectar que tiene un número relativamente superior de sitios o, particularmente, cuando se detecta, la detección tradicional usando la técnica PCR ha sido incapaz de cumplir tal requisito, de este modo, se requiere adicionalmente un procedimiento novedoso de investigación.

Sumario

30 Un problema técnico principal solucionado por la presente divulgación es proporcionar un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma y un aparato del mismo, que pueda detectar la microdelección en la región de sitio de secuencia marcada del cromosoma cuando se enfrenta a una gran cantidad de muestra con un coste limitado y también pueda detectar microdelecciones no informadas en la región de sitio de secuencia marcada del cromosoma.

35 Para solucionar el anterior problema técnico, una solución técnica usada por la presente divulgación es: proporcionar un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma, que comprende:

- 40 obtener una sonda de captura que se corresponde con secuencia de ADN de la región de sitio de secuencia marcada;
- hibridar la sonda de captura con una biblioteca de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN;
- secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación; y
- 45 analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

Preferentemente, la etapa de análisis de los datos de secuenciación usando el ensayo estadístico matemático comprende adicionalmente:

- 50 someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de profundidad de secuenciación normalizado; y
- detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

Preferentemente, la etapa de someter el valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización comprende adicionalmente:

5 dividir todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.

Preferentemente, la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

10 calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 15 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico; y
 obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

20 Preferentemente, en el que después de la etapa de obtener el resultado R1, el procedimiento comprende adicionalmente:

someter el resultado R1 a verificación experimental, para obtener un resultado de verificación;
 determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
 25 obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al segundo valor de corte de probabilidad.

Preferentemente, en el que la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

30 calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una media de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3;
 35 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es superior al primer valor de corte de relación;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y
 40 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

Preferentemente, la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

50 calcular un valor medio y una varianza basándose en todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 55 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico;
 obtener un resultado R3 que muestra la presencia de la microdelección en la región dada, si la probabilidad de

una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;
 calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;
 calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;
 formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación; y
 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

Preferentemente, la etapa de obtener la sonda de captura comprende adicionalmente:

encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica;
 seleccionar una secuencia de ADN cualificada de la región de sitio de secuencia marcada que cumple el requisito para designar la sonda de captura; y
 designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.

Preferentemente, la biblioteca de la pluralidad de muestras de ADN se construye mediante las siguientes etapas:

construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;
 mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada; y
 determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.

Preferentemente, en la etapa de construcción de una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, para cada una de la pluralidad de muestras de ADN únicas, la biblioteca de ADN se construye mediante la siguientes etapas:

romper un genoma de ADN en fragmentos de ADN que tienen un tamaño predeterminado mediante procedimiento físico o químico, para obtener fragmentos de ADN;
 someter los fragmentos de ADN a reparación de extremo, para obtener fragmentos de ADN con extremo reparado con un extremo fosforilado;
 añadir una base "A" a los fragmentos de ADN con extremo reparado en el extremo 3' para obtener fragmentos de ADN con la base "A" en el extremo 3';
 ligar un adaptador de índice a los fragmentos de ADN con la base "A" en el extremo 3', para obtener fragmentos de ADN ligados con el adaptador de índice;
 amplificar los fragmentos de ADN ligados con el adaptador de índice usando un cebador que comprende una secuencia específica al adaptador de índice, para obtener un producto de amplificación; y
 determinar el producto de amplificación como la biblioteca de ADN, si el producto de amplificación cumple el requisito de control de calidad.

Preferentemente, después de la etapa de secuenciación de la secuenciación del ADN capturado y antes de la etapa de análisis de los datos de secuenciación, el procedimiento comprende adicionalmente:
 someter los datos de secuenciación a un control de calidad.

Preferentemente, la etapa de someter de los datos de secuenciación a control de calidad comprende adicionalmente:

filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados;
 alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;
 filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación

calculado, para obtener segundos datos de secuenciación cualificados; y
 filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación
 calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada
 dada.

5 Preferentemente, la etapa de filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener
 los primeros datos de secuenciación cualificados que comprenden adicionalmente:

obtener un primer conjunto de datos de secuenciación filtrando datos no cualificados que contiene más de una
 proporción predeterminada de bases que tienen una baja calidad de los datos de secuenciación;
 10 obtener un segundo conjunto de datos de secuenciación filtrando datos no cualificados que contiene más del 10
 % de bases no determinadas a partir del primer conjunto de datos de secuenciación;
 obtener un tercer conjunto de datos de secuenciación filtrando datos no cualificados que contiene una secuencia
 de adaptador de secuenciación a partir del segundo conjunto de datos de secuenciación después de haber sido
 alineado a una biblioteca de secuencias de adaptador de secuenciación; y
 15 obtener los primeros datos de secuenciación cualificados filtrando datos no cualificados de los cuales contiene
 una secuencia exógena a partir del tercer conjunto de datos de secuenciación después de haber sido alienados a
 todas las secuencias exógenas introducidas.

Preferentemente, la etapa de filtrado de los segundos datos de secuenciación no cualificados que comprende
 adicionalmente:

20 obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando
 todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN en los primeros datos de
 secuenciación cualificados en un orden ascendente; y
 filtrar datos de secuenciación no cualificados de los cuales tiene un valor de profundidad de secuenciación más
 allá de un intervalo de $Q1-1.5IQR$ a $Q3+1.5IQR$, para obtener los segundos datos de secuenciación cualificados.

25 Preferentemente, la etapa de filtrado de los terceros datos de secuenciación no cualificados que comprende
 adicionalmente:

obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando
 todos los valores de profundidad de secuenciación de la región dada entre distintas muestras de ADN en los
 segundos datos de secuenciación cualificados en un orden ascendente; y
 30 filtrar datos de secuenciación no cualificados de los cuales una mediana de los valores de profundidad de
 secuenciación de la región dada entre las distintas muestras de ADN es cero o superior a $Q3+1.5IQR$, para
 obtener los terceros datos de secuenciación cualificados.

Como referencia, se describe otra solución técnica en el presente documento: proporcionar un aparato para la
 detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma, que puede
 comprender:

35 una sonda de captura que obtiene módulo, para obtener una sonda de captura que se corresponde con
 secuencia de ADN de la región de sitio de secuencia marcada;
 un módulo de hibridación, para hibridar la sonda de captura con una biblioteca de mezcla de una pluralidad de
 muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la
 pluralidad de muestras de ADN;
 40 un módulo de obtención de datos de secuenciación, para secuenciar la secuencia de ADN capturada de la región
 de sitio de secuencia marcada para obtener datos de secuenciación; y
 un módulo de obtención de resultados de microdelección, para analizar los datos de secuenciación usando un
 procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de
 la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de
 45 muestras de ADN.

Preferentemente, el módulo de obtención de resultados de microdelección comprende adicionalmente:

una unidad de normalización de valor de profundidad de secuenciación, para someter un valor de profundidad de
 secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de profundidad
 de secuenciación normalizado; y
 50 una unidad de obtención de resultados de microdelección, para detectar un valor atípico del valor de profundidad
 de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de
 secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que
 muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del
 cromosoma en cada una de la pluralidad de muestras de ADN.

55 Preferentemente, la unidad de normalización de valor de profundidad de secuenciación se usa en la división de
 todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una
 misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener

los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.

Preferentemente, la unidad de obtención de resultados de microdelección comprende adicionalmente:

- 5 un módulo de obtención de valor medio y varianza, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
- un módulo de obtención de curva de distribución normal, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
- 10 un módulo de cálculo de probabilidad, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal; y
- un primer módulo de determinación, para determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de
- 15 profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

Preferentemente, la unidad de obtención de microdelección aún comprende adicionalmente:

- 20 un módulo de determinación de valor de corte de probabilidad, para someter el resultado R1 a verificación experimental, para obtener un resultado de verificación y determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
- un segundo módulo de determinación, para obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al segundo valor de corte de probabilidad.

25 Preferentemente, la unidad de obtención de microdelección aún comprende adicionalmente:

- un primer módulo de obtención de relación D/S, para calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de
- 30 ADN;
- un primer módulo de obtención de relación D/R, para calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
- 35 un primer módulo de obtención de valor de corte de relación, para formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3;
- un tercer módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es superior al primer valor de corte de
- 40 relación;
- un cuarto módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y
- 45 un quinto módulo de determinación, para obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

Preferentemente, la unidad de obtención de microdelección aún comprende adicionalmente:

- 50 un módulo de obtención de valor medio y varianza, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
- un módulo de obtención de curva de distribución normal, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
- 55 un módulo de cálculo de probabilidad, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
- un sexto módulo de determinación, para determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R3 que muestra la presencia de la microdelección en la región

dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;

un segundo módulo de obtención de relación D/S, para calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;

un segundo módulo de obtención de relación D/R, para calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;

un segundo módulo de determinación de valor de corte de relación, para formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;

un séptimo módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;

un octavo módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación; y

un noveno módulo de determinación, para obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

Preferentemente, el módulo de obtención de sonda de captura comprende adicionalmente:

una unidad de detección de región, para encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica;

una unidad de selección de secuencia, para seleccionar una secuencia de ADN cualificada de la región de sitio de secuencia marcada que cumple el requisito para designar la sonda de captura; y

una unidad de obtención de sonda de captura, para designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.

Preferentemente, el aparato descrito en el presente documento puede comprender adicionalmente un módulo de construcción de biblioteca de mezcla, en la que el módulo de construcción de biblioteca de mezcla comprende adicionalmente:

una unidad de construcción de pluralidad de bibliotecas de ADN, para construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;

una unidad de mezcla, para mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada; y

una unidad de construcción de biblioteca de mezcla, para determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.

Preferentemente, el aparato descrito en el presente documento puede comprender adicionalmente un módulo de control de calidad, en la que el módulo de control de calidad comprende adicionalmente:

una primera unidad de filtrado, para filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados;

una unidad de cálculo de valor de profundidad de secuenciación, para alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;

una segunda unidad de filtrado, para filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados; y

una tercera unidad de filtrado, para filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

Para solucionar el anterior problema técnico, una solución técnica adicional usada por la presente divulgación es: proporcionar al ordenador un medio legible que comprenden un orden, configurado para realizar el procedimiento anterior por un procesador.

Las ventajas de la presente divulgación pueden comprender: siendo distintas de la técnica anterior, el procedimiento de detección de la microdelección en la región de sitio de secuencia marcada del cromosoma y el aparato del mismo, mediante el cual obtiene la sonda de captura que cubre una región de sitio de secuencia marcada completa de un cromosoma y captura secuencia de ADN de la región de sitio de secuencia marcada a partir de una pluralidad de muestras de ADN después de hibridarse con una bibliotecas de mezcla de la pluralidad de muestras de ADN, puede detectar de forma eficaz y precisa microdelecciones informadas o no informadas en una región de sitio de secuencia

marcada de un cromosoma con una gran cantidad de muestras de ADN; además el procedimiento del análisis de los datos de secuenciación usando un procedimiento estadístico matemático es científico, estable, con una alta sensibilidad, una tasa de falso positivos baja, que puede usarse en el análisis eficaz de microdeleciones.

Breve descripción de los dibujos

- 5 La Fig. 1 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.2 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- 10 La Fig.3 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización adicional de la presente divulgación;
- La Fig.4 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- 15 La Fig.5 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- La Fig.6 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.7 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- 20 La Fig.8 es un histograma que muestra una muestra filtrada que tiene un valor atípico de valor de profundidad de secuenciación mediante un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.9 es un histograma que muestra una muestra que tiene un valor de profundidad de secuenciación normal mediante un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- 25 La Fig.10 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- La Fig.11 es un histograma que muestra una región filtrada que tiene un valor atípico de valor de profundidad de secuenciación mediante un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.12 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización adicional de la presente divulgación;
- 35 La Fig.13 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.14 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- 40 La Fig.15 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- 45 La Fig.16 es un diagrama que muestra un ejemplo que ilustra un árbol de toma de decisiones construido mediante algoritmo ID3 en un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.17 es un diagrama de flujo parcial que muestra un análisis del árbol de toma de decisiones en un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- 50 La Fig.18 es un diagrama de flujo que muestra el análisis del árbol de toma de decisiones en un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;
- La Fig.19 es un diagrama de flujo que muestra un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización adicional de la presente divulgación;
- 55 La Fig.20 comprende un histograma y un gráfico que muestra una mediana del valor de profundidad de secuenciación de una muestra en un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- 60 La Fig.21 comprende un histograma y un gráfico que muestra una mediana del valor de profundidad de secuenciación de una región mediante un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;
- La Fig.22 es un diagrama que muestra una probabilidad cuando un valor de profundidad de secuenciación es un valor determinado en un procedimiento de detección de una microdeleción en una región de sitio de secuencia

marcada de un cromosoma de acuerdo con una realización de la presente divulgación;

La Fig.23 es un diagrama que muestra un umbral de probabilidad y una relación entre una expectativa del umbral de probabilidad y una observación del umbral de probabilidad en un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;

La Fig.24 es un diagrama que muestra una relación de una tasa de positivo verdadero con respecto a tasa de positivo falso en un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;

La Fig. 25 es un diagrama que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación;

La Fig.26 es un diagrama esquemático que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con otra realización de la presente divulgación;

La Fig.27 es un diagrama esquemático que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización adicional de la presente divulgación;

La Fig.28 es un diagrama que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con aún otra realización de la presente divulgación;

La Fig.29 es un diagrama esquemático que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con aún otra realización de la presente divulgación.

Descripción detallada

Se hará referencia, en detalle, a los dibujos y las realizaciones de la presente divulgación. Además, los términos "primero" y "segundo" se usan en el presente documento para fines descriptivos y no están previstos para indicar o implicar una importante o significancia relativa.

La Fig.1 es un diagrama de flujo que muestra un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación. Tal como se muestra en la Fig.1, el procedimiento comprende:

etapa 101: obtener una sonda de captura que se corresponde con secuencia de ADN de la región de sitio de secuencia marcada.

El sitio de secuencia marcada (STS) es una secuencia de ADN de copia única corta de conocida ubicación de mapa que sirve como marcador, que puede mapearse únicamente mediante amplificación de PCR, para producir un sitio para la realización de mapa, en concreto, detección con un orden de una serie de STS puede fabricar un mapa de una región genómica. Una sonda es un fragmento de ADN o ARN unicatenario corto (aproximadamente de 20 a 500 pb) para detectar una secuencia de ácido nucleico complementario del mismo.

etapa 102: hibridar la sonda de captura con una biblioteca de mezcla de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN.

La expresión "bibliotecas de mezcla" usada en el presente documento se refiere a todas las secuencias de ADN para secuenciar obtenidas mezclando una pluralidad de bibliotecas de ADN de muestras de ADN únicas. A continuación se describe, en detalle, etapas específicas para la construcción de la biblioteca de mezcla.

etapa 103: secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación.

etapa 104: analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

La estadística matemática es una rama de las matemáticas que se desarrolla con un desarrollo de teoría de probabilidad, estudiando cómo optimizar eficazmente, disponer y analizar datos afectados por factores aleatorios, mediante lo cual se puede inferir o predecir un problema en consideración, proporcionando bases y sugerencias para fabricar una determinada decisión o tomar una determinada acción.

En una realización, tal como se muestra en la Fig.2, la etapa de obtener la sonda de captura comprende adicionalmente:

etapa 201: encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica.

Dentro de una región de sitio de secuencia marcada de cromosoma Y en base de datos de UCSC, en una coordenada de ubicación de secuencia de referencia de genoma humano Hg19 (<http://genome.ucsc.edu/>), la secuencia de ADN de la región de sitio de secuencia marcada en el cromosoma Y puede encontrarse basándose en la coordenada de ubicación.

etapa 202: seleccionar una secuencia de ADN cualificada de la región de sitio de secuencia marcada que cumple el requisito para designar la sonda de captura.

La eficacia de captura de un chip puede verse afectada por la repetibilidad y contenido de GC de secuencia de ADN, que puede dar como resultado un error de captura. Por lo tanto, es muy importante seleccionar una

secuencia de ADN cualificada que cumpla el requisito de determinar la sonda de captura.
 etapa 203: designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.

5 Para capturar un fragmento de ADN, la sonda de captura tiene en general una longitud de 60 a 150 pb, contenido GC entre un intervalo del 40 % al 70 %. Cuando se determina una sonda para una secuencia de ADN de una misma región, se requiere una pluralidad de sondas para cubrir la secuencia de ADN completa, además se presenta un solapamiento entre dos de las sondas de capturas, en el que el solapamiento general tiene una longitud de 20 pb.

10 En una realización específica, una coordenada de ubicación de secuencia de ADN en la región de sitio de secuencia marcada de un cromosoma se encuentra en la base de datos genómica; y, a continuación, la ubicación coordinada se envía a una empresa que proporciona un servicio de captura de ADN; y, a continuación, se finaliza la determinación y síntesis de la sonda de captura por tal empresa.

Preferentemente, tal como se muestra en la Fig.3, la biblioteca de mezcla de la pluralidad de muestras de ADN se construye mediante las siguientes etapas:

15 etapa 301: construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;
 cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente, para distinguir bibliotecas derivadas a partir de distintas muestras de ADN en la etapa de secuenciación. Cada una de la pluralidad de bibliotecas de ADN tiene una secuencia base índice distinta que tiene una longitud de 6 pb a 8 pb en un extremo de la misma.

20 etapa 302: mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada.
 La pluralidad de bibliotecas de ADN puede mezclarse igualmente con una relación predeterminada.

etapa 303: determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.

Las bibliotecas de mezcla de la pluralidad de muestras de ADN se someten a cuantificación y a un control de calidad para retirar impureza exógena.

25 Preferentemente, tal como se muestra en la Fig.4, la etapa de construcción de una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, para cada una de la pluralidad de muestras de ADN únicas, la biblioteca de ADN se construye mediante la siguientes etapas:

30 etapa 401: romper un genoma de ADN en fragmentos de ADN que tienen un tamaño predeterminado mediante procedimiento físico o químico, para obtener fragmentos de ADN.

