

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 970 286**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12Q 1/6851 (2008.01)

C12Q 1/6869 (2008.01)

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.08.2018 PCT/US2018/045434**

87 Fecha y número de publicación internacional: **11.07.2019 WO19135790**

96 Fecha de presentación y número de la solicitud europea: **06.08.2018 E 18898428 (0)**

97 Fecha y número de publicación de la concesión europea: **22.11.2023 EP 3735470**

54 Título: **Plantillas de control de calidad para garantizar la validez de ensayos basados en secuenciación**

30 Prioridad:

05.01.2018 US 201862614236 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.05.2024

73 Titular/es:

**BILLIONTOONE, INC. (100.0%)
1035 O'Brien Drive
Menlo Park, CA 94025, US**

72 Inventor/es:

**TSAO, DAVID;
SILAS, SUKRIT y
ATAY, OGUZHAN**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 970 286 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

Plantillas de control de calidad para garantizar la validez de ensayos basados en secuenciación

Referencia cruzada a solicitudes relacionadas

Esta solicitud reivindica el beneficio de la solicitud provisional de EE. UU. de número de serie 62/614.236, presentada el 5 de enero de 2018.

Campo técnico

Esta descripción se refiere en general al campo de la secuenciación genética.

Antecedentes

La secuenciación de alto rendimiento (por ejemplo, la secuenciación de próxima generación (NGS)) se utiliza cada vez más para ensayos de diagnóstico, tanto para la secuenciación del genoma completo como del exoma, y en aplicaciones más especializadas como pruebas prenatales no invasivas (NIPT), biopsias líquidas y ensayos similares que detectan polimorfismos. En la secuenciación de alto rendimiento (por ejemplo, NGS), la contaminación cruzada es un problema importante para las aplicaciones clínicas, porque se puede procesar una pluralidad de muestras (por ejemplo, hasta 384 muestras, etc.) en la misma secuenciación. En particular, en los ensayos en los que las mutaciones o polimorfismos son infrecuentes, de modo que sus frecuencias alélicas representan sólo un pequeño porcentaje del total, la contaminación cruzada de otras muestras puede dar lugar a falsos positivos. Esto es particularmente cierto para la NIPT y las biopsias líquidas, donde una diferencia cuantitativa de menos de un pequeño porcentaje es la diferencia entre un resultado positivo y uno negativo.

Las prácticas estándar de preparación de bibliotecas para la secuenciación de alto rendimiento pueden requerir la amplificación de una muestra de ADN para la entrada inicial. Estas etapas de amplificación pueden exacerbar el efecto de la contaminación cruzada, ya que cualquier amplificación de un alelo mutante en el laboratorio puede contaminar las muestras y los experimentos posteriores, lo que comúnmente se conoce como contaminación por arrastre en la PCR. Para evitar este problema, algunos ensayos de diagnóstico estándar, como la qPCR, utilizan sistemas de prevención de transferencia de dUTP/UNG, en los que dUTP se sustituye por dTTP en la PCR, y los amplicones que contienen uracilo se degradan después del ensayo mediante el tratamiento con la enzima uracilo ADN glicosilasa. Sin embargo, no existe una solución similar para los ensayos basados en secuenciación de alto rendimiento (p. ej., los ensayos basados en NGS, etc.) a pesar de la necesidad aún más crítica debido a la mayor sensibilidad de la secuenciación de alto rendimiento (p. ej., NGS) y los pequeños cambios cuantitativos que miden los ensayos basados en la secuenciación de alto rendimiento.

Si bien eliminar por completo la contaminación cruzada en la secuenciación de alto rendimiento es difícil debido a la química asociada, poder rastrearla sería igualmente valioso. En los ejemplos, se puede añadir una secuencia diferente e identificable a cada muestra para rastrear su contaminación en otros pocillos. Sin embargo, tales ejemplos en los que cada usuario, cada experimento y cada muestra tienen una biblioteca diferente de secuencias pueden ser engorrosos y pueden requerir el mantenimiento de una gran pluralidad de bibliotecas distintas (por ejemplo, 384 bibliotecas distintas; varias bibliotecas distintas que corresponden al número de muestras que se procesan en la misma ejecución de secuenciación, etc.) cuando se utilizan para rastrear la contaminación cruzada de ensayos basados en secuenciación de alto rendimiento multiplexados (por ejemplo, los ensayos basados en NGS, etc.). Además, dichos ejemplos no podrían rastrear la transferencia de PCR de los experimentos anteriores, ya que se usarían las mismas bibliotecas en diferentes experimentos. Además, debido a la dificultad de mantener una gran pluralidad de bibliotecas distintas (por ejemplo, 384 bibliotecas distintas, etc.), las propias secuencias identificadoras pueden sufrir contaminación cruzada. Como tal, existe la necesidad de realizaciones nuevas y útiles de un método y/o sistema, tales como para rastrear la contaminación cruzada y al mismo tiempo superar estas limitaciones.

Tourlousse et al. (Nucleic Acids Research, 45 (4): e23, 2017), los documentos WO 2017/165864 y WO 2012/058316 se refieren a patrones de adición sintéticos para la secuenciación de alto rendimiento de amplicones de genes.

La invención se expone en las reivindicaciones adjuntas. Por consiguiente, la invención proporciona un método para analizar una muestra materna obtenida de una mujer embarazada para facilitar la determinación prenatal de la presencia de un trastorno genético, y el método comprende: añadir, a la muestra materna, un conjunto de moléculas plantilla de control de calidad (QCT) asociadas al trastorno genético, y el conjunto de moléculas QCT comprende: regiones asociadas a la diana con similitud de secuencia respecto de una región de secuencia diana de las moléculas diana endógenas, y regiones de variación con disimilitud de secuencia respecto de una región de secuencia de las moléculas diana endógenas, en donde las regiones de variación comprenden una región identificadora molecular incrustada que tiene una secuencia de nucleótidos única; generar una mezcla coamplificada basándose en la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que comprenden la región de secuencia diana de las moléculas diana endógenas; secuenciar la mezcla coamplificada; determinar computacionalmente un número único del conjunto de moléculas QCT, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT que se detectan a partir de las lecturas de secuencias de moléculas QCT identificadas durante la secuenciación, en donde las lecturas de secuencias de moléculas QCT corresponden al conjunto de moléculas QCT;

calcular la profundidad de la secuenciación de QCT promedio basándose en dividir el número de lecturas de secuencias de moléculas QCT por el número único de moléculas QCT; determinar el recuento absoluto de las moléculas diana endógenas basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio; determinar el recuento absoluto de moléculas de referencia endógenas basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio; y facilitar la determinación prenatal de la presencia del trastorno genético basándose en una comparación entre el recuento absoluto de las secuencias diana endógenas y el recuento absoluto de las secuencias de referencia endógenas.

Un aspecto de la invención proporciona un método para identificar la contaminación asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación de alto rendimiento, y el método comprende: generar un conjunto de moléculas plantilla de control de calidad (QCT), y cada molécula QCT comprende: una región asociada a una diana con similitud de secuencia respecto de una región de secuencia diana de una diana biológica, y una región de variación con disimilitud de secuencia respecto de una región de secuencia de la diana biológica en donde la región de variación comprende una región identificadora molecular incrustada que tiene una secuencia de nucleótidos única; y determinar computacionalmente un conjunto de grupos de lecturas de secuencias QCT basándose en las regiones de variación del conjunto de moléculas QCT, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT que se detectan a partir de las lecturas de secuencias de moléculas QCT identificadas durante la secuenciación, en donde las lecturas de secuencias de moléculas QCT corresponden al conjunto de moléculas QCT, en donde el conjunto de grupos de lecturas de secuencias QCT comprende las lecturas de secuencias de moléculas QCT derivadas de la secuenciación de alto rendimiento que corresponden a un conjunto de mezclas de QCT generadas basándose en el conjunto de moléculas QCT y un conjunto de muestras que comprenden la diana biológica, en donde la preparación de una biblioteca de secuenciación comprende la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que comprenden la diana biológica basándose en la similitud de secuencia de la región asociada a la diana y la región de secuencia diana de la diana biológica objetivo en donde el conjunto de moléculas QCT se añade a una concentración de menos de 0,00001 nanogramos de moléculas QCT amplificables por muestra a más de una muestra en la biblioteca de secuenciación, y en donde el conjunto de grupos de lecturas de secuencias QCT permiten la determinación del recuento de moléculas QCT que estaban en la mezcla antes de la coamplificación; y basándose en el conjunto de grupos de lecturas de secuencias QCT, determinar un parámetro de contaminación que describe la contaminación asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación de alto rendimiento.