En general, los fragmentos de ADN obtenidos tienen una longitud de 200 a 300 pb y la sonda capturada tiene una longitud de aproximadamente 80 pb, como la longitud de los fragmentos de ADN siendo como de 200 a 300 pb puede dar como resultado una eficacia de captura superior relativa; además, el valor de profundidad de secuenciación es en general de 200 a 300 pb mediante secuenciación de PE después de la captura de sonda.

35 etapa 402: someter los fragmentos de ADN a reparación de extremo, para obtener fragmentos de ADN con extremo reparado con un extremo fosforilado;

etapa 403: añadir una base "A" a los fragmentos de ADN con extremo reparado en el extremo 3' para obtener fragmentos de ADN con la base "A" en el extremo 3';

40 etapa 404: ligar un adaptador de índice a los fragmentos de ADN con la base "A" en el extremo 3', para obtener fragmentos de ADN ligados con el adaptador de índice;

etapa 405: amplificar los fragmentos de ADN ligados con el adaptador de índice usando un cebador que comprende una secuencia específica al adaptador de índice, para obtener un producto de amplificación; y

etapa 406: determinar el producto de amplificación como la biblioteca de ADN, si el producto de amplificación cumple el requisito de control de calidad.

45 Preferentemente, después de la etapa de secuenciación de la secuenciación del ADN capturado y antes de la etapa de análisis de los datos de secuenciación, el procedimiento comprende adicionalmente: someter los datos de secuenciación a un control de calidad. Tal como se muestra en la Fig.5, la etapa de someter de los datos de secuenciación a control de calidad comprende adicionalmente:

50 etapa 501: filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados; preferentemente la técnica de secuenciación de alto rendimiento puede ser la técnica de secuenciación Illumina Hiseq 2000 o puede ser otras técnicas de secuenciación de alto rendimiento que existen en la técnica anterior.

55 etapa 502: alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;

etapa 503: filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados; y

etapa 504: filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

5 Preferentemente, tal como se muestra en la Fig.6, en la etapa 501, la etapa de filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener los primeros datos de secuenciación cualificados que comprenden adicionalmente:

10 etapa 601: obtener un primer conjunto de datos de secuenciación filtrando datos no cualificados que contiene más de una proporción predeterminada de bases que tienen una baja calidad de los datos de secuenciación. El equipamiento de secuenciación distinto usa distintos procedimientos de cálculo del valor de calidad, en el que un estándar para secuencia con un bajo valor de calidad puede consultar empresas de equipos de secuenciación o referirse al estándar general en la técnica. En el ejemplo actual, una fórmula para calcular un valor de calidad usado por el secuenciador HiSeq2000 de la empresa Illumina es $Q = A - 64$, en el que Q es el valor de calidad de secuenciación de una base determinada, A se encuentra en código ASCII de caracteres que se corresponden al valor de calidad de secuenciación de la base determinada en una salida de archivo FQ por el secuenciador HiSeq2000. En el ejemplo actual, las bases que tienen un valor de calidad inferior a 5 se define como una base con un valor de calidad bajo. Si una secuencia contiene más del 50 % de bases que tienen el valor de calidad bajo, tal secuencia se determina como los primeros datos de secuenciación no cualificados, los cuales deben filtrarse para obtener los primeros datos de secuenciación cualificados.

15 etapa 602: obtener un segundo conjunto de datos de secuenciación filtrando datos no cualificados que contiene más del 10 % de bases no determinadas a partir del primer conjunto de datos de secuenciación;

20 etapa 603: obtener un tercer conjunto de datos de secuenciación filtrando datos no cualificados que contiene una secuencia de adaptador de secuenciación a partir del segundo conjunto de datos de secuenciación después de haber sido alineado a una biblioteca de secuencias de adaptador de secuenciación; y

25 etapa 604: obtener los primeros datos de secuenciación cualificados filtrando datos no cualificados de los cuales contiene una secuencia exógena a partir del tercer conjunto de datos de secuenciación después de haber sido alienados a todas las secuencias exógenas introducidas.

30 En la aplicación práctica, la técnica de secuenciación de alto rendimiento puede ser la técnica de secuenciación Illumina Hiseq 2000 o puede ser otras técnicas de secuenciación de alto rendimiento que existen en la técnica. Distintos secuenciadores o condiciones pueden usar distintos estándares para filtrar datos de secuenciación no cualificados, por ejemplo, puede usarse un determinado estándar cuando se secuencia por Illumina Hiseq 2000: una secuencia que contiene más del 50 % de bases que tiene un valor de calidad inferior a determinado umbral se considera como una secuencia no cualificada, en la cual el determinado umbral para valor de calidad bajo depende de la técnica de secuenciación específica; una secuencia que contiene más del 10 % de bases no determinadas (como base N en un resultado de secuenciación por Illumina Hiseq 2000) se considera como una secuencia no cualificada; excepto para una secuencia de adaptador de muestra, los datos de secuenciación cualificados se alinean a otras secuencias exógenas introducidas por el experimento, tales como diversas secuencias de adaptador. Si la secuencia exógena se presenta, entonces una secuencia completa que contiene la secuencia exógena se considera como ser no cualificada.

40 Preferentemente, tal como se muestra en la Fig.7, la etapa de filtrado de los segundos datos de secuenciación no cualificados que comprende adicionalmente:

45 etapa 701: obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN en los primeros datos de secuenciación cualificados en un orden ascendente.

50 En estadísticas descriptivas, los cuartiles de un conjunto clasificado de valores de datos en orden ascendente son los tres puntos que dividen el conjunto de datos en cuatro grupos iguales, comprendiendo cada grupo un cuarto de los datos. Un primer cuartil (designado Q1) también conocido como "un cuartil más pequeño", a saber, un cuartil inferior, iguala al 25 % de valor th en el conjunto clasificado de todos los valores de datos. Un segundo cuartil (designado Q2) también conocido como una mediana, iguala al 50% de valor th en el conjunto clasificado de todos los valores de datos. Un tercer cuartil (designado Q3) también conocido como "un cuartil más grande" un cuartil superior, iguala al 75% de valor th en el conjunto clasificado de todos los valores de datos. Un intervalo de intercuartil (designado IQR) es la diferencia entre el cuartil superior y el cuartil inferior. Cada uno del Q1, Q2 y Q3 se considera como un punto de demarcación independientemente de los valores de tales varianzas del mismo, mediante el cual todos los valores de datos se dividen en cuatro partes iguales. Una tendencia de las varianzas de los valores de datos puede analizarse comparando Q1 y Q3. El cuartil también se ha aplicado ampliamente en diseñar un diagrama de cajas en estadística. El diagrama de cajas se refiere a un diagrama compuesto de una caja y dos segmentos trazados con un conjunto de datos que son cinco magnitudes estadísticas (rasgos), no solo directamente reflejando la distribución del conjunto de datos, pero también siendo capaz de analizar una pluralidad de datos. Tales cinco magnitudes estadísticas son el valor máximo, el valor mínimo, la mediana y dos de los cuartiles.

60 etapa 702: filtrar datos de secuenciación no cualificados de los cuales tiene un valor de profundidad de secuenciación más allá de un intervalo de $Q1 - 1.5IQR$ a $Q3 + 1.5IQR$, para obtener los segundos datos de secuenciación cualificados.

En una realización de la presente divulgación, tal y como se muestra en las Fig.8 y Fig.9, la coordenada X indica una distribución de profundidad de región; la coordenada Y indica una secuencia de región para el mismo valor de profundidad. La Fig.8 es un histograma que muestra una muestra filtrada que tiene un valor atípico de valor de profundidad de secuenciación. La Fig.9 es un histograma que muestra una muestra que tienen un valor de profundidad de secuenciación.

5 Preferentemente, tal como se muestra en la Fig.10, la etapa de filtrado de los terceros datos de secuenciación no cualificados que comprende adicionalmente:

10 etapa 1001: obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la región dada entre distintas muestras de ADN en los segundos datos de secuenciación cualificados en un orden ascendente; y
 etapa 1002: filtrar datos de secuenciación no cualificados de los cuales una mediana de los valores de profundidad de secuenciación de la región dada entre las distintas muestras de ADN es cero o superior a $Q3+1.5IQR$, para obtener los terceros datos de secuenciación cualificados.

15 En otra realización de la presente divulgación, tal como se muestra en la Fig.21, la coordenada X indica la distribución de profundidad de la pluralidad de muestras para una misma región, la coordenada Y indica la frecuencia de región con un mismo valor de profundidad de secuenciación. La Fig.21 es un histograma que muestra una región filtrada que tiene un valor atípico de valor de profundidad de secuenciación.

Preferentemente, tal como se muestra en la Fig.12, la etapa de análisis de los datos de secuenciación usando el ensayo estadístico matemático comprende adicionalmente:

20 etapa 1201: someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de profundidad de secuenciación normalizado. Preferentemente, la etapa de someter el valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización comprende adicionalmente: dividir todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.
 25 etapa 1202: detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.
 30

En una realización preferida, tal como se muestra en la Fig.13, la etapa 1202 aún comprende adicionalmente:

35 etapa 1301: calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 etapa 1302: obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región; etapa 1303: calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 40 etapa 1304: determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

45 En otra realización preferente, tal como se muestra en la Fig.14, después de la etapa 1304 en el procedimiento comprende adicionalmente:

50 etapa 1401: someter el resultado R1 a verificación experimental, para obtener un resultado de verificación y determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
 etapa 1402: obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al segundo valor de corte de probabilidad.

55 En la aplicación práctica, en caso de que la mayoría de muestras no tengan microdelección, como para una determinada región, un valor de profundidad de secuenciación de una muestra que tiene una microdelección debe tener un valor atípico de una distribución normal obtenida con las muestras que no tienen microdelección. Se usa un procedimiento estadístico matemático para detectar un valor atípico de los valores de profundidad de secuenciación para cada región, mediante los cuales una región que tiene un valor atípico de los valores de profundidad de secuenciación se determina como una región que tiene una microdelección.

Un procedimiento de determinación si un valor de profundidad de secuenciación para una región dada es un valor atípico comprende: obtener una curva de distribución normal con todos los valores no atípicos para la región dada, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la región dada, después de filtrar los valores atípicos de los valores de profundidad de secuenciación; calcular una probabilidad (valor p.) para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal; determinar un primer valor de corte de probabilidad adecuado (valor de corte de p.) para estas probabilidades; y obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

Un procedimiento de determinación del valor de corte de probabilidad puede comprender: calcular la probabilidad (valor p.) para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación en la distribución normal de acuerdo con el procedimiento anteriormente mencionado; determinar la presencia o ausencia de microdelección dentro de más regiones de muestra con un valor de corte relajado; y someter estas regiones de muestra a verificación experimental, para determinar la presencia o ausencia de microdelección. Un procedimiento de verificación experimental puede comprender: designar una sonda adecuada; someter el ADN de tales regiones de muestra a reacción de PCR; determinar si el producto de amplificación de PCR es normal, para determinar la presencia o ausencia de microdelección. Después de obtener información que muestre si se presenta microdelección en tales regiones de muestra, puede seleccionarse un valor de corte de probabilidad adecuado, para tener datos con el mejor valor de falso negativo y valor de falso positivo.

En una realización preferida adicional, tal como se muestra en la Fig.15, la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

etapa 1501: calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;

etapa 1502: calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;

etapa 1503: formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3;

etapa 1504: obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es superior al primer valor de corte de relación;

etapa 1505: obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y

etapa 1506: obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

En aplicaciones prácticas, cuando se necesita detectar una pluralidad de muestras y regiones y una proporción de las regiones de muestra que tienen microdelección es relativamente más pequeña, la relación entre el valor de profundidad de secuenciación de la región que tiene microdelección y la mediana de todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN es normalmente más pequeña y la relación entre el valor de profundidad de secuenciación de la región que tiene microdelección y la mediana de los valores de profundidad de secuenciación de todas las regiones también es normalmente más pequeño. Por lo tanto, puede usarse un procedimiento basado en un árbol de decisiones usando un parámetro adecuado en la verificación de un valor atípico de valor de profundidad de secuenciación.

Etapas específicas del procedimiento basándose en el árbol de decisiones pueden comprender: calcular una primera relación entre un valor de profundidad de secuenciación de una región de muestra a someter a ensayo y una mediana de todos los valores de profundidad de secuenciación de una de la pluralidad de muestras de ADN; determinar un resultado que muestra una ausencia de microdelección si la primera relación es superior a un primer determinado valor de corte; calcular una segunda relación entre el valor de profundidad de secuenciación de la región de muestra a someter a ensayo y una mediana de la mediana de los valores de profundidad de secuenciación de todas las regiones, si la primera relación es inferior al primer determinado valor de corte; determinar el resultado que muestra una ausencia de microdelección si la segunda relación es superior a un segundo determinado valor de corte y determinar un resultado que muestra una presencia de microdelección si la segunda relación es inferior al segundo determinado valor de corte.

La determinación de un valor de corte en el ensayo de árbol de decisiones se calcula mediante algoritmo Dicotomizador Iterativo 3 (ID3) del árbol de decisiones (Mitchell, Tom M. Machine Learning. McGraw-Hill, 1997). El grano del algoritmo de ID3 comprende: durante la selección de un atributo entre internodos del árbol de decisión, la

obtención de información se toma como un estándar para la selección de atributos, de modo que cuando se somete cada uno de los nodos sin hoja a ensayos, puede obtenerse la máxima información de clase con respecto al informe sometido a ensayo.

Las ideas básicas para el algoritmo ID3 del árbol de decisión comprenden:

- 5 etapa 1, seleccionar un atributo como un nodo raíz del árbol de decisiones y crear ramificaciones con todos los valores del atributo;
- etapa 2: clasificar datos formados usando el árbol creado, en el que
- si todos los ejemplos de un nodo de hoja pertenecen a la misma clase, tal nodo de hoja se marca con la misma clase; entonces si todos los nodos de hoja se marcan con clases, se termina el algoritmo;
- 10 etapa 3: si aún existe un nodo de hoja sin marcar, se usa un atributo que no se presenta nunca en una trayectoria desde el nodo de hoja hasta el nodo de raíz para marcar tal nodo de hoja; y crear continuamente ramificaciones con todos los valores de tal atributo; repetir la etapa 2 del algoritmo.

En la etapa 1, seleccionar distintos atributos puede crear un árbol de decisiones distinto, de este modo, seleccionar un atributo adecuado puede crear un árbol de decisiones simple. En el algoritmo ID3, se selecciona en general un atributo mediante un procedimiento a base de información heurística. El atributo seleccionado mediante el procedimiento heurístico puede comprender la cantidad más alta de información, dando como resultado un árbol de decisiones que tiene ramificaciones mínimas. En la presente divulgación, se deben seleccionar dos distribuidos, en concreto, una relación T1 (D/S) entre el valor de profundidad de secuenciación de la región de muestra y una mediana del valor de profundidad de secuenciación de la región de muestra, una relación T2(D/R) entre el valor de profundidad de profundidad de secuenciación de la pluralidad de muestras de ADN de la misma región y una mediana del valor de profundidad de secuenciación de la pluralidad de muestras de ADN para una misma región.

Los atributos con máxima información se determinan mediante obtención de información. Un procedimiento de cálculo de obtención de información puede comprender:

D se asume una división de conjunto de formación por clases y una entropía de D se representa por

25
$$info(D) = - \sum_{i=1}^{T1} p_i \log_2(p_i)$$
, en el cual pi representa una probabilidad de la clase i-th que se presenta en el conjunto de formación, que puede estimarse mediante la división del número del conjunto de formación que pertenece a tal clase por el número total de todo el conjunto de formación. Un significado práctico de la entropía se refiere a una cantidad de información promedio requerida para la marcación de clase del conjunto de formación en D.

30 Asumiendo que el conjunto de formación en D se divide por un atributo A, entonces, la información que se espera de

$$info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} info(D_j)$$

la división de D por A es $gain(A) = info(D) - info_A(D)$, mientras que la obtención de información es

El algoritmo ID3 calcula una tasa de obtención de cada atributo en cada división requerida y se divide basándose en un atributo que tiene la tasa de obtención máxima.

35 Un ejemplo de detectar un recuento no real en una comunidad se muestra a continuación para ilustrar como crear un árbol de decisiones usando el algoritmo ID3. Para una mayor simplicidad, el conjunto de formación se asumió que contenía 10 elementos, que se muestran en la Tabla 1.

Tabla 1 detección de recuento no real en una comunidad

Densidad log	Densidad de compañeros	usando o sin usar un retrato de cabeza real	siendo o no siendo un recuento real
s	s	no	no
s	l	sí	sí
l	m	sí	sí
m	m	sí	sí
l	m	sí	sí
m	l	no	sí
m	s	no	no

(continuación)

Densidad log	Densidad de compañeros	usando o sin usar un retrato de cabeza real	siendo o no siendo un recuento real
l	m	no	Sí
m	s	no	sí
s	s	sí	no

s, m y l representan respectivamente pequeño, medio y grande. Asumiendo que L, F, H y R representan respectivamente la densidad log, densidad de compañeros, usando o sin usar un retrato de cabeza real y siendo o no siendo un recuento real. La obtención de información de cada atributo se calculó a continuación.

$$info(D) = -0,7\log_2 0,7 - 0,3\log_2 0,3 = 0,7 * 0,51 + 0,3 * 1,74 = 0,879$$

$$obtención(L) = 0,879 - 0,603 = 0,276$$

5 Por lo tanto, la obtención de información de densidad log fue de 0,276. La obtención de información de H y F respectivamente fue de 0,033 y 0,553 usando el mismo procedimiento. Como F tuvo la máxima obtención de información, F se seleccionó como el atributo para la primera división, un resultado después de la primera división se muestra como Fig.16.

Mientras que para resolver los datos que tienen un atributo de consecutividad similar a T1 y T2, los datos se sometieron en primer lugar a discretización. Un procedimiento simple del mismo fue dividir los valores del atributo A en dos partes: una parte Ai contiene valores del atributo A siendo inferiores a N y la otra parte Ai contiene valores del atributo A siendo superiores a N.

10 Para cualquiera uno de los atributos, todos los valores en un conjunto de datos es limitado. Asumiendo que los valores del atributo fueron (v1, v2, ...,vn).

Entonces, en tal conjunto de datos, hubo totalmente n-1 partes de valores de división, basándose en lo que se construyó en el árbol de decisiones.

Las etapas específicas de discretización pueden contener:

- 15 1) encontrar el valor mínimo (MIN) del atributo consecutivo, encontrar el valor máximo MAX del atributo consecutivo;
- 2) determinar puntos igualmente divididos N Ai en un intervalo (MIN, MAX), que fueron respectivamente $A_i = MIN + ((MAX - MIN) / N) * i$, en los que $i = 1, 2, 3, \dots$;
- 20 3) calcular un valor de obtención tomando (MIN, Ai) y (Ai, MAX) como el valor de intervalo respectivamente y comparando los valores de obtención calculados;
- 4) seleccionar Ak que tiene un valor de obtención máximo como un punto de rotura del atributo consecutivo y determinar dos intervalos del atributo como (MIN, Ak) y (Ak, MAX).

En un ejemplo de la presente divulgación, las etapas específicas de determinación del valor de corte basándose en el árbol de decisiones comprendían:

25 En primer lugar, someter la pluralidad de muestras seleccionada a verificación de PCR de la región parcial, para obtener el resultado que muestra la presencia de la microdelección, en la que "+" representaba que el resultado que mostraba la presencia de la microdelección fue positivo; "-" representaba que el resultado que mostraba la ausencia de la microdelección fue negativo

30 calcular la entropía basándose en una fórmula de $Info(D) = -(P+) \log_2 (P+) - (P-) \log_2 (P-)$, en la que p+ (P-) representaba una probabilidad de presencia (ausencia) de la microdelección en el grupo de formación completo, que podría estimarse dividiendo la cantidad de elemento que pertenece a tal clase por la cantidad total de elementos en los conjuntos de formación completos. El significado real de entropía se refería a una cantidad de información promedio requerida para la marcación de clase de los conjuntos de formación en D.

35 En segundo lugar, calcular una relación T1 entre un valor de profundidad de secuenciación normalizado de región de muestra y la mediana de todos los valores de profundidad de secuenciación de la de la pluralidad de muestras de ADN;

40 clasificar la relación T1 obtenida en un orden ascendente, para obtener V1, V2, V3, ..., Vn; tomando un valor medio de un par de valores clasificados consecutivos como el valor de corte $a1' = (V1 + V2) / 2$, $a2' = (V2 + V3) / 2, \dots, (n-1)' = (Vn-1 + Vn) / 2$; obtener el valor de corte mínimo a1' y el valor de corte máximo a(n-1)'

En tercer lugar, calcular una relación T2 entre un valor de profundidad de secuenciación normalizado de región de muestra y la mediana de los valores de profundidad de secuenciación de todas las regiones; clasificar la relación T2 obtenida en un orden ascendente, para obtener U1, U2, U3, ..., Un;

tomando un valor medio de un par de valores clasificados consecutivos como el valor de corte $b1'=(U1+U2)/2$, $b2'=(U2+U3)/2$..., $b(n-1)'=(Un-1+Un)/2$;

obtener el valor de corte mínimo $b1'$ y el valor de corte máximo $a(n-1)'$.

5 En cuarto lugar, calculas las entropías de distintos valores divididos, $a1'$, $a2'$, ..., obtenidos mediante ciclos basándose en:

$$info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} info(D_j)$$

mientras que la obtención de información de la misma fue:

$$obtención(A) = info(D) - info_A(D).$$

10 En quinto lugar, calculas las entropías de distintos valores divididos, $b1'$, $b2'$, ..., obtenidos mediante ciclos basándose en:

$$info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} info(D_j)$$

mientras que la obtención de información de la misma fue:

$$obtención(A) = info(D) - info_A(D).$$

15 En sexto lugar, comparar los grupos de las obtenciones máximas de obtención de información, en los que el atributo (a) que tiene un valor más grande se tomó como una raíz; clasificar información basándose en un valor de corte que tiene el valor de corte máximo de valor de obtención, que se mostró en la Fig.17.

20 En séptimo lugar, calcular la obtención máxima dividiendo los valores de profundidad de secuenciación de dos grupos de datos por una mediana de valores de profundidad de secuenciación en la región respectivamente de acuerdo con las etapas anteriormente mencionadas; clasificar información basándose en el valor de corte que tiene el valor de corte máximo de valor de obtención, que se mostró en la Fig.18, es decir, el árbol de decisiones determinado de la presente divulgación.

En una realización preferida adicional, tal como se muestra en la Fig.19, la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

- 25 etapa 1901: calcular un valor medio y una varianza basándose en todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
- etapa 1902: obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
- 30 etapa 1903: calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
- etapa 1904: determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R3 que muestra la presencia de la microdelección en la región dada, si la probabilidad de
- 35 una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;
- etapa 1905: calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;
- 40 etapa 1906: calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;
- etapa 1907: formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
- 45 etapa 1908: obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
- etapa 1909: obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación; y

etapa 1910: obtener el resultado que muestra la presencia de la microdeleción en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

5 En aplicaciones prácticas, para detectar de forma más precisa una región que tiene microdeleción, puede usarse una estrategia que combina los dos procedimientos anteriores, que comprende: predeterminar un valor de corte de valor p. relajado y filtrar los datos resultantes basándose en el árbol de decisiones, para obtener el resultado que muestra la presencia de microdeleción en la región de sitio de secuencia marcada del cromosoma.

10 El procedimiento de detección de la microdeleción en la región de sitio de secuencia marcada del cromosoma y el aparato del mismo, mediante el cual obtiene la sonda de captura que cubre una región de sitio de secuencia marcada completa de un cromosoma y captura secuencia de ADN de la región de sitio de secuencia marcada a partir de una pluralidad de muestras de ADN después de hibridarse con una bibliotecas de mezcla de la pluralidad de muestras de ADN, puede detectar de forma eficaz y precisa microdeleciones informadas o no informadas en una región de sitio de secuencia marcada de un cromosoma con una gran cantidad de muestras de ADN; además de la etapa de analizar mediante procedimiento estadístico matemático en la presente divulgación puede basarse en una distribución normal o puede basarse en un procedimiento de análisis de árbol de decisiones o puede basarse en una combinación de los mismos, seguido por una verificación experimental, tal procedimiento de análisis de información es científico, estable, con una alta sensibilidad, una tasa de falso positivos baja, que puede usarse en el análisis eficaz de microdeleciones.

20 Se hará referencia, en detalle, a los ejemplos de la presente divulgación. Se apreciará por los expertos en la técnica que los siguientes ejemplos son explicativos y no pueden interpretarse como limitantes del ámbito de la presente divulgación. Si la tecnología o condiciones específicas no se especifican en los ejemplos, se llevará a cabo una etapa de acuerdo con las técnicas o condiciones descritas en la literatura en la técnica (por ejemplo, haciendo referencia a J. Sambrook, y col. (traducido por Huang PT), Molecular Cloning: A Laboratory Manual, 3ª Ed., Science Press) o de acuerdo con las instrucciones de producto. Si los fabricantes o reactivos o instrumentos no se especifican, los reactivos o instrumentos pueden estar disponibles en el mercado, por ejemplo, en Illumina.

25 El ejemplo actual toma un ejemplo de detección de una microdeleción en una región de sitio de secuencia marcada de cromosoma Y, aunque no limitado con el cromosoma Y. Un total de 11 muestras que incluían 10 muestras de infertilidad y 1 muestra de ser humano sano se usaron en el ejemplo actual, que se sometieron a hibridación con un chip de Nimblegen (Roche) después de construir una biblioteca juntas. El número de muestras en el ejemplo actual se usó para ilustrar la presente divulgación, aunque no se limitó al número de muestras para un experimento.

30 Los reactivos usados en el ejemplo actual se muestran en la Tabla 2.

Tabla 2. Reactivos usados en el ejemplo actual de la presente divulgación

Reactivo	Modelo	Fabricante
Kit de purificación de PCR de QIAquick	28106	QIAGEN
Tampón de quinasa de polinucleótidos 10x	B904	Enzymatics
Conjunto de solución dNTP	N201L	Enzymatics
ADN polimerasa T4	P708L	Enzymatics
Enzima Klenow	P706L	Enzymatics
Quinasa de polinucleótidos T4	Y904L	Enzymatics
tampón azul 10x	B011	Enzymatics
25 mM dATP	28406501V	GE
Klenow (3'-5' exo-)	P701-LC-L	Enzymatics
Tampón de ligación rápida 2x	B101	Enzymatics
ADN ligasa (rápida) T4	L603-HC-L	Enzymatics
Adaptador de índice PE/PE oligo mezcla (40 µM)		Invitrogen
Kit de purificación de PCR de MiniElute	28006	QIAGEN
ADN polimerasa de Platinum® Pf	C11708-013	Invitrogen
Cebador de PCR PE 1.0/PE de cebador de índice 1.0 (10 pM)		Invitrogen

(continuación)

Reactivo	Modelo	Fabricante
Cebador de PCR PE 2.0/PE de cebador de índice 2.0 (index, 10 pM)		Invitrogen
Kit de oligonucleótidos de preparación de muestra de multiplexación	PE-400-1001	PE-400-1001
Tampón de hibridación 2xSC	5340721001	Nimblegen
Componente A de hibridación SC	5340721001	Nimblegen
Phusion Mix	F-531L	NEB
Kit de control Phix y de cebadores de secuenciación de multiplexación	PE-400-1002	Illumina
ADN Cot-1	15279-011	Invitrogen

I. Fragmentación de genoma de ADN

La etapa de construcción de la pluralidad de bibliotecas de ADN para la pluralidad de muestras de ADN únicas incluía: romper el genoma de ADN en fragmentos de ADN que tienen un tamaño predeterminado mediante un procedimiento físico o químico, para obtener fragmentos de ADN rotos.

- 5 El genoma de ADN Yan Huang (<http://yh.genomics.org.cn/>) que tiene un peso de 3 µg sin proteína, contaminación de degradación de ARN se tomó como un material inicial, que se rompió usando un equipo de fragmentación de ultrasonido de Covaris-S2 (Covaris, EE.UU.). Los parámetros usados para la fragmentación se muestran en la Tabla 3.

Tabla 3 Configuración de parámetros del equipo de fragmentación de ultrasonido de Covaris-S2

tratamiento 1	relación de carga (%)	10
	Intensidad	10
	ciclo/impulso	1000
	tiempo (s)	60
tratamiento 2	tiempo (s)	0
tratamiento 3	tiempo (s)	0
tratamiento 4	tiempo (s)	0
ciclos		4

- 10 La muestra fragmentada se sometió a ensayo de calidad mediante electroforesis, a continuación, la muestra cualificada se purificó mediante el Kit de purificación de PCR de QIAquick y, a continuación, la muestra purificada se disolvió en 32 µl en un tampón de elución. Las bandas principales de ADN cualificado se centralizan entre 200 pb a 300 pb.

II. Reparación de extremo de fragmentos de ADN

- 15 La etapa de construcción de la pluralidad de bibliotecas de ADN para la pluralidad de muestras de ADN únicas incluía: someter los fragmentos de ADN a reparación de extremo, para obtener fragmentos de ADN con extremo reparado con un extremo fosforilado.

El ADN obtenido a partir de la etapa previa se formuló en un sistema de reacción para la reparación de extremo de acuerdo con la siguiente composición en un tubo de centrifugación de 1,5 ml que se muestra en la Tabla 4.

- 20 Tabla 4 Sistema de reacción para reparación de extremo

Muestra de ADN	75 µl
Tampón de quinasa de polinucleótidos 10x	10 µl
Conjunto de solución dNTP (10 mM cada)	4 µl
ADN polimerasa T4	5 µl

(continuación)

Fragmento de Klenow	1 μ l
Quinasa de polinucleótidos T4	5 μ l
Volumen total	100 μ l

Después de mezclar suavemente, el tubo que contiene 100 μ l de la anterior mezcla de reacción se colocó en un Termomezclador (Eppendorf) establecido a 20 °C y se dejó la reacción durante 30 min. Después de finalizar la reacción, la muestra obtenida se purificó utilizando el kit de purificación de PCR QIAquick (Qiagen). A continuación, la muestra purificada se disolvió en 32 μ l de ddH₂O para una suficiente disolución.

5 III. Adición de una base "A" en los fragmentos de ADN con extremo reparado

La etapa de construcción de la pluralidad de bibliotecas de ADN para la pluralidad de muestras de ADN únicas incluía: añadir una base "A" a los fragmentos de ADN con extremo reparado en el extremo 3' para obtener fragmentos de ADN con la base "A" en el extremo 3'.

10 El ADN obtenido a partir de la etapa previa se formuló en un sistema de reacción para añadir una base A de acuerdo con la siguiente composición en un tubo de centrifugación de 1,5 ml que se muestra en la Tabla 5.

Tabla 5 Sistema de reacción para añadir una base A a los fragmentos de ADN con extremo reparado

ADN	32 μ l
tampón azul 10x	5 μ l
dATP (1mM)	10 μ l
Klenow (3'-5' exo-)	3 μ l
Volumen total	50 μ l

15 Después de mezclar suavemente, el tubo que contiene 50 μ l de la anterior mezcla de reacción se colocó en un Termomezclador (Eppendorf) establecido a 37°C y se dejó la reacción durante 30 min. Después de finalizar la reacción, la muestra obtenida se purificó utilizando el kit de purificación de PCR QIAquick (Qiagen). A continuación, la muestra purificada se disolvió en 15 μ l de ddH₂O para una suficiente disolución.

IV. Ligación de un adaptador de índice

La etapa de construcción de la pluralidad de bibliotecas de ADN para la pluralidad de muestras de ADN únicas incluía: ligar un adaptador de índice a los fragmentos de ADN con la base "A" en el extremo 3', para obtener fragmentos de ADN ligados con el adaptador de índice.

20 El ADN obtenido a partir de la etapa previa se formuló en un sistema de reacción para ligar un adaptador de índice de acuerdo con la siguiente composición en un tubo de centrifugación de 1,5 ml que se muestra en la Tabla 6.

Tabla 6. Sistema de reacción para ligar un adaptador de índice

ADN añadido con una base "A"	15 μ l
Tampón de ligación rápida 2x	25 μ l
Adaptador de índice PE/PE oligo mezcla (40 μ M)	5 μ l
ADN ligasa (rápida) T4	5 μ l
Volumen total	50 μ l

25 Después de mezclar suavemente, el tubo que contiene 50 μ l de la anterior mezcla de reacción se colocó en un Termomezclador (Eppendorf) establecido a 20°C y se dejó la reacción durante 15 min. Después de finalizar la reacción, la muestra obtenida se purificó utilizando el kit de purificación MiniElute (Qiagen). A continuación, la muestra purificada se disolvió en 25 μ l del tampón de elución.

V. PCR previa hibridación

30 La etapa de construcción de la pluralidad de bibliotecas de ADN para la pluralidad de muestras de ADN únicas incluía: amplificar los fragmentos de ADN ligados con el adaptador de índice usando un cebador que comprende una secuencia específica al adaptador de índice, para obtener un producto de amplificación.

Tomando la muestra de ADN en la etapa IV como una plantilla, se llevó a cabo la amplificación de PCR tomando la secuencia que contenía el adaptador como el cebador, en el que el sistema de amplificación se muestra en la Tabla 7.

Tabla 7. Sistema de amplificación de PCR

Secuencia de adaptador que contiene ADN	25 µl
Tampón de amplificación depfx 10x	10 µl
MgSO ₄ (50mM)	4 µl
mezcla de dNTP (10mM)	4 µl
ADN polimerasa de Platinum® Pf	2 µl
Cebador de PCR PE 1.0/PE de cebador de índice 1.0 (10 pM)	10 µl
Cebador de PCR PE 2.0/PE de cebador de índice 2.0 (index, 10 pM)	10 µl
ddH ₂ O	35 µl
Volumen total	100 µl

Condición de reacción de PCR:

94 °C	2 min	} 4 ciclos
94 °C	15 s	
62 °C	30 s	
72 °C	30 s	
72 °C	5 min	

El producto de amplificación obtenido se purificó usando un Kit de purificación de PCR de QIAquick con 30 µl de tampón de elución.

5 VII. Construcción de una biblioteca de mezcla

La biblioteca de mezcla de la pluralidad de muestras de ADN se construyó mediante las siguientes etapas: mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada, tal como con una cantidad igual. En aplicaciones prácticas, la relación adecuada podría determinarse de acuerdo con un requisito para construir la biblioteca de mezcla.

10 VII. Hibridación de una región diana con la sonda de captura

El procedimiento de detección de la microdelección en la región de sitio de secuencia marcada de un cromosoma incluía: hibridar la sonda de captura con una biblioteca de mezcla de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN.

15 1) 450 µg de ADN de COT-1, 3 µg de la biblioteca de mezcla de la etapa VI, 1 nmoles de bloque de adaptador 1 de índice y bloque de adaptador 2 de índice (Kit de oligonucleótidos de preparación de muestra de multiplexación, Illumina) se añadieron a 1,5 ml de tubo de centrifugación. El tubo que contenía la mezcla anterior en 1) se colocó sobre un SpeedVac (Thermo) configurado a 60 °C.

20 2) El tubo de centrifugación que contenía la mezcla secada se añadió con 11,2 µl de agua purificada para una suficiente disolución de ADN, seguido por la adición con 18,5 µl de tampón de hibridación y 7,3 µl de hibridación SC. Después de mezclarla la mezcla obtenida se colocó en un hibridizador (Nimblegen) establecido a 95 °C y se dejó desnaturalizar el ADN durante 10 min.

25 3) Después de agitarse, la muestra obtenida de 2) se centrifugó a la velocidad máxima durante 30 segundos y a continuación se colocó sobre un hibridizador Nimblegen) establecido a 42 °C para preparar la hibridación.

4) El procedimiento de hibridación podría referirse al procedimiento de hibridación de chip de la empresa Nimblegen (NimbleGen Arrays User's Guide, Versión 3.1, 7 de julio de 2009, Roche NimbleGen, Inc.). La cantidad de muestra a someter a ensayo fue de 35 µl, la hibridación se realizó a 42 °C durante de 64 a 72 horas. El producto hibridizado se eluyó con 900 µl de 160mM de NaOH y el producto eluido se purificó usando el Kit de purificación de PCR de MinElute, finalmente, el producto purificado se eluyó con 80 µl de tampón de elución.