Un aspecto de la invención proporciona un método para la caracterización asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, y el método comprende: generar un conjunto de moléculas plantilla de control de calidad (QCT), y cada molécula QCT comprende: una región asociada a una diana que coincide con una región de secuencia diana de una molécula diana; y una región de variación que no coincide con una región correspondiente de la molécula diana, y la región de variación comprende un identificador molecular incrustado que tiene una secuencia de nucleótidos única; generar una mezcla que comprende el conjunto de moléculas QCT y moléculas de ácido nucleico de una muestra; coamplificar la mezcla; secuenciar la mezcla coamplificada para generar datos de secuenciación; determinar computacionalmente un conjunto de grupos de lecturas de secuencias QCT, mediante el uso de los datos de secuenciación, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT identificadas en los datos de secuenciación, determinar computacionalmente el recuento de moléculas QCT que estaban en la mezcla antes de la coamplificación basándose en los grupos de lecturas de secuencias QCT; y basándose en el recuento de moléculas QCT, determinar la abundancia absoluta de la molécula diana asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación. Las características opcionales de la invención se exponen en las reivindicaciones dependientes adjuntas.

Breve descripción de las figuras

Las FIGURAS 1A-1D incluyen representaciones en diagramas de flujo de variaciones de las realizaciones de un método;

La FIGURA 2 incluye una representación en diagrama de flujo de una variación de una realización de un método;

La FIGURA 3 incluye una representación en diagrama de flujo de una variación de una realización de un método;

Las FIGURAS 4A-4D incluyen representaciones gráficas de los resultados de partes de validación de una variación de una realización de un método, en particular en relación con la contaminación cruzada y la asignación incorrecta de índices;

Las FIGURAS 5A-5B incluyen un ejemplo específico de resultados de experimentos que validan el uso de moléculas QCT para el recuento molecular;

La FIGURA 6 incluye un ejemplo específico de resultados asociados con aspectos de calidad asociados con la gestión de técnicos y/o la gestión de laboratorio;

Las FIGURAS 7A-7C incluyen representaciones gráficas de los resultados de validar partes de una variación de una realización de un método, en particular en relación con la cuantificación de moléculas QCT;

Las FIGURAS 8A-8B incluyen representaciones gráficas de los resultados de validar partes de una variación de una realización de un método, en particular en relación con la cuantificación de dianas biológicas;

La FIGURA 9 incluye un ejemplo específico del uso de moléculas QCT para medir equivalentes genómicos ensayables;

- 5 La FIGURA 10 incluye un ejemplo específico de determinación de los parámetros de contaminación;

La FIGURA 11 incluye un ejemplo específico del uso de moléculas QCT con una pluralidad de muestras para determinar errores de procesamiento de muestras;

La FIGURA 12 incluye un ejemplo específico del uso de moléculas QCT en diferentes etapas;

Las FIGURAS 13A-13B incluyen ejemplos específicos de caracterización asociada con saltos de índices;

- 10 La FIGURA 14 incluye un ejemplo específico para facilitar la medición de los niveles de contaminación reales asociados con el uso de cebadores de índice dual único;

Las FIGURAS 15A-15D incluyen ejemplos específicos asociados a facilitar el diagnóstico de un trastorno de un solo gen;

- 15 Las FIGURAS 16A-16B incluyen ejemplos específicos asociados a facilitar el diagnóstico de una anomalía cromosómica.

Descripción de las realizaciones

La siguiente descripción de las realizaciones está destinada a permitir que cualquier experto en la técnica las realice y las utilice.

1. Descripción general.

- 20 Como se muestra en las FIGURAS 1A-1D y las FIGURAS 2-3, las realizaciones de un método 100 (por ejemplo, para la caracterización asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, etc.) pueden incluir la generación de un conjunto de moléculas plantilla de control de calidad (QCT) (por ejemplo, cada molécula QCT que incluye una región asociada a la diana, una región de variación, etc.) S110; determinar (por ejemplo, computacionalmente, etc.) un conjunto de grupos de lecturas de secuencias QCT (por ejemplo, que corresponden al conjunto de moléculas QCT, etc.) basándose en el conjunto de moléculas QCT (por ejemplo, basándose en las regiones de variación del conjunto de moléculas QCT, etc.) S120; y/o basándose en el conjunto de grupos de lecturas de secuencias QCT, determinar un parámetro relacionado con la secuenciación (por ejemplo, un parámetro de contaminación, un parámetro de recuento de moléculas, etc.) asociado con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación S130.

- 30 Adicional o alternativamente, las realizaciones del método 100 pueden incluir preparar una o más bibliotecas de secuencias S112; secuenciar con una o más bibliotecas de secuencias S114; facilitar (por ejemplo, ayudar, determinar, proporcionar, etc.) uno o más diagnósticos de una o más afecciones (por ejemplo, trastornos genéticos, etc.) S140 (por ejemplo, basándose en uno o más parámetros relacionados con la secuenciación, etc.); facilitar (por ejemplo, ayudar, determinar, proporcionar, administrar, etc.) el tratamiento para una o más afecciones, tal como basándose en los parámetros relacionados con la secuenciación, los diagnósticos y/u otros componentes adecuados S150; y/o cualquier otro proceso adecuado.

- 40 En un ejemplo específico, el método 100 (por ejemplo, para facilitar el diagnóstico prenatal de un trastorno genético a partir de una muestra materna asociada con una mujer embarazada, etc.) puede incluir: añadir, a la muestra materna, un conjunto de moléculas QCT asociadas con el trastorno genético, y el conjunto de moléculas QCT incluye: regiones asociadas a la diana con similitud de secuencia respecto de una región de secuencia diana de las moléculas diana endógenas (por ejemplo, asociadas con el trastorno genético, etc.), y regiones de variación (por ejemplo, que incluyen regiones de un identificador molecular integrado (EMI) que incluyen un conjunto de bases "N" variables, donde cada base "N" se selecciona de cualquiera de una base "A", una base "G", una base "T" y una base "C", etc.) con disimilitud de secuencia con respecto a una región de secuencia de las moléculas diana endógenas; generar una mezcla coamplificada basándose en la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico (por ejemplo, ácidos nucleicos; fragmentos de ácidos nucleicos, etc.) que incluye la región de secuencia diana; secuenciar la mezcla coamplificada; determinar computacionalmente un número único del conjunto de moléculas QCT, basándose en el número de regiones de variación que son distintas y se detectan a partir de las lecturas de secuencias de moléculas QCT de la secuenciación, donde las lecturas de secuencias de moléculas QCT corresponden al conjunto de moléculas QCT; calcular la profundidad de la secuenciación de QCT promedio basándose en dividir el número de lecturas de secuencias de moléculas QCT por el número único de moléculas QCT; determinar el recuento absoluto de las moléculas diana endógenas basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio; determinar el recuento absoluto de moléculas de referencia endógenas basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la

profundidad de la secuenciación de QCT promedio; y facilitar el diagnóstico prenatal del trastorno genético basándose en una comparación entre el recuento absoluto de las secuencias diana endógenas y el recuento absoluto de las secuencias de referencia endógenas.

En un ejemplo específico, el método 100 (por ejemplo, para la caracterización, tal como la identificación de contaminación, asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, etc.) puede incluir generar un conjunto de moléculas QCT, y cada molécula QCT incluye una región de variación (por ejemplo, incluye una o más regiones EMI, etc.) y/o una región asociada a la diana (por ejemplo, con similitud de secuencia respecto de una región de secuencia diana de una diana biológica, etc.); determinar computacionalmente un conjunto de grupos de lecturas de secuencias QCT basándose en las regiones de variación del conjunto de moléculas QCT, tal como cuando el conjunto de grupos de lecturas de secuencias QCT incluye las lecturas de secuencias de moléculas QCT derivadas de la secuenciación que corresponde a una mezcla de QCT generada basándose en el conjunto de moléculas QCT y una muestra que incluye la diana biológica (p. ej., una muestra que incluye moléculas diana endógenas que corresponden a la diana biológica, etc.), tal como cuando la preparación de una biblioteca de secuenciación incluye la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que incluyen la diana biológica (por ejemplo, basándose en la similitud de secuencia de la región asociada a la diana y la región de secuencia diana de la diana biológica, etc.); y basándose en el conjunto de grupos de lecturas de secuencias QCT, determinar un parámetro relacionado con la secuenciación asociado con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación (por ejemplo, determinar un parámetro de contaminación que describe la contaminación asociada con al menos una de las preparaciones de la biblioteca de secuenciación y la secuenciación de alto rendimiento, etc.).

En un ejemplo específico, como se muestra en la FIGURA 2, el método 100 (por ejemplo, para garantizar la validez de ensayos basados en secuenciación basándose en moléculas QCT, etc.) puede incluir: generar una biblioteca de QCT o mezclas de bibliotecas de QCT de moléculas QCT (p. ej., ADN QCT que incluye regiones asociadas a las dianas con una alta similitud con un gen de interés para permitir la coamplificación mediante el uso de los cebadores de PCR representados como flechas negras en la FIGURA 2; ADN QCT que incluye regiones de variación con diferencias de secuencia en comparación con el gen de interés, donde la región de variación puede incluir una región EMI que incluye bases "N" que pueden adoptar aleatoriamente bases "A", "C", "T" o "G", donde se pueden generar hasta 4⁴ secuencias EMI únicas con "NNNN", donde se puede encontrar la probabilidad de que dos moléculas QCT tengan el mismo EMI mediante el uso de la solución al problema del cumpleaños para calcular la probabilidad de colisión de hash, donde se pueden mostrar las subsecciones de las diferencias de las secuencias QCT y HBB en la FIGURA 2; ADN QCT que incluye una región identificadora de QCT (ID de QCT) para distinguir las bibliotecas de QCT y las secuencias de los genes de interés en las lecturas de secuenciación; etc.; preparar una biblioteca de secuenciación basándose en las moléculas QCT y una o más muestras que incluyen las dianas biológicas (por ejemplo, HBB, como se muestra en la FIGURA 2, etc.), tal como añadiendo la biblioteca de QCT al ADN humano; aplicar enfoques computacionales para agrupar las lecturas de secuencias de moléculas QCT (por ejemplo, basándose en la similitud de las secuencias EMI; donde el número de grupos de EMI corresponde al número absoluto de moléculas QCT añadidas a la muestra, etc.) y asignar los grupos a diferentes identificadores de muestras (por ejemplo, correspondientes a diferentes muestras; correspondientes a diferentes compartimentos de muestras utilizados en la secuenciación, etc.); y utilizar dichos datos para evaluar las métricas de control de calidad como contaminación cruzada, asignación incorrecta de índices, errores del usuario (por ejemplo, al ejecutar el ensayo), incumplimiento de los parámetros del ensayo (por ejemplo, una cantidad demasiado pequeña de ADN de entrada, equivalentes genómicos accesibles en una muestra, etc.), y/o para cuantificar la cantidad de la diana biológica de entrada a la que se puede acceder mediante un ensayo.

Las realizaciones del método 100 y/o del sistema 200 pueden funcionar para cuantificar con precisión la abundancia de las dianas biológicas, rastrear y/o cuantificar con precisión el grado de contaminación (por ejemplo, la contaminación cruzada entre diferentes muestras, diferentes experimentos; los verdaderos niveles de contaminación asociados con el uso de cebadores de índice dual único, etc.), identificar errores del usuario en la ejecución de ensayos basados en secuenciación, monitorizar la asignación incorrecta de índices de secuenciación, determinar el incumplimiento de los parámetros del ensayo, identificar y/o facilitar la eliminación de cebadores contaminantes y/o con saltos de índices, y/o mejorar cualquier aspecto adecuado asociado a la preparación y/o secuenciación de la biblioteca de secuenciación, tal como para mejorar el diagnóstico y/o la terapia.

En un ejemplo, en la presente memoria se describe que se puede añadir un solo reactivo (por ejemplo, que incluye un conjunto de moléculas QCT, etc.) a un conjunto de muestras para rastrear la contaminación cruzada y otros errores del usuario añadiendo un solo reactivo a todas las muestras. Por ejemplo, la adición de moléculas QCT, cuando va acompañada de un proceso de análisis matemático y computacional personalizado basándose en las soluciones al problema del cumpleaños para la colisión de hash, puede rastrear la contaminación cruzada entre diferentes usuarios, diferentes experimentos y diferentes muestras al mismo tiempo. En los ejemplos, se puede añadir una única biblioteca de QCT a todas las muestras (por ejemplo, asociada a la secuenciación de alto rendimiento, etc.), para mejorar la facilidad de uso y la comodidad. En un ejemplo específico, se pueden añadir distintas bibliotecas de QCT (por ejemplo, que corresponden a diferentes regiones identificadoras de QCT, como ID de QCT, etc.) en diferentes etapas de la preparación de la muestra para rastrear cualquier error del usuario o pérdida de la muestra de entrada. En los ejemplos, se puede aplicar un enfoque automático de huella digital al dispensar, donde cada muestra puede identificarse

mediante las moléculas QCT (por ejemplo, basándose en las regiones de variación de las moléculas QCT; las regiones identificadoras de QCT de las moléculas QCT, etc.). En los ejemplos, se puede medir la contaminación debida al arrastre en la PCR, cuya contaminación puede ser un problema en los entornos clínicos y/u otros contextos. En ejemplos específicos, se pueden usar moléculas QCT para asignar una huella molecular a cada tubo de PCR, y el arrastre en la PCR se puede detectar y cuantificar manteniendo una base de datos de todas las regiones de variación (por ejemplo, secuencias EMI de las regiones EMI, etc.) asociadas a cada tubo de PCR, realizado en una ubicación o sala de laboratorio determinada. El arrastre en la PCR en los ensayos posteriores se puede identificar mediante la búsqueda computacional de la huella de la región de variación (por ejemplo, las similitudes de las huellas de EMI, etc.) en la base de datos histórica.

Las realizaciones pueden usarse adicional o alternativamente para garantizar la calidad frente a un problema importante para la secuenciación de alto rendimiento (por ejemplo, NGS, etc.), concretamente, el "cambio de índice" o la asignación incorrecta de índices. En los ejemplos, incluso en ausencia de contaminación cruzada, las lecturas o señales de secuenciación (por ejemplo, hasta un 5-10%, etc.) de una muestra se pueden asignar incorrectamente a otra muestra cuando se multiplexan en la misma celda de flujo. En los ejemplos, se puede realizar un método conveniente de huella digital al dispensar para cuantificar con precisión el grado de asignación incorrecta en cada muestra. En un ejemplo específico, como se muestra en las FIGURAS 4A-4D, el efecto acumulativo de la contaminación cruzada en pocillos cercanos y la asignación incorrecta de índices se puede detectar con una sensibilidad >90%. En un ejemplo específico, como se muestra en la FIGURA 4A, se puede preparar una biblioteca Illumina Truseq HT utilizando las 96 combinaciones de índices i7 e i5, donde cada pocillo corresponde a un experimento de secuenciación de amplicones de HBB con 400, 200, 100 o 0 moléculas QCT añadidas a cada pocillo; en cada pocillo se muestra la fracción de lecturas de QCT identificadas como contaminación cruzada, donde, en el experimento, la contaminación cruzada y la asignación incorrecta de índices oscilaron entre <1% y 13%; debido a 0 moléculas QCT en las columnas D710-D712, la fracción en estos pocillos indica la sensibilidad con la que las variaciones de las realizaciones del método 100 pueden detectar la contaminación cruzada; y donde la FIGURA 4B ilustra el número y la fuente de las lecturas contaminantes encontradas en D702/D504, indicadas por X (izquierda); y el número y destino de las lecturas contaminantes que se originan en D702/D504, indicado por O (derecha); y donde la FIGURA 4C ilustra un análisis análogo al de la FIGURA 4B, pero para el pocillo D707/D504; y donde la FIGURA 4D ilustra la fuente de contaminación para los pocillos D710/D504 y D711/D504, y donde no se halló que se originaran lecturas contaminantes en estos pocillos, lo que coincide con la ausencia de moléculas QCT añadidas a estos pocillos.