30 VIII. PCR después de captura

La biblioteca capturada se sometió a amplificación de PCR con un sistema de reacción de 150 µl de Phusion Mix, 4.2 µk de cada uno de los cebadores corriente arriba y cebadores corriente abajo (Kit de control Phix y de cebadores de secuenciación de multiplexación), 80 µl del producto purificado en el tampón de elución y 85 µl de ddH₂O. Después de dividirse en 6 tubos, conteniendo los tubos la mezcla se sometieron a amplificación de PCR durante 16 ciclos. Tras la reacción de PCR, los 6 tubos que contenían productos amplificados se mezclaron juntos y el producto de amplificación mezclado se sometió a purificación de perlas usan el Kit de purificación de PCR de QIAquick. El producto purificado que tenía una longitud de 300 a 450 pb se eluyó con 50 µl de tampón de elución.

VIII. Detección de biblioteca

El tamaño de longitud y contenido de los fragmentos insertados contenidos en la biblioteca obtenida se detectaron mediante sistema de análisis de Bioanalizador (Agilent, Santa Clara, EE.UU.). Y la concentración de la biblioteca obtenida se cuantificó de forma precisa mediante Q-PCR.

5 X. Análisis de secuenciación y de datos

El procedimiento de detección de la microdelección en la región de sitio de secuencia marcada del cromosoma incluía:

secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación; y

10 analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

La biblioteca cualificada obtenida en la etapa VIII se sometió a secuenciación en ordenador, que podría referirse a la Guía de Usuario HiSeq 2000. Catálogo n.º SY-940-1001 Pare n.º 15011190 Rev B, Illumina.

15 El procedimiento del análisis de información en el ejemplo actual de la presente divulgación se muestra a continuación:

1. recibir datos de secuenciación obtenidos mediante técnica de secuenciación de alto rendimiento y someter los datos de secuenciación a control de calidad:

la etapa de someter de los datos de secuenciación a control de calidad comprendía adicionalmente:

20 filtrar los datos no cualificados en los datos de secuenciación, para obtener datos de secuenciación cualificados.

En el ejemplo actual de la presente divulgación, se usó la técnica de secuenciación de alto rendimiento Hiseq 2000 de Illumina. Después de recibirse, los datos de secuenciación se sometieron a filtración, para filtrar secuencia no cualificada. La secuencia no cualificada cumplió al menos un criterio:

25 contenía más del 50 % de bases que tenían una calidad de secuenciación inferior a 5;

contenía más del 10 % de bases siendo como N;

contenía una secuencia de adaptador de secuenciación después de alinearse a una biblioteca de secuencias de adaptador de secuenciación.

2. Cálculo de valor de profundidad de secuenciación de región de muestra:

la etapa de someter de los datos de secuenciación a control de calidad comprendía adicionalmente:

30 alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;

35 En el ejemplo actual, los datos de secuenciación obtenidos mediante la técnica de secuenciación de alto rendimiento se alinearon con una secuencia de referencia de genoma humano usando SOAPaligner. Se usó el HG19 (<http://genome.ucsc.edu/>) como la secuencia de referencia de genoma humano. El valor de profundidad de secuenciación de la región de muestra se sometió a cálculo estadístico.

3. Cálculo de valor de secuenciación para cada muestra

40 la etapa de someter de los datos de secuenciación a control de calidad comprendía adicionalmente:

filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados.

La etapa de filtrado de los segundos datos de secuenciación no cualificados que comprendía adicionalmente:

45 obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN en los primeros datos de secuenciación cualificados en un orden ascendente; y

filtrar datos de secuenciación no cualificados de los cuales tiene un valor de profundidad de secuenciación más allá de un intervalo de $Q1-1.5IQR$ a $Q3+1.5IQR$, para obtener los segundos datos de secuenciación cualificados.

50 Tal como se muestra en la Fig.20, se filtró una muestra que tiene unos datos de secuenciación siendo como el valor atípico. En Fig.20, el panel izquierdo era un histograma de una mediana de valor de profundidad de secuenciación de una muestra, el panel derecho fue un diagrama de cajas que muestra una mediana de valor de profundidad de secuenciación de una muestra. Podría observarse de ahí que la mediana de los valores de profundidad de secuenciación de la pluralidad de muestras de ADN se concentraba principalmente dentro de 42X a 60X.

4. Cálculo del valor de secuenciación para cada región

55 La etapa de someter de los datos de secuenciación a un control de calidad comprendía adicionalmente:

filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

la etapa de filtrado de los terceros datos de secuenciación no cualificados que comprendía adicionalmente:

- 5 obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la región dada entre distintas muestras de ADN en los segundos datos de secuenciación cualificados en un orden ascendente; y
 filtrar datos de secuenciación no cualificados de los cuales una mediana de los valores de profundidad de
 10 secuenciación de la región dada entre las distintas muestras de ADN es cero o superior a $Q3+1.5IQR$, para obtener los terceros datos de secuenciación cualificados.

Tal como se muestra en la Fig.21, el panel izquierdo era un histograma que mostraba una mediana de valor de profundidad de secuenciación de una región; y el panel derecho fue un diagrama de cajas que muestra una mediana de valor de profundidad de secuenciación de la región. Podría observarse de ahí que la mediana de los valores de profundidad de secuenciación la región de muestras se concentraba principalmente dentro de 35X a 75X.

- 15 5. después de filtrar la región o muestra que tiene un valor atípico de valor de profundidad atípico, los datos de secuenciación de cada muestra se sometieron a normalización, en la que todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región se dividieron por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región. Y, entonces, los valores de profundidad
 20 normalizados se sometieron a ensayo de valor extremo mediante una curva de distribución normal, es decir, la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

- calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 25 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 30 determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico; y
 obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad;
 35 someter el resultado R1 a verificación experimental, para obtener un resultado de verificación;
 determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
 obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al
 40 segundo valor de corte de probabilidad.

Para cada región después de filtrarse el valor atípico del valor de profundidad de secuenciación, seleccionar los datos de secuenciación de la pluralidad de muestras de ADN después de filtrarse el valor atípico del valor de profundidad de secuenciación, la curva de distribución normal con todos los valores no atípicos para la misma región se obtuvo basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de
 45 cada una de la pluralidad de muestras de ADN para la misma región. La probabilidad de valor p. para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico se obtuvo basándose en la curva de distribución normal. Tal como se muestra en la Fig.22, la coordenada X fue muestra, la coordenada Y fue $-\log_{10}$ (valor p); un diagrama, tal como se muestra en la Fig.23, mostró un resultado que ilustra los valores de expectativa y valores de observación del valor p. Mediante tal procedimiento, se obtuvo la probabilidad del valor de profundidad de secuenciación de cada muestra de AND para cada región. 20 regiones de muestras que tienen probabilidades inferiores se sometieron a verificación experimental de reacción de PCR. El valor de corte de probabilidad para detectar la microdelección se determinó basándose en el resultado verificado y la precisión de detección de la microdelección se sometió a evaluación, mediante etapas específicas:

- 55 seleccionar la región parcial de algunas muestras, que incluían la presencia o la ausencia de la microdelección, así como una región con una amplitud más grande de valor p;
 designando cebadores para ambos extremos de distintas regiones y sometiendo las muestras de ADN a amplificación de PCR y, a continuación, sometiendo el producto amplificado a análisis de electroforesis para obtener un resultado que muestre la presencia de la microdelección; y finalmente obtener el valor de corte del valor P que tiene una significancia estadística. Como se muestra en la Tabla 8, para obtener la tasa de positivo verdadero y tasa de falso positivo correspondiente a distinta profundidad de secuenciación, se obtuvo un valor de
 60 área bajo la curva (AUC) como de hasta 0,9968254 verificando el procedimiento de detección de la microdelección a base de técnica de secuenciación de alto rendimiento. Tal como se muestra en la Fig.24, la

coordenada X fue la tasa de falso positivo, la coordenada Y fue la tasa de positivo verdadero. La Fig.24 mostró que podrían seleccionarse distintas probabilidades de valores de corte basándose en distintas tasas de positivo verdadero y tasas de falso positivo.

5 Tabla 8. Distintas tasas de positivo verdadero y tasas de falso positivo obtenidas mediante la selección de distintos valores de corte después de tomar el logaritmo $-\log_{10}$ con el valor p

relación de corte	positivo verdadero	falso positivo
0	1	1
1	1	0,20952381
2	1	0,123809524
3	1	0,076190476
4	1	0,038095238
5	1	0,028571429
6	0,974358974	0,028571429
7	0,897435897	0,019047619
8	0,871794872	0,00952381
9	0,846153846	0,00952381
10	0,846153846	0
11	0,820512821	0
12	0,512820513	0
13	0,512820513	0
14	0,512820513	0
15	0,512820513	0
16	0,512820513	0
17	0,512820513	0
18	0,512820513	0
19	0,512820513	0
20	0,487179487	0
21	0,256410256	0
22	0,256410256	0
23	0	0

6. Además, la presencia o ausencia de microdelección en la región de muestra fue determinar mediante árbol de decisiones en vista de los valores de profundidad de secuenciación de la región de muestra.

La etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

- 10 calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
- 15 calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
- 20 formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3;
- obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras

de ADN, si la primera relación D/S es superior al primer valor de corte de relación;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y
 5 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

En el ejemplo actual, se predeterminó un valor de corte relajado de valor $p \cdot 10^{-6}$ y el resultado positivo se sometió a clasificación mediante el árbol de decisiones, para disminuir adicionalmente la tasa de falso positivo y, a continuación, se obtuvo un sitio de microdelección en la región de muestra.

La etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprendía adicionalmente:

calcular un valor medio y una varianza basándose en todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 15 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 20 determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico;
 obtener un resultado R3 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;
 25 calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;
 calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;
 30 formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
 35 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación; y
 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

La presente divulgación también proporciona un medio legible por ordenador que comprende un orden, configurado para realizar el procedimiento de acuerdo con el procedimiento anteriormente mencionado por un procesador, que se omite por consideración de concisión.

La Fig.25 es un diagrama esquemático que muestra un aparato para detectar una microdelección en una región de sitio de secuencia marcada de un cromosoma de acuerdo con una realización de la presente divulgación. Tal como se muestra en la Fig.25, el aparato de la presente divulgación puede comprender:
 un dispositivo de obtención de sonda de captura 2501, un dispositivo de hibridación 2502, un dispositivo de obtención de datos de secuenciación 2503 y un dispositivo de obtención de resultados de microdelección 2504.

El dispositivo de obtención de sonda de captura 2501 es para obtener una sonda de captura que se corresponde con secuencia de ADN de la región de sitio de secuencia marcada.

El sitio de secuencia marcada (STS) es una secuencia de ADN de copia única corta de conocida ubicación de mapa que sirve como marcador, que puede mapearse únicamente mediante amplificación de PCR, para producir un sitio para la realización de mapa, en concreto, detección con un orden de una serie de STS puede fabricar un mapa de una región genómica. Una sonda es un fragmento de ADN o ARN uncatenario corto (aproximadamente de 20 a 500 pb) para detectar una secuencia de ácido nucleico complementario del mismo.

El dispositivo de hibridación 2502 es para hibridar la sonda de captura con una biblioteca de mezcla de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN.

La expresión "bibliotecas de mezcla" usada en el presente documento se refiere a todas las secuencias de ADN para secuenciar obtenidas mezclando una pluralidad de bibliotecas de ADN de muestras de ADN únicas. Anteriormente se describe, en detalle, etapas específicas para la construcción de la biblioteca de mezcla.

5 El dispositivo de obtención de datos de secuenciación 2503 es para secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación.

El dispositivo de obtención de resultados de microdelección 2504 es para analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

10 Preferentemente, el dispositivo de obtención de sonda de captura 2501 comprende adicionalmente:

una unidad de detección de región, para encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica (GDB);

una unidad de selección de secuencia, para seleccionar una secuencia de ADN cualificada de la región de sitio de secuencia marcada que cumple el requisito para designar la sonda de captura; y

15 una unidad de obtención de sonda de captura, para designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.

La base de datos genómica (GDB, <http://www.gdb.org/>) es una base de datos para guardar y procesar mapa genómico para un proyecto genómico humano (HGP). La enciclopedia de construcciones de DDB para genoma humano, además de construir el mapa genómico, también desarrolla un procedimiento de descripción de contenido genómico de nivel de secuencia, que comprende la variación y descripción de secuencia con otras funciones y fenotipo. La base de datos GDB guarda datos mediante un modelo de objeto, que proporciona servicio de búsqueda en un objeto de datos basado en Web. Un usuario para buscar diversos tipos de objetos y en vista del mapa genómico por medio de modo de gráficos. Por ejemplo, en la base de datos UCSC, la secuencia de ADN en la región de sitio de secuencia marcada en el cromosoma Y puede encontrarse basándose en la coordenada de ubicación de secuencia de referencia de genoma humano Hg19 (<http://genome.ucsc.edu/>). En ejemplos específicos, la coordenada de ubicación en la que se encontró la secuencia de ADN en la región de sitio de secuencia marcada de un cromosoma en la base de datos genómica, que se envió a Roche-NimbleGen u otras empresas que proporcionan servicio de captura de ADN. La designación y síntesis se completó por tales empresas.

30 Preferentemente, el aparato de la presente divulgación comprende adicionalmente un dispositivo de construcción de biblioteca de mezcla, en la que el dispositivo de construcción de biblioteca de mezcla comprende adicionalmente:

una unidad de construcción de pluralidad de bibliotecas de ADN, para construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;

una unidad de mezcla, para mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada; y

35 una unidad de construcción de biblioteca de mezcla, para determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.

Preferentemente, el aparato de la presente divulgación comprende adicionalmente: un dispositivo de control de calidad, en el que el dispositivo de control de calidad comprende adicionalmente:

40 una primera unidad de filtrado, para filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados;

una unidad de cálculo de valor de profundidad de secuenciación, para alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;

45 una segunda unidad de filtrado, para filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados; y

una tercera unidad de filtrado, para filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

50

Preferentemente, el dispositivo de obtención de resultados de microdelección comprende adicionalmente:

una unidad de normalización de valor de profundidad de secuenciación, para someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de profundidad de secuenciación normalizado; y

55 una unidad de obtención de resultados de microdelección, para detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que

muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

5 Preferentemente, la unidad de normalización de valor de profundidad de secuenciación se usa específicamente en la división de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.

Preferentemente, tal como se muestra en la Fig.26, la unidad de obtención de resultados de microdelección comprende adicionalmente:

10 un módulo de obtención de valor medio y varianza 2601, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 un módulo de obtención de curva de distribución normal 2602, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 15 un módulo de cálculo de probabilidad 2603, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal; y
 un primer módulo de determinación 2604, para determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de
 20 profundidad de secuenciación específico y obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

En una realización preferida, Tal como se muestra en la Fig.27, la unidad de obtención de microdelección aún comprende adicionalmente:

25 un módulo de determinación de valor de corte de probabilidad 2701, para someter el resultado R1 a verificación experimental, para obtener un resultado de verificación y determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
 un segundo módulo de determinación 2702, para obtener un resultado R2 que muestra la presencia de la
 30 microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al segundo valor de corte de probabilidad.

Tal como se muestra en la Fig.28, en otra realización preferente, la unidad de obtención de microdelección aún comprende adicionalmente:

35 un primer módulo de obtención de relación D/S 2801, para calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 un primer módulo de obtención de relación D/R 2802, para calcular una primera relación D/R entre el valor de
 40 profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 un primer módulo de obtención de valor de corte de relación 2803, para formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la
 45 primera relación D/R usando algoritmo ID3;
 un tercer módulo de determinación 2804, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es superior al primer valor de corte de relación;
 un cuarto módulo de determinación 2805, para obtener el resultado que muestra la ausencia de la microdelección
 50 en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y
 un quinto módulo de determinación 2806, para obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

55 Tal como se muestra en la Fig.29, en una realización preferente adicional, la unidad de obtención de microdelección aún comprende adicionalmente:

un módulo de obtención de valor medio y varianza 2901, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de

ADN para una misma región;

un módulo de obtención de curva de distribución normal 2902, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;

5 un módulo de cálculo de probabilidad 2903, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;

un sexto módulo de determinación 2904, para determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R3 que muestra la presencia de la microdeleción en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;

10 un segundo módulo de obtención de relación D/S 2905, para calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;

15 un segundo módulo de obtención de relación D/R 2906, para calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;

20 un segundo módulo de determinación de valor de corte de relación 2907, para formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;

un séptimo módulo de determinación 2908, para obtener el resultado que muestra la ausencia de la microdeleción en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;

25 un octavo módulo de determinación 2909, para obtener el resultado que muestra la ausencia de la microdeleción en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación; y

un noveno módulo de determinación 2910, para obtener el resultado que muestra la presencia de la microdeleción en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

El aparato de detección de la microdeleción en la región de sitio de secuencia marcada del cromosoma y el aparato del mismo, mediante el cual obtiene la sonda de captura que cubre una región de sitio de secuencia marcada completa de un cromosoma y captura secuencia de ADN de la región de sitio de secuencia marcada a partir de una pluralidad de muestras de ADN después de hibridarse con una bibliotecas de mezcla de la pluralidad de muestras de ADN, puede detectar de forma eficaz y precisa microdeleciones informadas o no informadas en una región de sitio de secuencia marcada de un cromosoma con una gran cantidad de muestras de ADN; además el procedimiento de análisis estadístico matemático usado por la presente divulgación puede basarse en una distribución normal, un procedimiento de análisis de árbol de decisiones o una combinación de los mismos seguido por una verificación experimentos, que es científica, estable, con una alta sensibilidad, una tasa de falso positivos baja, usándose en el análisis eficaz de microdeleciones.

Aunque se han mostrado y descrito realizaciones explicativas, se apreciaría por los expertos en la técnica que las realizaciones anteriores no pueden interpretarse para que limiten la presente divulgación y pueden realizarse cambios, alternativas y modificaciones en las realizaciones sin alejarse del espíritu, principios y ámbito de la presente divulgación.

En lo sucesivo, se describen realizaciones preferentes para facilitar una comprensión más profunda de la invención:

1. Un procedimiento de detección de una microdeleción en una región de sitio de secuencia marcada de un cromosoma, que comprende:

50 obtener una sonda de captura que se corresponde con secuencia de ADN de la región de sitio de secuencia marcada;

hibridar la sonda de captura con una biblioteca de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN; secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación; y

55 analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdeleción en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

2. El procedimiento de la realización 1, en el que la etapa de análisis de los datos de secuenciación usando el ensayo estadístico matemático comprende adicionalmente:

60 someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a

- normalización, para obtener un valor de profundidad de secuenciación normalizado;
 detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia
 marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento
 estadístico matemático, para obtener el resultado que muestra la presencia o la ausencia de la microdelección
 en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.
- 5
3. El procedimiento de la realización 2, en el que la etapa de someter el valor de profundidad de secuenciación
 de la región de sitio de secuencia marcada a normalización comprende adicionalmente:
 dividir todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para
 una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para
 obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.
- 10
4. El procedimiento de la realización 2, en el que la etapa de detectar el valor atípico del valor de profundidad de
 secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:
- calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación
 normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose
 en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la
 pluralidad de muestras de ADN para la misma región;
 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de
 profundidad de secuenciación específico, basándose en la curva de distribución normal;
 determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la
 pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico; y
 obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de
 una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación
 específico es inferior al primer valor de corte de probabilidad.
- 15
- 20
- 25
5. El procedimiento de la realización 4, en el que después de la etapa de obtener el resultado R1, el
 procedimiento comprende adicionalmente: someter el resultado R1 a verificación experimental, para obtener un
 resultado de verificación;
- determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el
 segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad;
 obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de
 una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior
 al segundo valor de corte de probabilidad.
- 30
6. El procedimiento de la realización 2, en el que la etapa de detectar el valor atípico del valor de profundidad de
 secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:
- calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de
 ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una
 de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio
 de secuencia marcada de la pluralidad de muestras de ADN;
 calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de
 ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las
 regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la
 pluralidad de muestras de ADN;
 formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un
 segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3,
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de
 muestras de ADN, si la primera relación D/S es superior al primer valor de corte de relación;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de
 muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera
 relación D/R es superior al segundo valor de corte de relación;
 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de
 muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera
 relación D/R es inferior al segundo valor de corte de relación.
- 35
- 40
- 45
- 50
7. El procedimiento de la realización 2, en el que la etapa de detectar el valor atípico del valor de profundidad de
 secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:
- calcular un valor medio y una varianza basándose en todos los valores de profundidad normalizados de cada
 una de la pluralidad de muestras de ADN para una misma región;
 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose
 en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la
- 55

- pluralidad de muestras de ADN para la misma región;
 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico;
 5 obtener un resultado R3 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;
 calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado R3;
 10 calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;
 15 formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación;
 20 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.
- 25 8. El procedimiento de la realización 1, en el que la etapa de obtener la sonda de captura comprende adicionalmente:
 encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica;
 seleccionar una secuencia de ADN cualificada de la región de sitio de secuencia marcada que cumple el requisito para designar la sonda de captura;
 30 designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.
9. El procedimiento de la realización 1, en el que la biblioteca de mezcla de la pluralidad de muestras de ADN se construye mediante las siguientes etapas:
 35 construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;
 mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada;
 determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.
- 40 10. El procedimiento de la realización 8, en el que la etapa de construcción de una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, para cada una de la pluralidad de muestras de ADN únicas, la biblioteca de ADN se construye mediante la siguientes etapas:
 romper un genoma de ADN en fragmentos de ADN que tienen un tamaño predeterminado mediante procedimiento físico o químico, para obtener fragmentos de ADN;
 45 someter los fragmentos de ADN a reparación de extremo, para obtener fragmentos de ADN con extremo reparado con un extremo fosforilado;
 añadir una base "A" a los fragmentos de ADN con extremo reparado en el extremo 3' para obtener fragmentos de ADN con la base "A" en el extremo 3';
 ligar un adaptador de índice a los fragmentos de ADN con la base "A" en el extremo 3', para obtener fragmentos de ADN ligados con el adaptador de índice;
 50 amplificar los fragmentos de ADN ligados con el adaptador de índice usando un cebador que comprende una secuencia específica al adaptador de índice, para obtener un producto de amplificación; y
 determinar el producto de amplificación como la biblioteca de ADN, si el producto de amplificación cumple el requisito de control de calidad.
- 55 11. El procedimiento de la realización 1, en el que después de la etapa de secuenciación de la secuenciación del ADN capturado y antes de la etapa de análisis de los datos de secuenciación, el procedimiento comprende adicionalmente:
 someter los datos de secuenciación a un control de calidad.
12. El procedimiento de la realización 10, en el que la etapa de someter de los datos de secuenciación a control de calidad comprende adicionalmente:

filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados;
 alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;
 filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados;
 filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

13. El procedimiento de la realización 11, en el que la etapa de filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener los primeros datos de secuenciación cualificados que comprenden adicionalmente:

obtener un primer conjunto de datos de secuenciación filtrando datos no cualificados que contiene más de una proporción predeterminada de bases que tienen una baja calidad de los datos de secuenciación;
 obtener un segundo conjunto de datos de secuenciación filtrando datos no cualificados que contiene más del 10 % de bases no determinadas a partir del primer conjunto de datos de secuenciación;
 obtener un tercer conjunto de datos de secuenciación filtrando datos no cualificados que contiene una secuencia de adaptador de secuenciación a partir del segundo conjunto de datos de secuenciación después de haber sido alineado a una biblioteca de secuencias de adaptador de secuenciación;
 obtener los primeros datos de secuenciación cualificados filtrando datos no cualificados de los cuales contiene una secuencia exógena a partir del tercer conjunto de datos de secuenciación después de haber sido alienados a todas las secuencias exógenas introducidas.

14. El procedimiento de la realización 11, en el que la etapa de filtrado de los segundos datos de secuenciación no cualificados que comprende adicionalmente:

obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN en los primeros datos de secuenciación cualificados en un orden ascendente;
 filtrar datos de secuenciación no cualificados de los cuales tiene un valor de profundidad de secuenciación más allá de un intervalo de $Q1-1.5IQR$ a $Q3+1.5IQR$, para obtener los segundos datos de secuenciación cualificados.

15. El procedimiento de la realización 11, en el que la etapa de filtrado de los terceros datos de secuenciación no cualificados que comprende adicionalmente:

obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la región dada entre distintas muestras de ADN en los segundos datos de secuenciación cualificados en un orden ascendente;
 filtrar datos de secuenciación no cualificados de los cuales una mediana de los valores de profundidad de secuenciación de la región dada entre las distintas muestras de ADN es cero o superior a $Q3+1.5IQR$, para obtener los terceros datos de secuenciación cualificados.

Ejemplo de referencia 16. Un aparato para la detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma, que comprende:

una sonda de captura que obtiene módulo, para obtener una sonda de captura que se corresponde con secuencia de ADN de la región de sitio de secuencia marcada;
 un módulo de hibridación, para hibridar la sonda de captura con una biblioteca de mezcla de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN;
 un módulo de obtención de datos de secuenciación, para secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación; y
 un módulo de obtención de resultados de microdelección, para analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

Ejemplo de referencia 17. El aparato del ejemplo de referencia 15, en el que el módulo de obtención de resultados de microdelección comprende adicionalmente:

una unidad de normalización de valor de profundidad de secuenciación, para someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de

profundidad de secuenciación normalizado;

una unidad de obtención de resultados de microdelección, para detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

Ejemplo de referencia 18. El aparato del ejemplo de referencia 16, en el que la unidad de normalización de valor de profundidad de secuenciación se usa en la división de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.

Ejemplo de referencia 19. El aparato del ejemplo de referencia 16, en el que la unidad de obtención de resultados de microdelección comprende adicionalmente:

un módulo de obtención de valor medio y varianza, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;

un módulo de obtención de curva de distribución normal, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;

un módulo de cálculo de probabilidad, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal; y

un primer módulo de determinación, para determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad.

Ejemplo de referencia 20. El aparato del ejemplo de referencia 19, en el que la unidad de obtención de microdelección aún comprende adicionalmente:

un módulo de determinación de valor de corte de probabilidad, para someter el resultado R1 a verificación experimental, para obtener un resultado de verificación y determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad;

un segundo módulo de determinación, para obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al segundo valor de corte de probabilidad.

Ejemplo de referencia 21. El aparato del ejemplo de referencia 16, en el que la unidad de obtención de microdelección aún comprende adicionalmente:

un primer módulo de obtención de relación D/S, para calcular una primera relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;

un primer módulo de obtención de relación D/R, para calcular una primera relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;

un primer módulo de obtención de valor de corte de relación, para formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3;

un tercer módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es superior al primer valor de corte de relación;

un cuarto módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es superior al segundo valor de corte de relación; y

un quinto módulo de determinación, para obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la primera relación D/S es inferior al primer valor de

corte de relación y la primera relación D/R es inferior al segundo valor de corte de relación.

Ejemplo de referencia 22. El aparato del ejemplo de referencia 16, en el que la unidad de obtención de microdelección aún comprende adicionalmente:

- 5 un módulo de obtención de valor medio y varianza, para calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
- un módulo de obtención de curva de distribución normal, para obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la
- 10 misma región;
- un módulo de cálculo de probabilidad, para calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
- un sexto módulo de determinación, para determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestra de ADN en una región dada con el valor de profundidad de secuenciación específico y obtener un resultado R3 que muestra la presencia de la
- 15 microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al tercer valor de corte de probabilidad;
- un segundo módulo de obtención de relación D/S, para calcular una segunda relación D/S entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN en el resultado
- 20 R3;
- un segundo módulo de obtención de relación D/R, para calcular una segunda relación D/R entre el valor de profundidad normalizador de la pluralidad de muestras de ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de todas las regiones en el resultado R3;
- un segundo módulo de determinación de valor de corte de relación, para formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
- 30 un séptimo módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
- un octavo módulo de determinación, para obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor
- 35 de corte de relación y la segunda relación D/R es superior al cuarto valor de corte de relación;
- un noveno módulo de determinación, para obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el resultado R3 es inferior al cuarto valor de corte de relación.

40 Ejemplo de referencia 23. El aparato del ejemplo de referencia 16, en el que el módulo de obtención de sonda de captura comprende adicionalmente:

- una unidad de detección de región, para encontrar la secuencia de ADN de la región de sitio de secuencia marcada dentro de una base de datos genómica;
- una unidad de selección de secuencia, para seleccionar una secuencia de ADN cualificada de la región de
- 45 sitio de secuencia marcada que cumple el requisito para designar la sonda de captura; y
- una unidad de obtención de sonda de captura, para designar y sintetizar la sonda de captura basándose en la secuencia de ADN cualificada.

50 Ejemplo de referencia 24. El aparato del ejemplo de referencia 16, que comprende adicionalmente un módulo de construcción de biblioteca de mezcla, en la que el módulo de construcción de biblioteca de mezcla comprende adicionalmente:

- una unidad de construcción de pluralidad de bibliotecas de ADN, para construir una pluralidad de bibliotecas de ADN para una pluralidad de muestras de ADN únicas, en la que cada una de la pluralidad de bibliotecas de ADN tiene un adaptador distinto respectivamente;
- una unidad de mezcla, para mezclar la pluralidad de bibliotecas de ADN con una relación predeterminada; y
- 55 una unidad de construcción de biblioteca de mezcla, para determinar la mezcla resultado como la biblioteca de mezcla, si la mezcla resultante cumple un requisito de control de calidad.

Ejemplo de referencia 25. El aparato del ejemplo de referencia 15, que comprende adicionalmente un módulo de control de calidad, en la que el módulo de control de calidad comprende adicionalmente:

una primera unidad de filtrado, para filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de secuenciación cualificados;

5 una unidad de cálculo de valor de profundidad de secuenciación, para alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de muestras de ADN;

10 una segunda unidad de filtrado, para filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación calculado, para obtener segundos datos de secuenciación cualificados; y una tercera unidad de filtrado, para filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia marcada dada.

26. Un medio legible que comprende un orden, configurado para realizar el procedimiento de acuerdo con cualquiera una de las realizaciones 1 a 15 mediante un procesador.

15

REIVINDICACIONES

1. Un procedimiento de detección de una microdelección en una región de sitio de secuencia marcada de un cromosoma, que comprende:

5 obtener una sonda de captura correspondiente a la secuencia de ADN de la región de sitio de secuencia marcada;
 hibridar la sonda de captura con una biblioteca de una pluralidad de muestras de ADN, para capturar la secuencia de ADN de la región de sitio de secuencia marcada a partir de la pluralidad de muestras de ADN;
 secuenciar la secuencia de ADN capturada de la región de sitio de secuencia marcada para obtener datos de secuenciación; y
 10 analizar los datos de secuenciación usando un procedimiento estadístico matemático, para obtener un resultado que muestra una presencia o una ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.

2. El procedimiento de la reivindicación 1, en el que la etapa de análisis de los datos de secuenciación usando el ensayo estadístico matemático comprende adicionalmente:

15 someter un valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización, para obtener un valor de profundidad de secuenciación normalizado;
 detectar un valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada basándose en el valor de profundidad de secuenciación normalizado usando el procedimiento estadístico matemático, para obtener el resultado que muestra la presencia o la ausencia de la microdelección en la región de sitio de secuencia marcada del cromosoma en cada una de la pluralidad de muestras de ADN.
 20

3. El procedimiento de la reivindicación 2, en el que la etapa de someter el valor de profundidad de secuenciación de la región de sitio de secuencia marcada a normalización comprende adicionalmente:

dividir todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN para una misma región por un valor de profundidad promedio de cada una de la pluralidad de muestras de ADN, para obtener los valores de profundidad normalizados de la pluralidad de muestras de ADN para la misma región.
 25

4. El procedimiento de la reivindicación 2, en el que la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

calcular un valor medio y una varianza basándose en todos los valores de profundidad de secuenciación normalizados de cada una de la pluralidad de muestras de ADN para una misma región;
 30 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la pluralidad de muestras de ADN para la misma región;
 calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de profundidad de secuenciación específico, basándose en la curva de distribución normal;
 35 determinar un primer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad de muestras de ADN en una región dada con el valor de profundidad de secuenciación específico;
 obtener un resultado R1 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación específico es inferior al primer valor de corte de probabilidad,
 40 en el que opcionalmente, después de la etapa de obtener el resultado R1, el procedimiento comprende adicionalmente:

someter el resultado R1 a verificación experimental, para obtener un resultado de verificación;
 determinar un segundo valor de corte de probabilidad basándose en el resultado de verificación, en el que el segundo valor de corte de probabilidad es inferior al primer valor de corte de probabilidad; y
 45 obtener un resultado R2 que muestra la presencia de la microdelección en la región dada, si la probabilidad de una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad específico es inferior al segundo valor de corte de probabilidad.

5. El procedimiento de la reivindicación 2, en el que la etapa de detectar el valor atípico del valor de profundidad de secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:

50 calcular una primera relación D/S entre el valor de profundidad normalizado de la pluralidad de muestras de ADN de una región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la pluralidad de muestras de ADN, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 calcular una primera relación D/R entre el valor de profundidad normalizado de la pluralidad de muestras de ADN de la región dada y una mediana de los valores de profundidad de secuenciación de todas las regiones, basándose en el valor de profundidad normalizado de la región de sitio de secuencia marcada de la pluralidad de muestras de ADN;
 55 formar un primer valor de corte de relación con la primera relación D/S usando algoritmo ID3 y formar un

- segundo valor de corte de relación con la primera relación D/R usando algoritmo ID3,
 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la primera relación D/S es superior al primer valor de corte de relación;
- 5 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es
 superior al segundo valor de corte de relación;
 obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la primera relación D/S es inferior al primer valor de corte de relación y la primera relación D/R es
 inferior al segundo valor de corte de relación.
- 10 6. El procedimiento de la reivindicación 2, en el que la etapa de detectar el valor atípico del valor de profundidad de
 secuenciación de la región de sitio de secuencia marcada comprende adicionalmente:
- calcular un valor medio y una varianza basándose en todos los valores de profundidad normalizados de cada una
 de la pluralidad de muestras de ADN para una misma región;
- 15 obtener una curva de distribución normal con todos los valores no atípicos para la misma región, basándose en
 el valor medio calculado y varianza de todos los valores de profundidad normalizados de cada una de la
 pluralidad de muestras de ADN para la misma región;
- calcular una probabilidad para cada una de la pluralidad de muestras de ADN en cada región con un valor de
 profundidad de secuenciación específico, basándose en la curva de distribución normal;
- 20 determinar un tercer valor de corte de probabilidad basándose en la probabilidad para cada una de la pluralidad
 de muestras de ADN en una región dada con el valor de profundidad de secuenciación específico;
- obtener un resultado R3 que muestra la presencia de la microdelección en la región dada, si la probabilidad de
 una de la pluralidad de muestras de ADN en la región dada con el valor de profundidad de secuenciación
 específico es inferior al tercer valor de corte de probabilidad;
- 25 calcular una segunda relación D/S entre el valor de profundidad normalizado de la pluralidad de muestras de
 ADN de la región dada y una mediana de todos los valores de profundidad de secuenciación de cada una de la
 pluralidad de muestras de ADN en el resultado R3;
- calcular una segunda relación D/R entre el valor de profundidad normalizado de la pluralidad de muestras de
 ADN de la región dada y una mediana de los valores de profundidad de secuenciación de todas las regiones en
 el resultado R3;
- 30 formar un tercer valor de corte de relación con la segunda relación D/S usando algoritmo ID3 y formar un cuarto
 valor de corte de relación con la segunda relación D/R usando algoritmo ID3;
- obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la segunda relación D/S es superior al tercer valor de corte de relación;
- 35 obtener el resultado que muestra la ausencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R es
 superior al cuarto valor de corte de relación;
- obtener el resultado que muestra la presencia de la microdelección en la región dada de la pluralidad de muestras
 de ADN, si la segunda relación D/S es inferior al tercer valor de corte de relación y la segunda relación D/R en el
 resultado R3 es inferior al cuarto valor de corte de relación.
- 40 7. El procedimiento de la reivindicación 1, en el que después de la etapa de secuenciación de la secuenciación del
 ADN capturado y antes de la etapa de análisis de los datos de secuenciación, el procedimiento comprende
 adicionalmente:
- someter los datos de secuenciación a un control de calidad,
 en el que opcionalmente, la etapa de someter de los datos de secuenciación a control de calidad comprende
 45 adicionalmente:
- filtrar los primeros datos no cualificados en los datos de secuenciación, para obtener primeros datos de
 secuenciación cualificados;
- 50 alinear la primera secuenciación cualificada a una secuencia de genoma de referencia mediante software de
 alineación de secuencia corta y calcular un primer parámetro de correlación del valor de profundidad de
 secuenciación de cada una de la pluralidad de muestras de ADN y un segundo parámetro de correlación del
 valor de profundidad de secuenciación para una región de sitio de secuencia marcada entre la pluralidad de
 muestras de ADN;
- filtrar segundos datos de secuenciación no cualificados basándose en el primer parámetro de correlación
 calculado, para obtener segundos datos de secuenciación cualificados;
- 55 filtrar terceros datos de secuenciación no cualificados basándose en el segundo parámetro de correlación
 calculado, para obtener terceros datos de secuenciación cualificados de la región de sitio de secuencia
 marcada dada,
- en el que opcionalmente, la etapa de filtrado de los segundos datos de secuenciación no cualificados comprende
 adicionalmente:
- 60

obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la pluralidad de muestras de ADN en los primeros datos de secuenciación cualificados en un orden ascendente;

- 5 filtrando datos de secuenciación no cualificados de los cuales tiene un valor de profundidad de secuenciación más allá de un intervalo de $Q1-1,5IQR$ a $Q3+1,5IQR$, para obtener los segundos datos de secuenciación cualificados,

en el que opcionalmente, la etapa de filtrado de los terceros datos de secuenciación no cualificados comprende adicionalmente:

- 10 obtener un cuartil inferior Q1, un cuartil superior Q3 y un intercuartil IQR mediante función cuartil y clasificando todos los valores de profundidad de secuenciación de la región dada entre distintas muestras de ADN en los segundos datos de secuenciación cualificados en un orden ascendente;

filtrar datos de secuenciación no cualificados de los cuales una mediana de los valores de profundidad de secuenciación de la región dada entre las distintas muestras de ADN es cero o superior a $Q3+1,5IQR$, para obtener los terceros datos de secuenciación cualificados.