Las realizaciones pueden permitir adicional o alternativamente un recuento molecular preciso de una diana biológica (por ejemplo, basándose en el uso de un conjunto de regiones de variación de un conjunto de moléculas QCT, etc.), tal como cuando la profundidad de lectura es suficiente (por ejemplo, profundidad de lectura mayor de veinte por molécula QCT distinta, etc.), lo que puede ayudar a obtener una cuantificación precisa de la diana en los enfoques que utilizan la adición de secuencias complejas. Las realizaciones pueden cuantificar la diana biológica accesible que se analiza cuando la profundidad de lectura es suficientemente alta. En un ejemplo, relacionado con la detección de mutaciones para el uso en pruebas prenatales no invasivas mediante secuenciación de amplicones, dicha profundidad de lectura alta para una secuencia compleja se obtiene cuando se añaden <400 moléculas QCT a cada muestra, como cuando 96 de dichas muestras se multiplexan en un análisis de tipo MiSeq (por ejemplo, como se muestra en la FIGURA 7C). Sin embargo, se puede añadir cualquier número adecuado de moléculas QCT a una o más muestras para facilitar el recuento molecular y/u otra funcionalidad adecuada. En los ejemplos específicos, como se muestra en las FIGURAS 5A-5B, el número de lecturas de secuencias y el número de moléculas (por ejemplo, determinado basándose en las lecturas de secuencias y el procesamiento asociado con las moléculas QCT, etc.) se pueden correlacionar, y la relación entre el número de moléculas y las lecturas puede diferir entre 2 y 3 veces, lo que indica las mejoras asociadas al uso de las moléculas QCT para determinar el número de moléculas en una muestra determinada (por ejemplo, mejoras en la fiabilidad con respecto al uso del número de lecturas por sí solo, etc.).

En los ejemplos (por ejemplo, de cuantificación del recuento absoluto de moléculas para una o más dianas biológicas, etc.), el método 100 y/o el sistema 200 se pueden usar a) para determinar parámetros para el uso en algoritmos para determinar el resultado diagnóstico de ensayos, b) para rastrear la pérdida de ADN de entrada en diferentes etapas de los experimentos o ensayos, c) para devolver un resultado ininterpretable cuando el número de moléculas diana es demasiado bajo (por ejemplo, para determinar cuándo un ensayo no es fiable, etc.), d) diseñar ensayos para detectar la variación del número de copias en un locus particular o entre loci, y/o e) ayudar en la toma de decisiones terapéuticas y clínicas basándose en los resultados de los ensayos de diagnóstico.

Las realizaciones pueden evaluar y/o mejorar adicional o alternativamente varios aspectos de calidad asociados con la gestión de técnicos y/o la gestión de laboratorios (por ejemplo, la gestión de laboratorios clínicos, etc.). Como se muestra en la FIGURA 6, en un ejemplo específico, el método 100 y/o el sistema 200 se pueden usar para identificar el procesamiento de muestras problemáticas por parte de diferentes técnicos o laboratorios, donde las Muestras A01 a A06 frente a las Muestras B31 a B35 se procesaron en dos laboratorios diferentes con diferentes prácticas de separación pre/post-PCR; antes del procesamiento se añadió a cada muestra el mismo volumen de moléculas QCT del mismo kit, que correspondía aproximadamente a ~200 moléculas; "num_seqs" indica el número de grupos de EMI distintos que se identificaron para cada muestra; "contam_frac" indica la fracción total de lecturas contaminantes que se identificaron en cada muestra; "collision_frac" identifica el grado en que se encuentran dos grupos de EMI válidos en dos muestras diferentes; "contam_collision_frac" combina las dos métricas anteriores; "ident_frac" es el número de

lecturas que se asignan a EMI válidos dividido por el número total de lecturas para esa muestra en particular; "reads_per_qctmol" indica la profundidad de lectura promedio para los EMI; se utilizó un umbral de filtro para estas métricas derivadas para identificar las muestras que pasan o no pasan el control de calidad (QC); donde solo 1 de cada 6 muestras pasó las métricas de control de calidad para el Laboratorio A, mientras que 5 de cada 5 muestras pasaron las métricas de control de calidad en el Laboratorio B; y dónde se pueden utilizar estos resultados para cambiar la forma en que se puede realizar el procesamiento de las muestras y la separación pre/post-PCR (por ejemplo, en qué lugar del Laboratorio A, en la siguiente ejecución con mejoras en el procesamiento de muestras, las muestras pasaron las mismas métricas de control de calidad, etc.). Como se muestra en las FIGURAS 7A-7C, en los ejemplos específicos, al incluir múltiples especies de QCT que se dispensan de la misma mezcla, se puede medir el error aleatorio en las pipetas mediante la correlación de los recuentos absolutos de moléculas QCT (por ejemplo, como se muestra en la FIGURA 7C), y/o el error sistemático de la pipeta y/o cuantificación puede ser rastreable adicional o alternativamente (por ejemplo, como se muestra en la FIGURA 7A, tal como basándose en una comparación del panel central frente a los paneles izquierdo y derecho). En los ejemplos específicos, como se muestra en las FIGURAS 7A-7C, se puede determinar la cuantificación absoluta de las moléculas QCT añadidas. En un ejemplo específico, como se muestra en la FIGURA 7A, las bibliotecas de QCT1, QCT2 y QCT3 (por ejemplo, que corresponden a diferentes conjuntos de moléculas QCT, etc.) se pueden preparar, combinar y añadir a las reacciones de PCR a 100, 200, o 400 moléculas por biblioteca de QCT; Los EMI para cada biblioteca de QCT se pueden agrupar agregando las lecturas de secuencias EMI con como máximo 2 cambios de bases; las barras de error pueden representar la media +/- desviación estándar para 24 réplicas; y las líneas del gráfico pueden representar un ajuste mediante regresión lineal con un sombreado correspondiente al intervalo de confianza de la media del 95%. En un ejemplo específico, como se muestra en la FIGURA 7B, para determinar la solidez del recuento de QCT respecto de la profundidad de lectura, las lecturas de secuenciación se pueden reducir seleccionando aleatoriamente la mitad del total de lecturas; el número de grupos de EMI recuperados de las lecturas de secuenciación submuestreadas se puede representar gráficamente con respecto al conjunto de datos completo; el color de los puntos puede representar la profundidad de lectura submuestreada por grupo de EMI, donde la línea negra tiene una pendiente = 1, intersección = 0; El análisis de QCT es sólido cuando la profundidad de lectura por molécula QCT es superior a 20, lo que puede contribuir a la fiabilidad del recuento molecular; y cuando el número de grupos de QCT es 400, la profundidad de lectura muestreada es inferior a 20 por molécula. En un ejemplo específico, como se muestra en la FIGURA 7C, los recuentos de moléculas QCT pueden no estar correlacionados entre bibliotecas de QCT (por ejemplo, como se esperaba, etc.); donde se puede mostrar un diagrama de dispersión del número de grupos de QCT3 frente a los grupos de QCT1 para cada réplica de PCR de la FIGURA 7A al nivel de entrada de 100 moléculas QCT.

Las realizaciones pueden implementar adicional o alternativamente bibliotecas de QCT en diferentes etapas de preparación de bibliotecas de secuenciación (por ejemplo, las etapas de preparación de las muestras) y/o etapas de secuenciación para rastrear la pérdida de muestra. En un ejemplo específico, si se dispensa un primer conjunto de moléculas QCT (por ejemplo, moléculas QCT1; primeras moléculas QCT que incluyen una primera región identificadora de QCT compartida, etc.) en el punto de recogida de la muestra, y se dispensa una cantidad igual de un segundo conjunto de moléculas QCT (p. ej., moléculas QCT2; segundas moléculas QCT que incluyen una segunda región identificadora de QCT compartida, etc.) después de la purificación de la muestra; el rendimiento de la purificación se puede evaluar mediante comparaciones de los recuentos de moléculas para el primer conjunto de moléculas QCT y el segundo conjunto de moléculas QCT (por ejemplo, los recuentos de moléculas QCT1 frente a QCT2, etc.).

Las realizaciones pueden determinar adicional o alternativamente la porción de material biológico que es accesible mediante el ensayo, tal como mediante la cuantificación de las dianas biológicas basándose en el uso de las moléculas QCT, lo que puede mejorar al medir el material genómico total disponible y calcular la concentración esperada de la diana biológica, debido a que no todos los objetivos son accesibles mediante los ensayos. En un ejemplo específico, esto puede deberse al corte del ADN hasta una distribución de tamaños cortos, como en el caso del ADN libre circulante que se analiza en las aplicaciones de pruebas prenatales no invasivas (NIPT) para la determinación de enfermedades genéticas del feto y las aplicaciones de biopsias líquidas, donde se analiza el ADN tumoral circulante. En un ejemplo específico, en estas aplicaciones, dependiendo de la diana de interés, menos del 25% del ADN puede ser accesible, donde, como se muestra en las FIGURAS 8A-8B, la determinación de los equivalentes genómicos del ADN de entrada se puede determinar mediante el uso de las moléculas QCT, como por ejemplo cuando el ADN genómico humano se puede digerir mediante enzimas de restricción Alu o Hpy, que cortan el exterior y el interior del gen de interés, respectivamente; luego, las moléculas QCT pueden añadirse a 9 ng-36 ng de ADN digerido (correspondiente a 2.500-10.000 equivalentes genómicos), amplificarse mediante PCR y secuenciarse en un aparato MiSeq; los equivalentes genómicos (G.E.) de ADN humano en cada reacción de PCR se pueden medir mediante análisis asociados a las moléculas QCT al aplicar partes de las realizaciones del método 100; las reacciones de PCR se pueden realizar por duplicado, donde la FIGURA 8A ilustra una línea de ajuste lineal y el sombreado es el IC del val.pe-12779495% de la media, y donde la medición del ADN de entrada es coherente en todas las réplicas y en las series de dilución, pero es sistemáticamente mayor que la medición de Qubit mediante un factor constante; y donde, como se muestra en la FIGURA 8B, el ADN genómico humano puede cortarse hasta una distribución de tamaños con un pico de 100-150 pb; luego, las moléculas QCT se pueden añadir a 2,3 ng-36 ng de ADN cortado, y se pueden medir los equivalentes genómicos del ADN cortado, con un tamaño de amplicón de ~150 pb, y donde la FIGURA 8B ilustra la pendiente de la línea que indica la fracción de moléculas que pueden amplificarse debido a un corte aleatorio. En un ejemplo específico, como se muestra en la FIGURA 9, se pueden usar moléculas QCT para medir los equivalentes

genómicos ensayables, que pueden ser diferentes para cada ensayo, e incluso para el mismo ensayo con huellas diferentes; donde la región que rodea la misma mutación se amplificó a partir de ADN cortado para formar un producto de PCR de 150 pb frente a un producto de PCR de 72 pb (izquierda frente a derecha), y se utilizaron moléculas QCT para medir el número de moléculas amplificadas en ambos casos; donde 18 nanogramos (ng) de ADN genómico que corresponden a 5000 equivalentes genómicos de entrada se cortaron a una longitud promedio de ~170 pb (por ejemplo, la longitud promedio del ADN libre circulante) y se incluyeron en todos los casos (n=8 para 150 pb y n=4 para 72 pb); y de acuerdo con los modelos teóricos, la cantidad de moléculas que se pueden amplificar es significativamente menor que el ADN de entrada y puede tener una diferencia de hasta 2 veces entre diferentes huellas para la misma masa de ADN de entrada; y donde la FIGURA 9 puede indicar por qué otras mediciones del ADN de entrada (como la concentración) pueden no ser suficientes para diagnósticos moleculares precisos que requieren información molecular, como cuando una disminución de ~2x en el recuento molecular aumentaría el ruido de Poisson en aproximadamente un 40%, que puede ser la diferencia entre una precisión del 95% (2 sigma) y del 99% (3 sigma).

Las realizaciones del método 100 y/o del sistema 200 se pueden usar en asociación con una o más afecciones (por ejemplo, en asociación con la caracterización, el diagnóstico, el tratamiento y/o la realización de procesos relacionados con una o más afecciones, etc.), donde las afecciones pueden incluir y/o estar asociadas de otro modo con una o más de: pruebas prenatales no invasivas (NIPT) (p. ej., en relación con la detección genética para detectar la presencia de anomalías cromosómicas, incluida la aneuploidía, como la trisomía 21 o el síndrome de Down, la trisomía 18 o el síndrome de Edwards, la trisomía 13 o el síndrome de Patau, aneuploidías de los cromosomas sexuales como el síndrome de Turner, otras aneuploidías adecuadas; anomalías cromosómicas incluido el síndrome de DiGeorge; en relación con la detección genética de trastornos de un solo gen, etc.); otras pruebas prenatales; análisis de aneuploidía y/u otros análisis adecuados fuera de un contexto prenatal; trastornos genéticos (p. ej., trastornos de un solo gen, incluida la anemia de células falciformes; anomalías cromosómicas; trastornos asociados con la amplificación de genes; delección de genes; anomalías cromosómicas parciales; síndrome de delección 22q11.2 o síndrome de DiGeorge; síndrome de Charcot-Marie-Tooth, fibrosis quística, enfermedad de Huntington; distrofia muscular de Duchenne; hemofilia, talasemia, etc.), otras aplicaciones asociadas con anomalías cromosómicas (p. ej., ADN cromosómico adicional, ausente, irregular, etc.), cáncer (p. ej., a través de análisis asociados con cualquier oncogén adecuado, biomarcadores de cáncer y/u otras dianas asociadas al cáncer; mediante análisis asociados a biopsias líquidas), y/o cualesquiera otras afecciones adecuadas. En un ejemplo, el método 100 puede incluir determinar el recuento de moléculas diana (por ejemplo, correspondiente a una cantidad de moléculas diana en una muestra; basándose en el uso de moléculas QCT, etc.) para facilitar el diagnóstico asociado con al menos una de las pruebas prenatales no invasivas y biopsias líquidas. Las afecciones incluyen adicional o alternativamente: afecciones psiquiátricas y del comportamiento (por ejemplo, un trastorno psicológico; depresión; psicosis; etc.); afecciones relacionadas con la comunicación (por ejemplo, trastorno del lenguaje expresivo; tartamudez; trastorno fonológico; trastorno de autismo; afecciones de la voz; afecciones de la audición; afecciones oculares; etc.); afecciones relacionadas con el sueño (por ejemplo, insomnio, apnea del sueño, etc.); afecciones relacionadas con el sistema cardiovascular (por ejemplo, arteriopatía coronaria; presión arterial alta; etc.); afecciones relacionadas con el metabolismo (por ejemplo, diabetes, etc.), afecciones reumatóides (por ejemplo, artritis, etc.); afecciones relacionadas con el peso (por ejemplo, obesidad, etc.); afecciones relacionadas con el dolor; afecciones relacionadas con el sistema endocrino; enfermedad crónica; y/o cualquier otro tipo de afecciones adecuadas.

La secuenciación (por ejemplo, en relación con S112) asociada con una o más realizaciones del método 100 y/o sistema 200 incluye preferiblemente la secuenciación de alto rendimiento, que puede incluir y/o estar asociada con uno cualquiera o más de: NGS, tecnologías asociadas a NGS, secuenciación masiva de firmas en paralelo, secuenciación Polony, pirosecuenciación 454, secuenciación Illumina, secuenciación SOLiD, secuenciación de semiconductores Ion Torrent, secuenciación de nanobolas de ADN, secuenciación de una sola molécula con Heliscope, secuenciación en tiempo real de una sola molécula (SMRT), secuenciación de ADN con nanoporos, cualquier número de generación de tecnologías de secuenciación (por ejemplo, tecnologías de secuenciación de segunda generación, tecnologías de secuenciación de tercera generación, tecnologías de secuenciación de cuarta generación, etc.), secuenciación asociada a amplicones (por ejemplo, secuenciación de amplicones dirigida), secuenciación asociada a metagenomas, secuenciación por síntesis, secuenciación por corrientes de túnel, secuenciación por hibridación, secuenciación por espectrometría de masas, técnicas basadas en microscopía y/o cualquier tecnología adecuada relacionada con la secuenciación de alto rendimiento. Adicional o alternativamente, la secuenciación puede incluir cualquier tecnología de secuenciación adecuada (por ejemplo, secuenciación de Sanger, secuenciación capilar, etc.).