- 15 8. Un medio legible por ordenador que comprende instrucciones configuradas para realizar el procedimiento de acuerdo con una cualquiera una de las reivindicaciones 1 a 7 mediante un procesador.

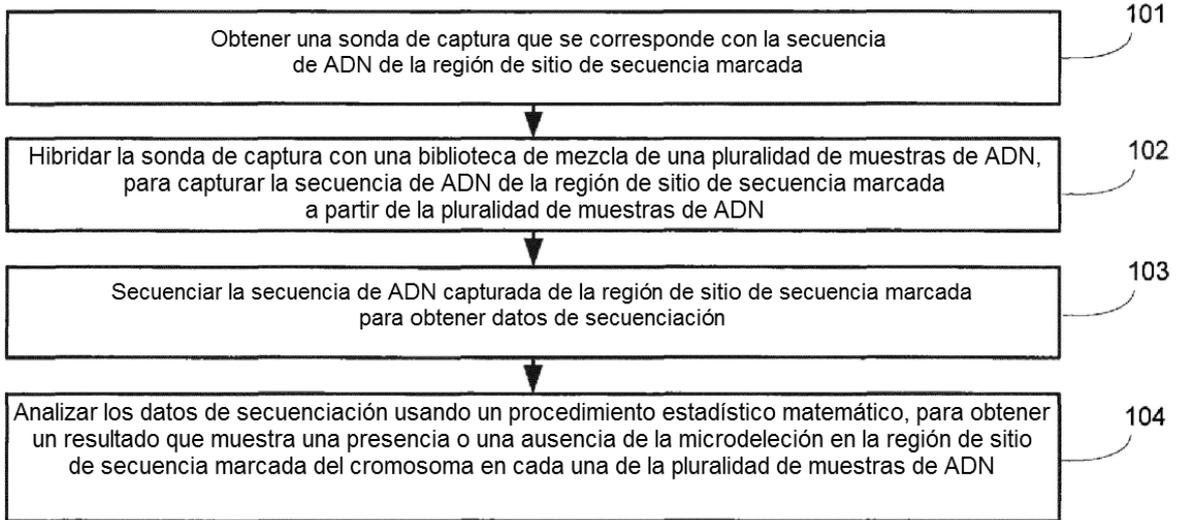


Fig.1

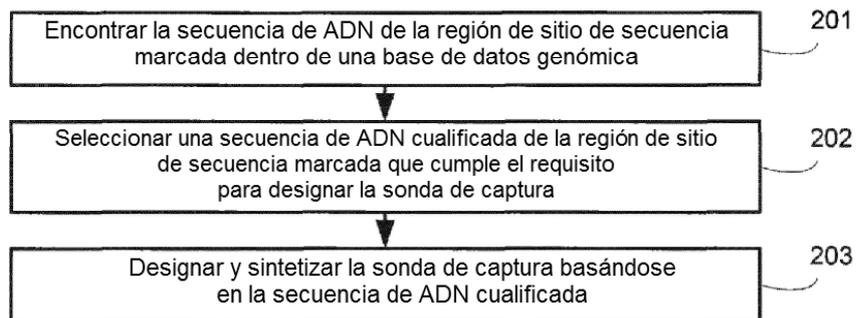


Fig.2

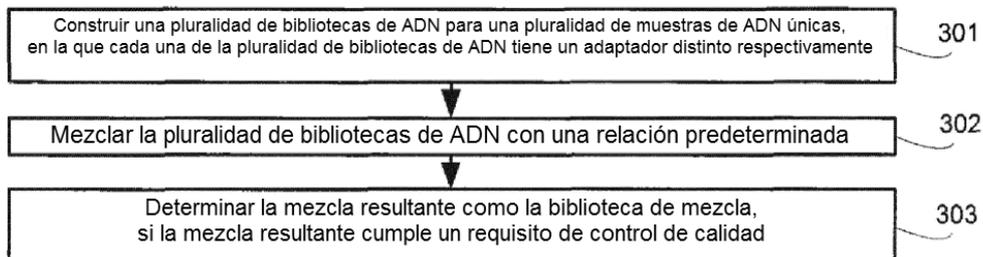


Fig.3

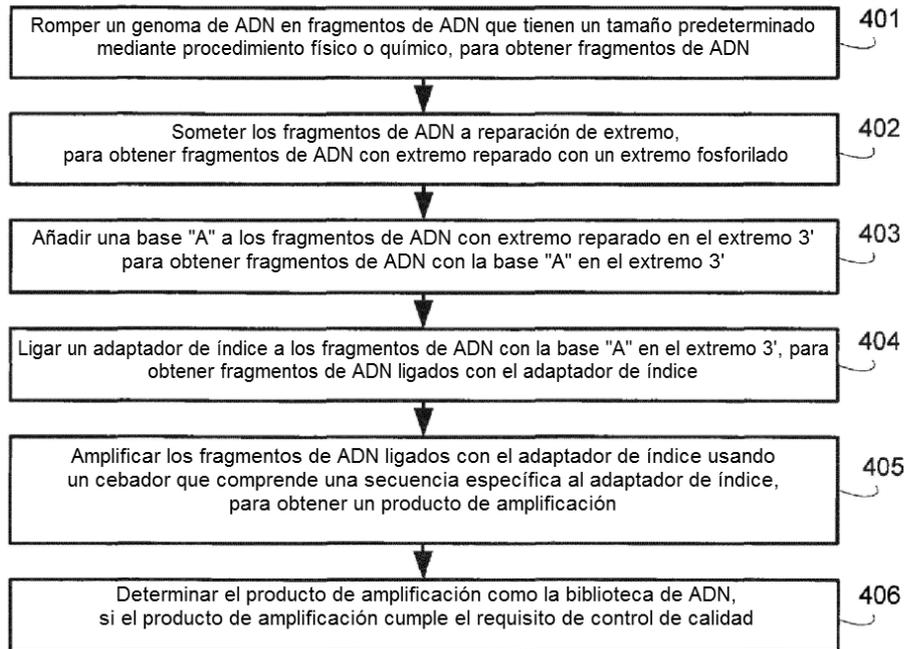


Fig.4

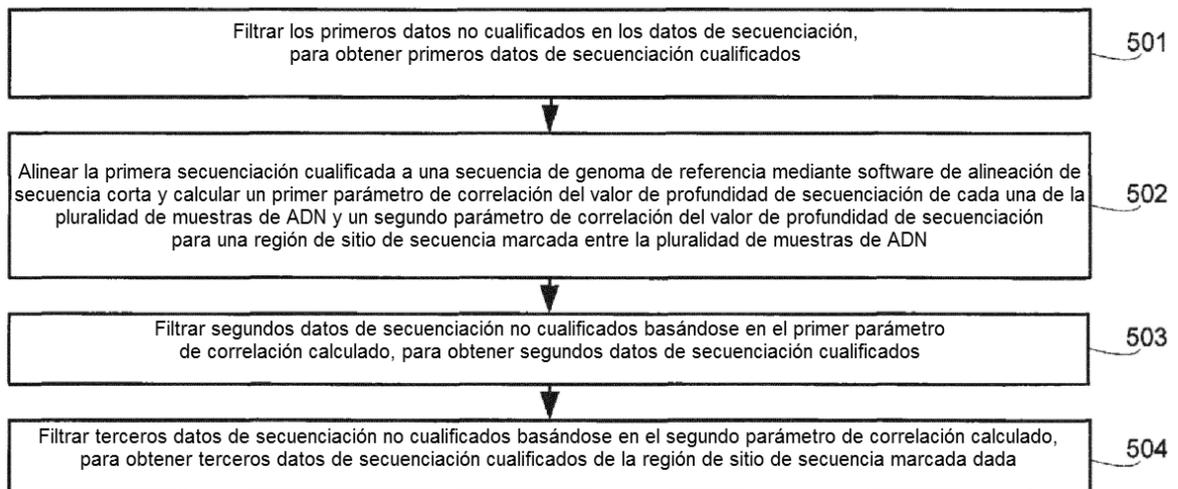


Fig.5

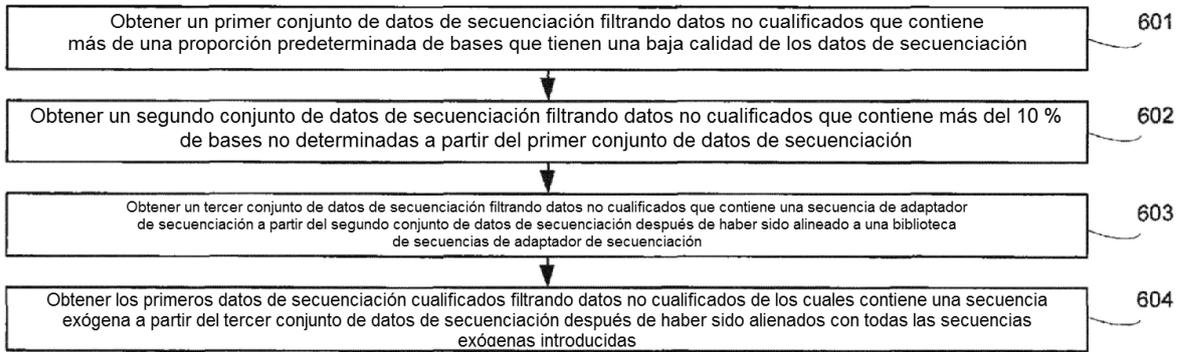


Fig.6

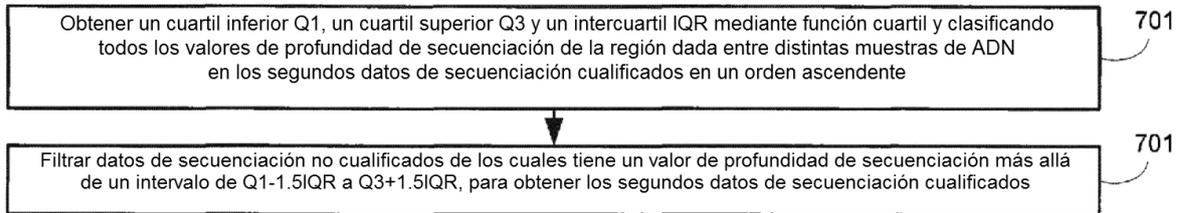


Fig.7