Uno o más ejemplos y/o partes de las realizaciones del método 100 y/o los procesos descritos en la presente memoria se pueden realizar de forma asincrónica (por ejemplo, secuencialmente), simultáneamente (por ejemplo, en paralelo; procesando simultáneamente muestras biológicas de manera múltiple y automatizada; procesando simultáneamente de manera computacional las lecturas de secuencia para mejorar la capacidad de procesamiento del sistema, etc.), en relación temporal con un evento desencadenante, y/o en cualquier otro orden adecuado en cualquier momento y frecuencia adecuados mediante y/o utilizando uno o más ejemplos de las realizaciones del sistema 200, los componentes y/o las entidades descritos en la presente memoria.

De manera adicional o alternativa, los datos descritos en la presente memoria (por ejemplo, grupos, parámetros relacionados con la secuenciación, identificadores, profundidades de lectura, lecturas de secuencias, determinaciones

de regiones de secuencias, diseños de moléculas QCT, diseños de cebadores, etc.) pueden asociarse con cualquier indicador temporal adecuado (por ejemplo, segundos, minutos, horas, días, semanas, períodos de tiempo, puntos de tiempo, marcas de tiempo, etc.), que incluyen uno o más: indicadores temporales que indican cuándo se recopilaron, determinaron, transmitieron, recibieron y/o procesaron de otro modo los datos; indicadores temporales que proporcionan contexto al contenido descrito por los datos, tales como indicadores temporales que indican la secuencia de las etapas de preparación y/o secuenciación de la biblioteca de secuenciación; cambios en los indicadores temporales (por ejemplo, datos a lo largo del tiempo; cambios en los datos; patrones de datos; tendencias de los datos; extrapolación de datos y/u otras predicciones, etc.); y/o cualquier otro indicador adecuado relacionado con el tiempo.

De manera adicional o alternativa, los parámetros, métricas, entradas, salidas y/u otros datos adecuados descritos en la presente memoria pueden asociarse con tipos de valores que incluyen uno o más de: puntuaciones, valores binarios, clasificaciones, niveles de confianza, identificadores (por ejemplo, identificadores de muestras, identificadores de moléculas QCT, etc.), valores a lo largo de un espectro y/o cualquier otro tipo de valores adecuado. Cualquier tipo adecuado de datos descritos en la presente memoria puede usarse como entradas, generarse como salidas y/o manipularse de cualquier manera adecuada para cualquier componente adecuado asociado a las realizaciones del método 100 y/o sistema 200.

Las realizaciones del sistema 200 pueden incluir adicional o alternativamente una red de manipulación de muestras configurada para generar moléculas (por ejemplo, moléculas QCT; bibliotecas de QCT, etc.), procesar muestras biológicas y/o realizar otros procesos adecuados; un sistema de secuenciación configurado para secuenciar material genético procesado a partir de mezclas generadas a partir de muestras biológicas y moléculas QCT; un sistema informático (por ejemplo, un sistema informático remoto; un sistema informático local, etc.) configurado para analizar las lecturas de secuencias, determinar grupos de lecturas de secuencias QCT, determinar parámetros relacionados con la secuenciación, facilitar diagnósticos, facilitar el tratamiento y/o realizar otros procesos adecuados (por ejemplo, procesos computacionales); y/o cualquier otro componente adecuado. Los componentes del sistema 200 pueden integrarse física y/o lógicamente de cualquier manera (por ejemplo, con cualquier distribución adecuada de funcionalidad entre los componentes, tal como en relación con partes de las realizaciones del método 100, etc.). Sin embargo, el método 100 y el sistema 200 se pueden configurar de cualquier manera adecuada.

2.1 Generación de moléculas QCT.

Las realizaciones del método 100 pueden incluir generar un conjunto de moléculas QCT S110, que puede funcionar para generar moléculas para ser usadas (por ejemplo, añadidas, procesadas, secuenciadas, etc.) en una o más etapas (por ejemplo, pasos, fases, períodos, períodos de tiempo, etc.) de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación (por ejemplo, secuenciación de alto rendimiento, etc.), tal como para facilitar el procesamiento computacional posterior (por ejemplo, la determinación de grupos de lecturas de secuencias QCT para facilitar la determinación de parámetros relacionados con las secuencias, etc.).

Las moléculas QCT incluyen preferiblemente regiones asociadas a la diana (por ejemplo, una o más regiones asociadas a la diana por molécula QCT, etc.). Como se muestra en la FIGURA 2, las regiones asociadas a la diana incluyen preferiblemente una similitud de secuencia (por ejemplo, similitud de secuencia completa; similitud de secuencia que satisface una condición umbral; similitud de secuencia de un número específico de bases, etc.) con una o más regiones de la secuencia diana de una o más moléculas diana (por ejemplo, moléculas diana endógenas; que corresponden a una o más dianas biológicas, etc.), pero pueden incluir adicional o alternativamente cualquier asociación adecuada con cualquier componente adecuado de una o más moléculas diana. Las regiones asociadas a la diana preferiblemente permiten la coamplificación de las moléculas QCT correspondientes (por ejemplo, que incluyen las regiones asociadas a la diana, etc.) y moléculas de ácido nucleico (por ejemplo, ácidos nucleicos, fragmentos de ácidos nucleicos, etc.) que incluyen la región de secuencia diana, lo que puede facilitar una precisión mejorada en el recuento molecular (por ejemplo, al determinar los parámetros de recuento de moléculas; teniendo en cuenta los sesgos de amplificación, etc.), pero puede permitir adicional o alternativamente cualquier proceso adecuado asociado con la preparación de una biblioteca de secuenciación, la secuenciación y/o porciones de las realizaciones del método 100. En un ejemplo, la preparación de una biblioteca de secuenciación (por ejemplo, realizar la preparación de una biblioteca de secuenciación S112) puede incluir la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que incluyen la diana biológica, basándose en la similitud de secuencia de la región asociada a la diana y la región de secuencia diana de la diana biológica, y donde la determinación del parámetro relacionado con la secuenciación puede incluir la determinación de un recuento de moléculas diana que describe el número de moléculas de la diana biológica asociada con la secuenciación, basándose en el conjunto de grupos de lecturas de secuencias QCT.

En las variaciones, las moléculas QCT pueden omitir las regiones asociadas a la diana. Por ejemplo, las moléculas QCT se pueden usar con componentes de muestras que incluyen dianas biológicas, sin asociación con las dianas (por ejemplo, sin tener una similitud predeterminada con las regiones de secuencias diana de las dianas biológicas) y/o sin la correspondiente coamplificación con los componentes de las muestras (por ejemplo, moléculas de ácido nucleico que incluyen las regiones de las secuencias diana, etc.). En los ejemplos, las moléculas QCT se pueden preprocesar para adaptarlas a la secuenciación, como cuando las moléculas QCT preprocesadas se pueden añadir a una muestra procesada adecuada para la secuenciación, para cosecuenciarla sin necesidad de coamplificación (p.ej., para mejorar la facilidad de uso). Las moléculas QCT que omiten las regiones asociadas a la diana se pueden usar preferiblemente

para facilitar la determinación de los parámetros de contaminación, pero se pueden usar adicional o alternativamente para facilitar cualquier determinación de parámetros relacionados con la secuenciación adecuada. En un ejemplo específico, el conjunto de moléculas QCT se puede adaptar para una secuenciación posterior (por ejemplo, secuenciación de alto rendimiento como NGS, etc.), donde la generación del conjunto de moléculas QCT puede incluir

5 amplificar un primer subconjunto de moléculas QCT (por ejemplo, que incluyen cada uno una primera región identificadora de QCT compartida, etc.) del conjunto de moléculas QCT; y amplificar un segundo subconjunto de moléculas QCT (por ejemplo, que incluyen cada una una segunda región identificadora de QCT compartida, etc.) del conjunto de moléculas QCT, donde las lecturas de secuenciación de las moléculas QCT derivan de la secuenciación correspondiente a: una mezcla de QCT generada basándose en el primer subconjunto de moléculas QCT y la muestra

10 que incluye la diana biológica (por ejemplo, que incluye las primeras moléculas diana que corresponden a la diana biológica, etc.), y una mezcla de QCT adicional generada basándose en el segundo subconjunto de moléculas QCT y una muestra adicional que incluye la diana biológica (por ejemplo, que incluye las segundas moléculas diana que corresponden a la diana biológica, etc.), donde la muestra y la muestra adicional corresponden respectivamente a un primer compartimento de muestras y un segundo compartimento de muestras de los compartimentos de muestras.

15 Sin embargo, las regiones asociadas a la diana y/o las moléculas QCT que omiten las regiones asociadas a la diana se pueden configurar de cualquier manera adecuada.

Las moléculas QCT incluyen preferiblemente una o más regiones de variación (por ejemplo, una o más regiones de variación por molécula QCT; regiones de variación adyacentes; regiones de variación separadas, etc.). Como se muestra en la FIGURA 2, una región de variación incluye preferiblemente una disimilitud de secuencia (por ejemplo, disimilitud de secuencia completa; disimilitud de un número específico de bases; disimilitud de secuencia parcial, etc.) con una o más regiones de secuencia (por ejemplo, distintas de una región de secuencia diana, etc.) de las moléculas diana. Una región de variación puede incluir adicional o alternativamente una o más regiones EMI. En una variación, una región EMI puede incluir un conjunto de bases "N" variables (por ejemplo, una o más bases "N" variables, etc.), donde cada base "N" se selecciona (por ejemplo, se selecciona aleatoriamente; se selecciona según distribuciones estadísticas y/o probabilidades predeterminadas, etc.) de cualquiera de una base "A", una base "G", una base "T" y una base "C". En una variación, una región EMI puede incluir una región sintetizada (por ejemplo, en una micromatriz; mediante el uso de síntesis basada en silicio, etc.) que incluye una o más bases específicas (por ejemplo, bases diseñadas y sintetizadas, etc.), tales como regiones sintetizadas diseñadas para facilitar la determinación del grupo de lecturas de secuencias QCT (por ejemplo, maximizando la distancia de Hamming por pares entre regiones EMI, etc.). En las variaciones, una molécula QCT puede incluir adicional o alternativamente una pluralidad de regiones EMI (por ejemplo, una región de variación que incluye una pluralidad de regiones EMI; regiones EMI adyacentes; regiones EMI separadas; regiones EMI que incluyen bases "N" variables; regiones EMI que incluyen regiones sintetizadas, etc.). Por ejemplo, cada región de variación del conjunto de moléculas QCT puede incluir una región identificadora molecular incrustada que incluye un conjunto de bases "N" variables, donde cada base "N" se selecciona de cualquiera de una base "A", una base "G", una base "T" y una base "C", donde cada molécula QCT del conjunto de moléculas QCT incluye además una región EMI adicional que incluye un conjunto adicional de bases "N" variables, donde la región EMI adicional está separada de la región EMI por una región de secuencia de la molécula QCT, tal como donde el conjunto de bases "N" variables y el conjunto adicional de bases "N" variables pueden incluir cada uno un número determinado (por ejemplo, predeterminado) de bases "N" (por ejemplo, más de tres bases "N", mayor que cualquier

20 número adecuado de bases "N", un número exacto de bases "N", etc.), y donde la determinación de un parámetro relacionado con la secuenciación (por ejemplo, un parámetro de contaminación) puede basarse en grupos de lecturas de secuencias QCT derivados basándose en las regiones EMI y las regiones EMI adicionales del conjunto de moléculas QCT (por ejemplo, basados en las lecturas de secuencias EMI distintas que corresponden a pares de una región EMI y una región EMI adicional, etc.). En una variación, una región de variación puede incluir adicional o

25

30

35

40

45

alternativamente un sintetizado.

En las variaciones, como se muestra en la FIGURA 2, una molécula QCT puede incluir una región identificadora de QCT que identifica la molécula QCT (y/u otras moléculas QCT adecuadas), tal como una región identificadora de QCT compartida (por ejemplo, una región de secuencia compartida, con disimilitud con respecto a una o más regiones de secuencia de las moléculas diana, etc.) que identifica las moléculas QCT que pertenecen a un conjunto de moléculas QCT (por ejemplo, donde diferentes regiones identificadoras de QCT son únicas para diferentes conjuntos de moléculas QCT, etc.). En un ejemplo, la región de variación de cada molécula QCT de un primer conjunto de moléculas QCT puede incluir una primera región EMI separada de una segunda región EMI por al menos una primera región identificadora de QCT, donde cada molécula QCT adicional de un segundo conjunto de moléculas QCT puede incluir una primera región EMI adicional separada de una segunda región EMI adicional por al menos una segunda región identificadora de QCT. En un ejemplo, la primera región EMI, la segunda región EMI, la primera región EMI adicional y la segunda región EMI adicional pueden incluir un conjunto de bases "N" variables, y donde cada base "N" se selecciona de cualquiera de una base "A", una base "G", una base "T" y una base "C", y donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT puede incluir la determinación del conjunto de grupos de lecturas de secuencias QCT basándose en la primera y la segunda regiones identificadoras de QCT, y en la primera región EMI, la segunda región EMI, la primera región EMI adicional y la segunda región EMI adicional. En un ejemplo, para cada molécula QCT del primer conjunto de moléculas QCT, la secuencia de la molécula QCT correspondiente se caracteriza por una similitud de secuencia completa con una primera plantilla de secuencia de la diana biológica excepto por la primera región identificadora de QCT, la primera región EMI, y la segunda región EMI; y donde, para cada molécula QCT adicional del segundo conjunto de moléculas QCT, la secuencia de la molécula

50

55

60

QCT adicional correspondiente se caracteriza por una similitud de secuencia completa con una segunda plantilla de secuencia excepto por la segunda región identificadora de QCT, la primera región EMI adicional y la segunda región EMI adicional. En un ejemplo específico, las secuencias de la molécula QCT pueden ser idénticas a la secuencia de la molécula diana (por ejemplo, una o más regiones de la secuencia de la molécula diana, etc.), excepto por dos secciones separadas de secuencias 5N interrumpidas por una región identificadora de QCT distinta previamente determinada (por ejemplo, una secuencia identificadora única, etc.). En un ejemplo específico, se pueden usar regiones identificadoras de QCT (por ejemplo, una secuencia de ID de QCT única, como se muestra en la FIGURA 2, etc.), para permitir el uso de múltiples bibliotecas de QCT que se pueden añadir en una etapa para el control interno o en etapas diferentes para el seguimiento de la pérdida de dianas biológicas de entrada u otros errores del usuario. Adicional o alternativamente, las regiones identificadoras de QCT se pueden configurar de cualquier manera adecuada. Sin embargo, las moléculas QCT pueden incluir cualquier combinación adecuada de cualquier tipo adecuado de regiones (por ejemplo, donde diferentes moléculas QCT incluyen el mismo o diferente tipo y/o número de regiones; con cualquier similitud y/o disimilitud de secuencia adecuada con las regiones de secuencia de las moléculas diana, etc.).