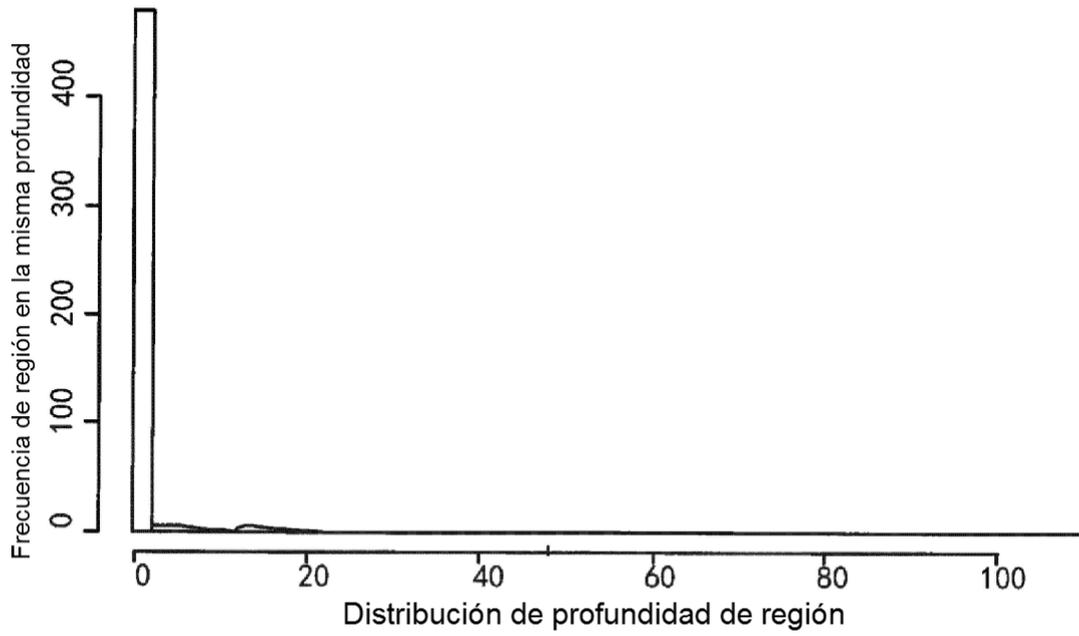


Fig.8

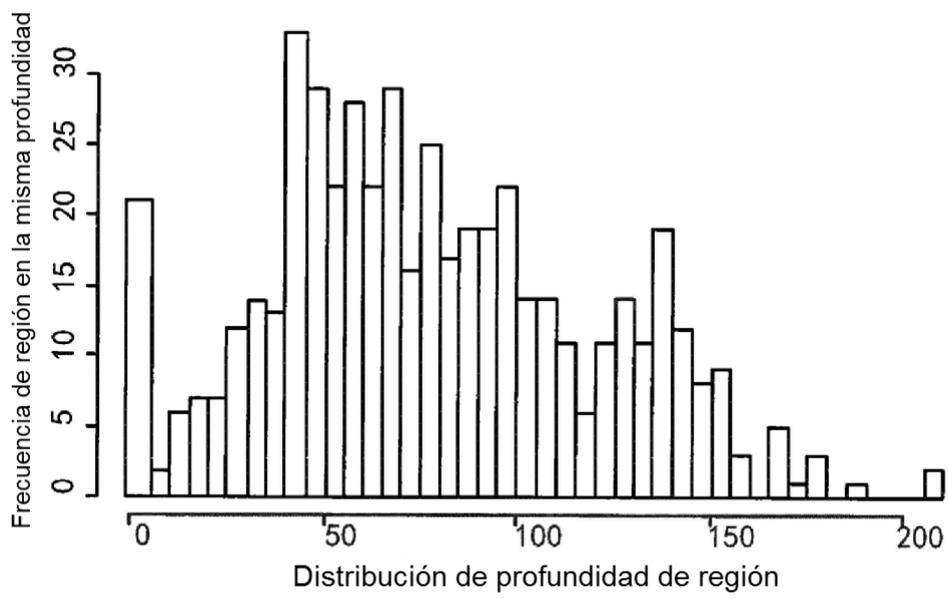


Fig.9

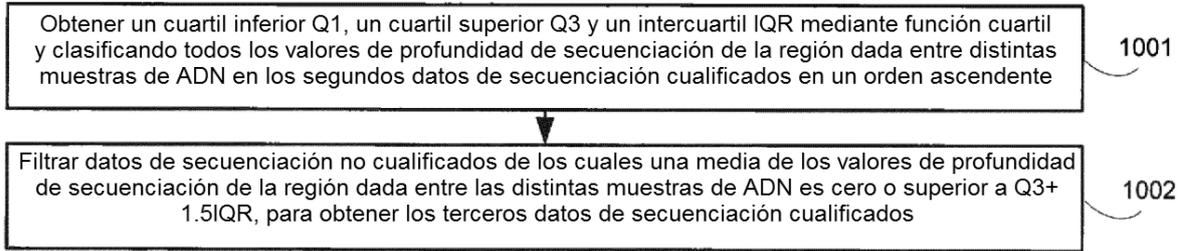


Fig.10

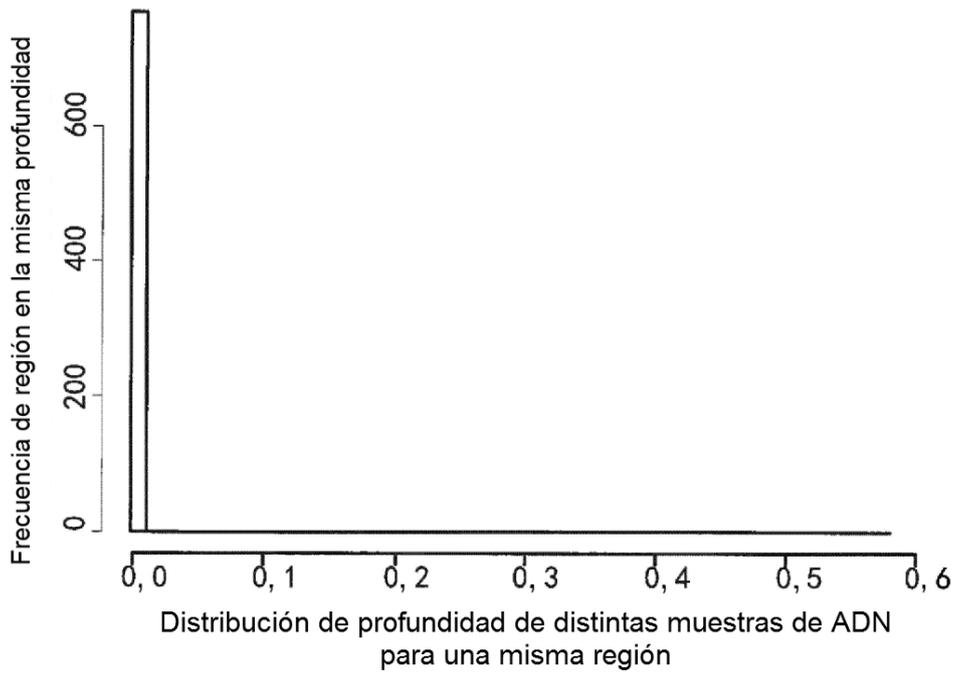


Fig.11

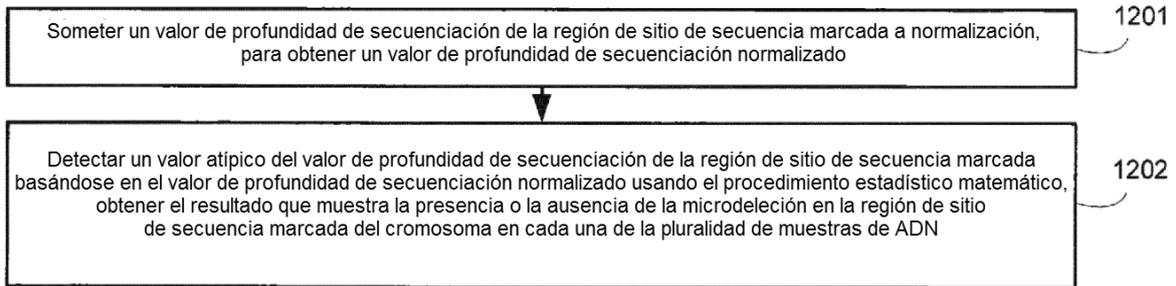


Fig.12

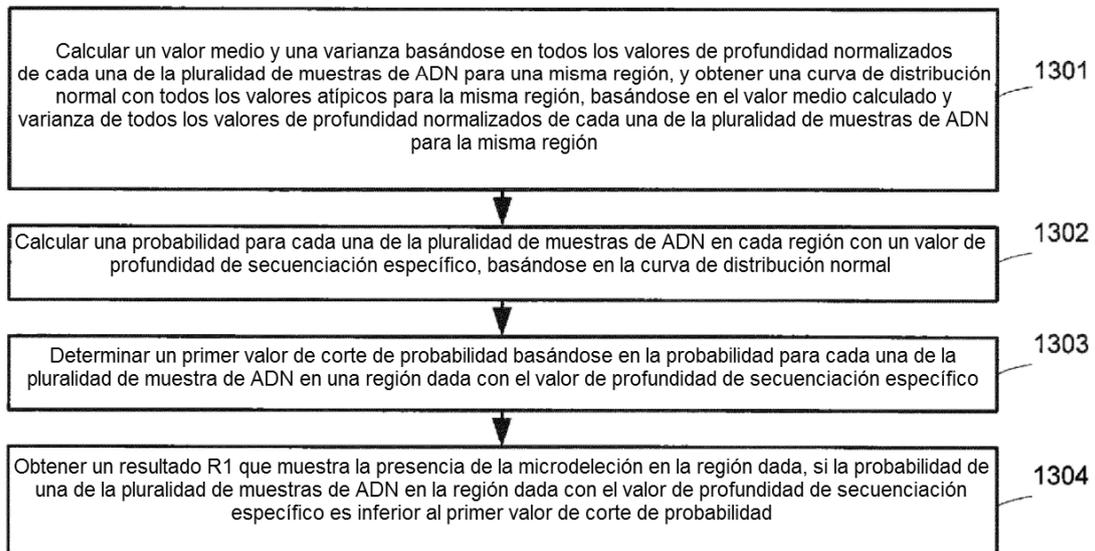


Fig.13

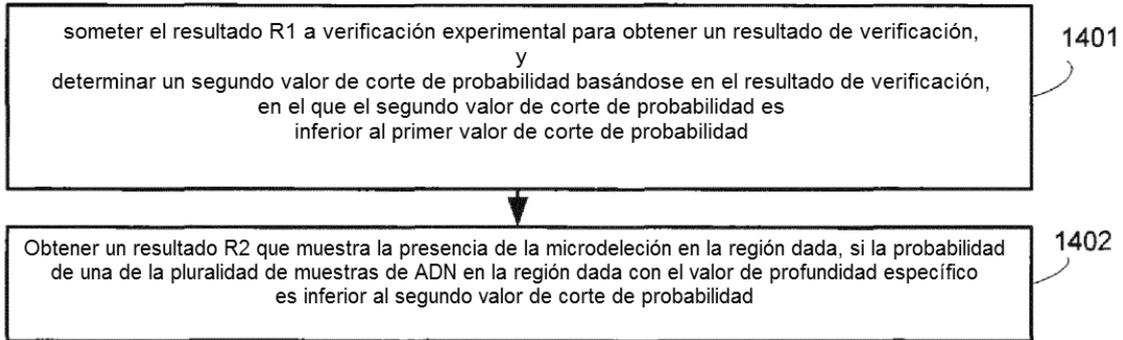


Fig.14

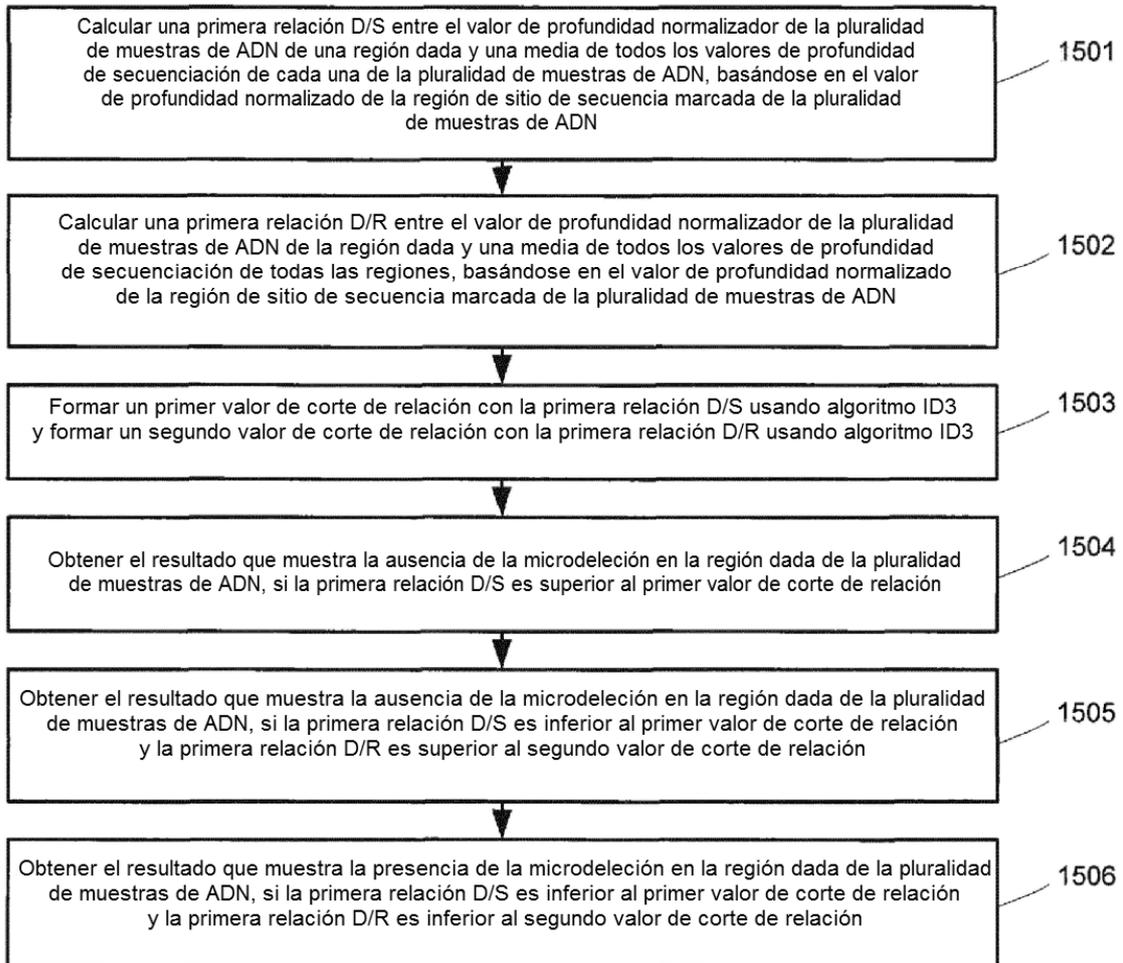


Fig.15

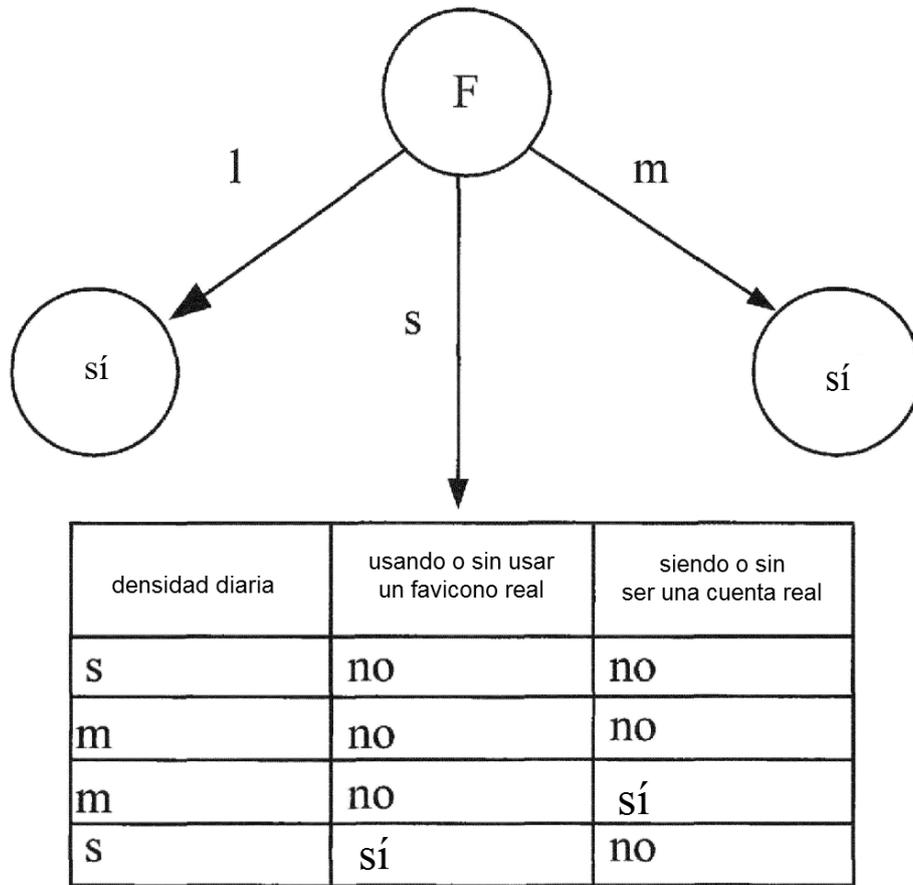


Fig.16

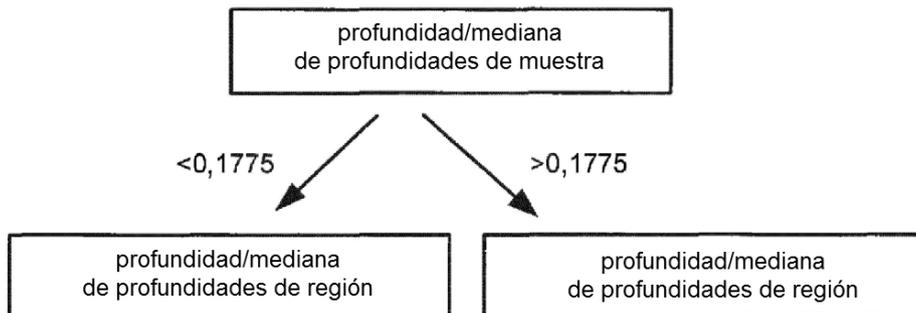


Fig.17

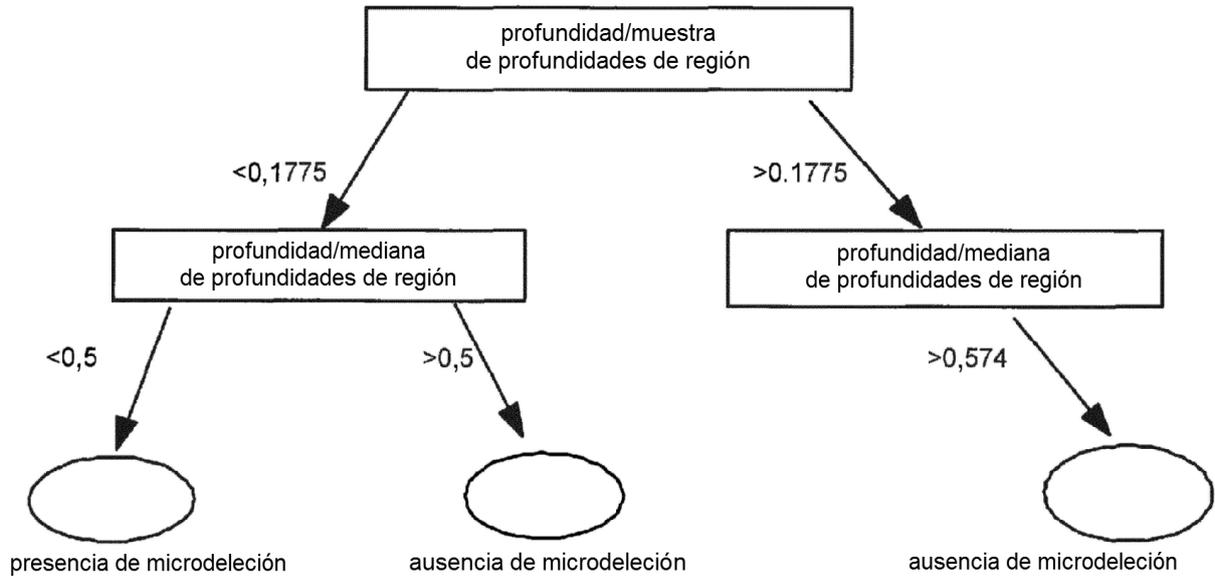


Fig.18

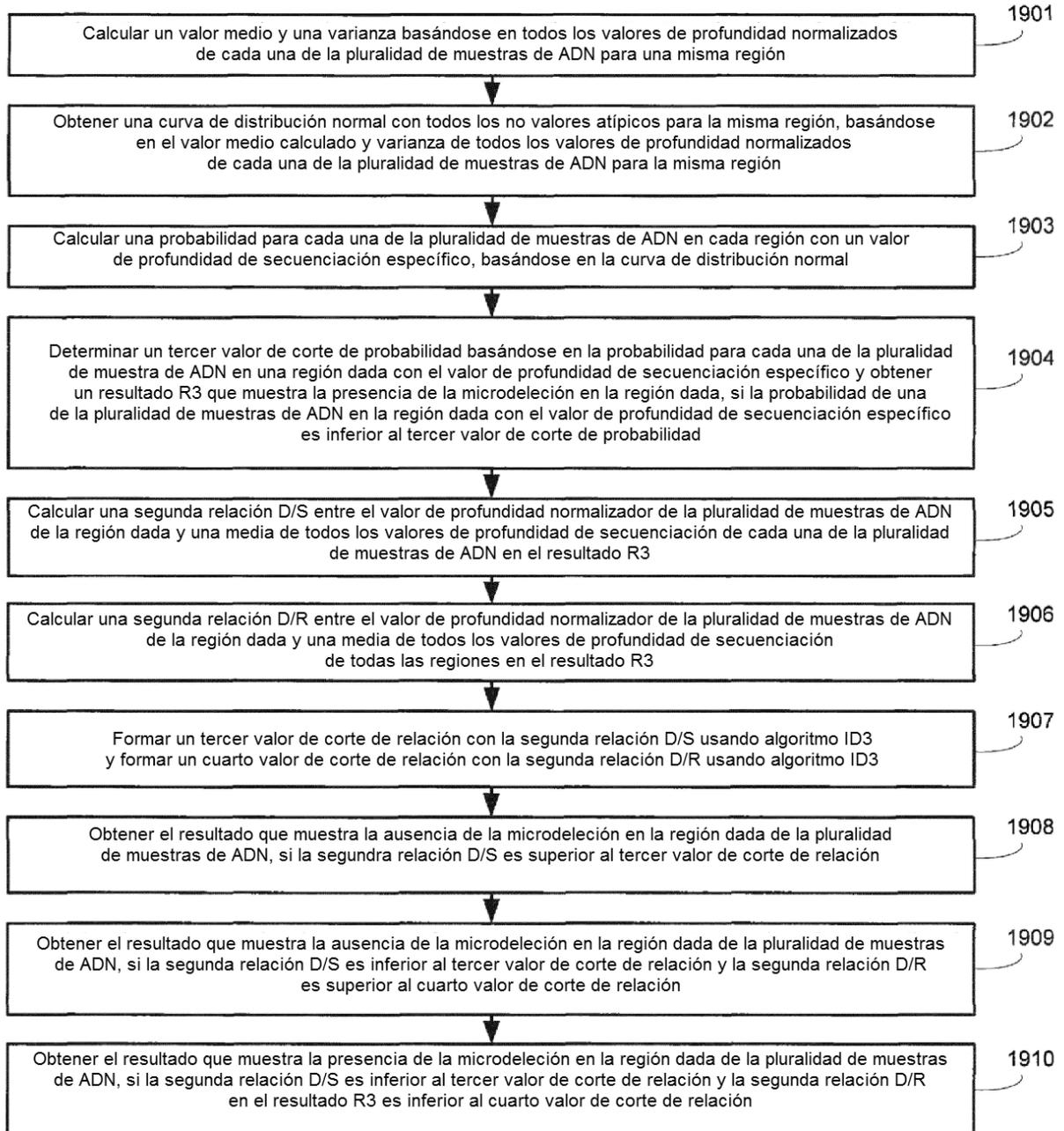


Fig.19

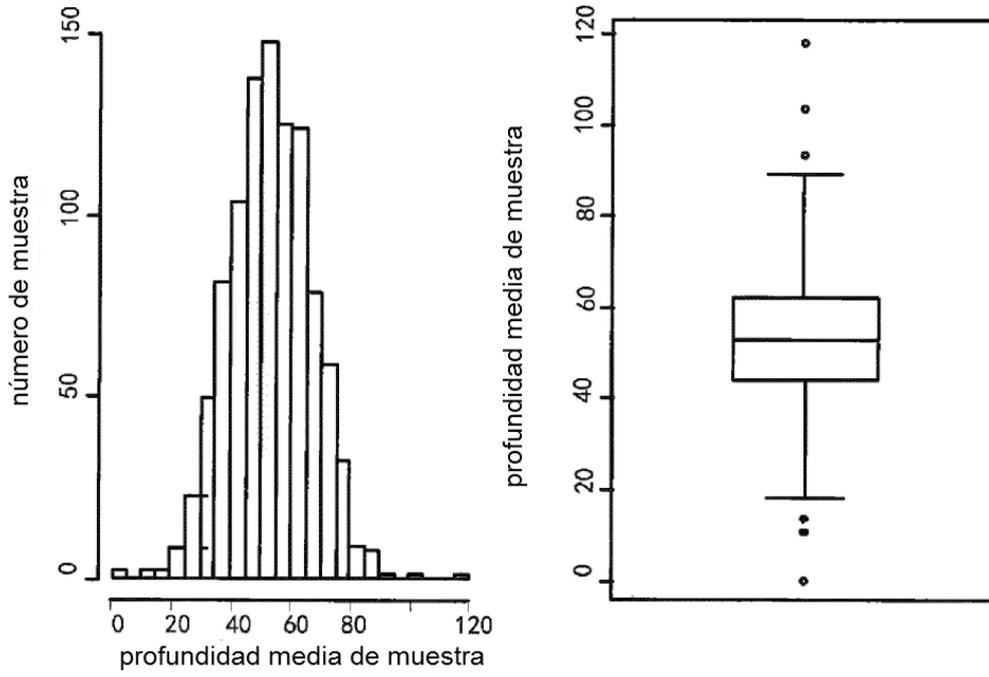


Fig.20

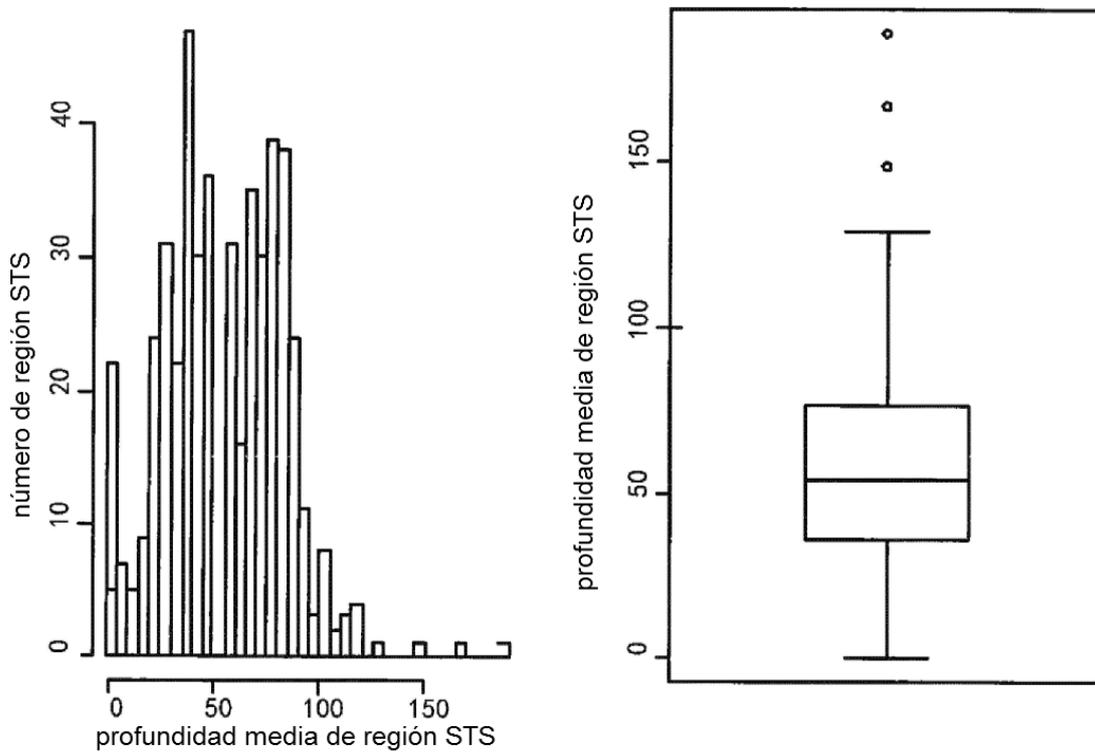


Fig.21

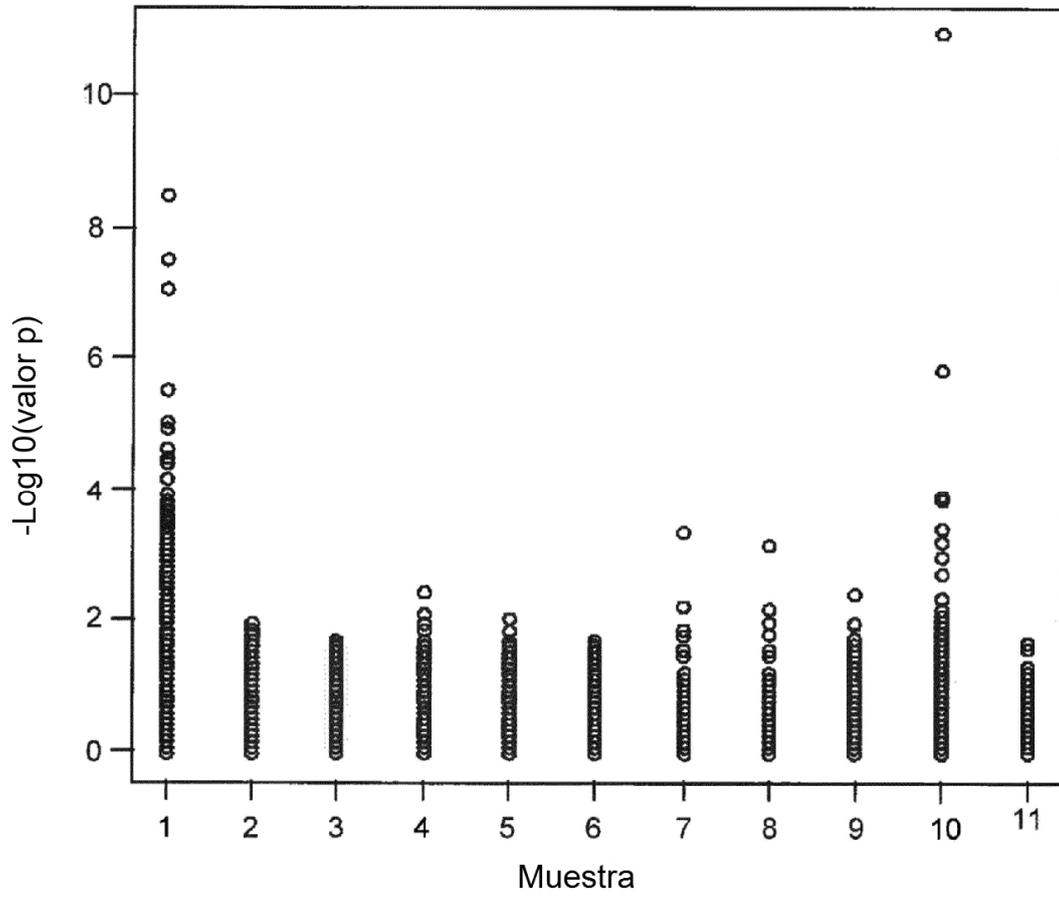


Fig.22

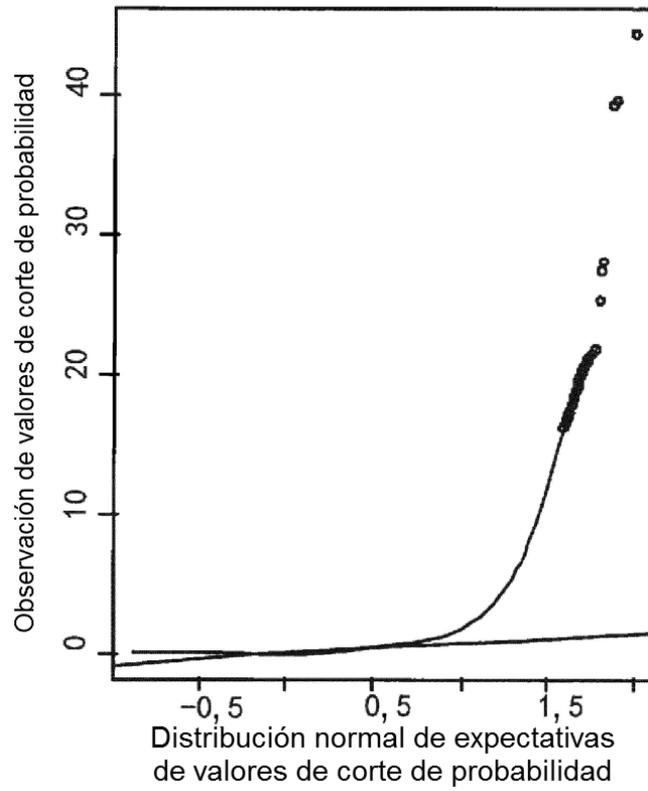


Fig.23

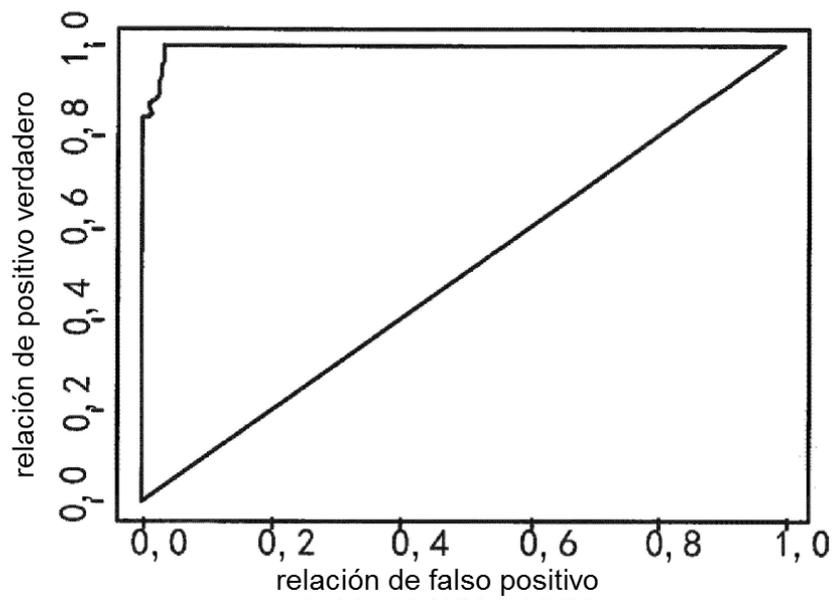


Fig.24

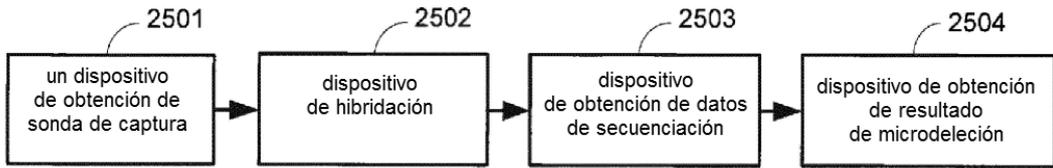


Fig.25

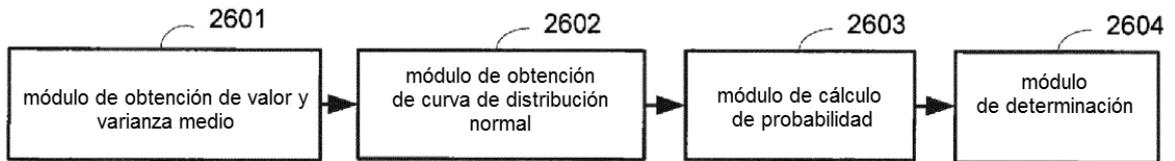


Fig.26

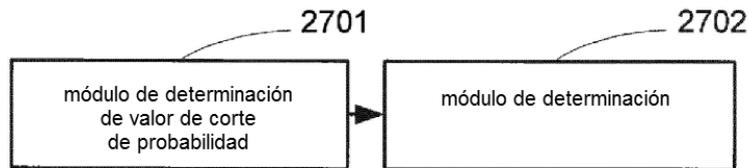


Fig.27

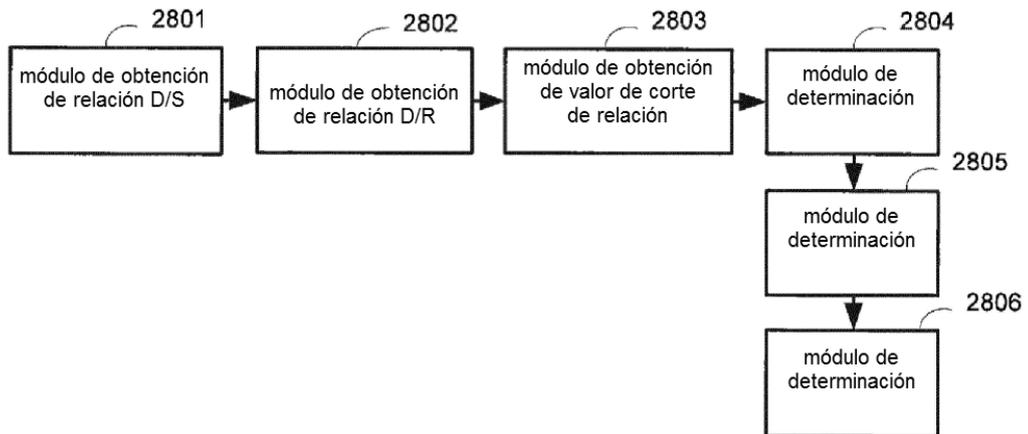


Fig.28

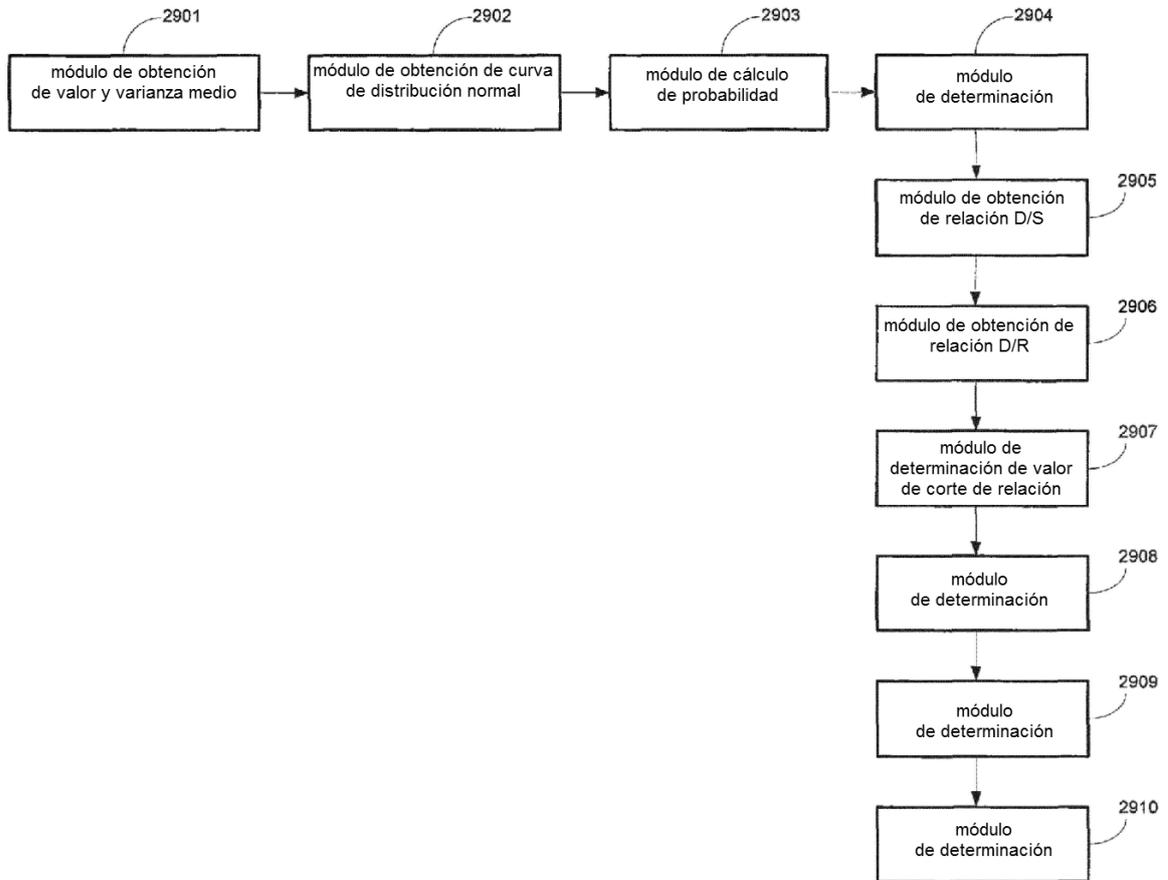


Fig.29