En las variaciones, el método 100 puede incluir adicional o alternativamente la generación de una o más bibliotecas de QCT (por ejemplo, cada biblioteca de QCT que incluye moléculas QCT, etc.), donde una biblioteca de QCT puede incluir múltiples conjuntos de moléculas QCT donde cada conjunto de moléculas QCT se puede identificar mediante una región identificadora de QCT diferente. En un ejemplo, la generación de una biblioteca de QCT puede incluir amplificar diferentes conjuntos de moléculas QCT (por ejemplo, para la preparación para la secuenciación, como cuando las moléculas QCT se amplifican antes de la adición a uno o más componentes de una muestra para generar una mezcla de QCT, etc.). En los ejemplos, la generación de una biblioteca de QCT puede incluir determinar una cantidad de moléculas QCT para incluirlas en la biblioteca de QCT. En un ejemplo específico, se pueden usar las soluciones al problema del cumpleaños para determinar el número máximo de moléculas QCT únicas que deben incluirse en cada muestra dada una diversidad particular de moléculas QCT, como donde, para 4^{10} secuencias, que se pueden generar mediante 10 bases N variables en una molécula QCT, se pueden usar hasta 1200 moléculas QCT con una probabilidad de $\sim 0,5$ de una única colisión de EMI válida ($\exp(-1200 \cdot 1199/2/4^{10}) \sim 0,5$), y donde con 200 moléculas QCT, la probabilidad de una única colisión válida es del $\sim 2\%$. En un ejemplo específico, la generación de una biblioteca de QCT puede incluir generar una biblioteca de QCT adaptada para la utilización (por ejemplo, en una única etapa de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación de alto rendimiento, etc.) de menos de 0,00001 nanogramos (y/u otras cantidades adecuadas) de moléculas QCT amplificables para cada muestra de un conjunto de muestras. Sin embargo, la determinación del número de moléculas QCT a incluir en una biblioteca de QCT y la generación de las bibliotecas de QCT se pueden realizar de cualquier manera adecuada.

En un ejemplo, las bibliotecas de QCT se pueden generar sintetizando secuencias oligonucleotídicas de cadena complementaria hacia secuencias oligonucleotídicas monocatenarias que contienen secuencias de "N" variables. En un ejemplo específico, se pueden generar bibliotecas de QCT bicatenarias resuspendiendo e hibridando los ultrámeros de QCT con una secuencia cebadora complementaria, prolongando las secuencias mediante el uso del fragmento Klenow (exo-) y tratando con exonucleasa I. El producto final se puede purificar para eliminar las moléculas de ADN monocatenarias no utilizadas y las bibliotecas de QCT se pueden cuantificar mediante ensayos fluorométricos como el ensayo Qubit HS, a partir del cual se puede calcular el número de moléculas QCT que se añadirán a cada muestra utilizando el peso molecular esperado de las moléculas QCT bicatenarias.

Sin embargo, la generación de moléculas QCT S110 se puede realizar de cualquier manera adecuada.

2.2 Determinación de un conjunto de grupos de lecturas de secuencias QCT.

Las realizaciones del método 100 pueden incluir la determinación de uno o más grupos S120 de lectura de secuencias QCT, que pueden funcionar para agrupar las lecturas de secuencias de moléculas QCT (por ejemplo, después de la preparación y secuenciación de la biblioteca de secuenciación, etc.) para facilitar la determinación de parámetros relacionados con la secuenciación.

Los grupos de lecturas de secuencias QCT incluyen preferiblemente las lecturas de secuencias de moléculas QCT (por ejemplo, derivadas de la secuenciación que corresponde a una o más mezclas de QCT generadas en base a uno o más conjuntos de moléculas QCT y una o más muestras que incluyen la diana biológica, etc.), pero pueden incluir adicional o alternativamente cualquier lectura adecuada y/o componente asociado con la secuenciación.

Las lecturas de secuencias de moléculas QCT se pueden agrupar computacionalmente, como para la determinación de la identidad de uno o más conjuntos de moléculas QCT que se han dispensado en cada muestra. Se pueden utilizar varios enfoques de agrupamiento computacional, que incluyen, entre otros, el análisis de componentes principales, K-medias, agrupamiento jerárquico y/o cualquier enfoque de agrupamiento basado en la identidad de secuencia. Adicional o alternativamente, el agrupamiento, los análisis computacionales asociados con el agrupamiento (por ejemplo, el preprocesamiento, filtrado, etc.) y/o cualquier otra parte adecuada de las realizaciones del método 100 pueden aplicar enfoques de inteligencia artificial (por ejemplo, enfoques de aprendizaje automático, etc.) que incluyen uno o más de: aprendizaje supervisado (p. ej., el uso de regresión logística, uso de redes neuronales de retropropagación, uso de bosques aleatorios, árboles de decisión, etc.), aprendizaje no supervisado (p. ej., el uso de

un algoritmo Apriori, el uso del agrupamiento de K-medias), aprendizaje semisupervisado, un algoritmo de aprendizaje profundo (por ejemplo, redes neuronales, una máquina de Boltzmann restringida, un método de red de creencias profundas, un método de red neuronal convolucional, un método de red neuronal recurrente, un método de autocodificador automático apilado, etc.), aprendizaje por refuerzo (por ejemplo, mediante el uso de un algoritmo de Q-learning, mediante el uso de aprendizaje por diferencia temporal), un algoritmo de regresión (por ejemplo, mínimos cuadrados ordinarios, regresión logística, regresión por pasos, splines de regresión adaptativa multivariante, suavizado de diagramas de dispersión estimados localmente, etc.), un método basado en ejemplos (por ejemplo, k-vecinos más cercanos, cuantificación de vectores de aprendizaje, mapa autoorganizado, etc.), un método de regularización (por ejemplo, regresión de crestas, operador de selección y contracción mínima absoluta, red elástica, etc.), un método de aprendizaje de árbol de decisión (por ejemplo, árbol de clasificación y regresión, dicotomizador iterativo 3, C4.5, detección automática de interacción chi-cuadrado, muñón de decisión, bosque aleatorio, splines de regresión adaptativa multivariante, máquinas de aumento de gradiente, etc.), un método bayesiano (por ejemplo, Bayes ingenuo, estimadores promediados de una dependencia, red de creencias bayesianas, etc.), un método de kernel (por ejemplo, una máquina de vectores de soporte, una función de base radial, un análisis discriminante lineal, etc.), un método de agrupamiento (por ejemplo, agrupamiento de k-medias, maximización de expectativas, etc.), un algoritmo de aprendizaje de reglas asociadas (por ejemplo, un algoritmo Apriori, un algoritmo Eclat, etc.), un modelo de red neuronal artificial (por ejemplo, un método Perceptron, un método de retropropagación, un método de red Hopfield, un método de mapa autoorganizado, un método de cuantificación de vectores de aprendizaje, etc.), un método de reducción de dimensionalidad (por ejemplo, análisis de componentes principales, regresión parcial de mínimos cuadrados, mapeo de Sammon, escalado multidimensional, búsqueda de proyecciones, etc.), un método de conjunto (por ejemplo, impulso, agregación de arranque, AdaBoost, generalización apilada, método de máquina de impulso de gradiente, método de bosque aleatorio, etc.) y/o cualquier enfoque de inteligencia artificial adecuado.

La determinación de los grupos de lecturas de secuencias QCT se basa preferiblemente en una o más regiones (por ejemplo, regiones de variación; regiones identificadoras de QCT, etc.) de las moléculas QCT (por ejemplo, basándose en las lecturas de secuencias que corresponden a las regiones de las moléculas QCT, etc.), pero puede basarse adicional o alternativamente en cualquier dato adecuado. En un ejemplo específico, después de que las moléculas QCT (p. ej., plantillas de control de calidad asociadas a la diana, etc.) se combinen con los componentes de la muestra, y se amplifique la diana biológica (p. ej., moléculas de ácido nucleico que incluyen la región de secuencia diana, etc.) mediante el uso de cebadores complementarios tanto de la región de secuencia diana como de la secuencia de la molécula QCT (por ejemplo, regiones asociadas a la diana de las moléculas QCT, etc.), las moléculas pueden indexarse para la multiplexación, secuenciarse y las lecturas de la secuenciación pueden separarse basándose en sus índices de multiplexación. En un ejemplo específico, las lecturas indexadas se pueden agrupar en diferentes grupos de QCT por las regiones identificadoras de QCT (por ejemplo, las secuencias ID de QCT, etc.) o se pueden identificar basándose en las coincidencias de secuencias exactas con las secuencias QCT esperadas (excepto las regiones de variación como la región EMI, etc.). En un ejemplo, la determinación (por ejemplo, computacionalmente, etc.) de un conjunto de grupos de lecturas de secuencias QCT puede incluir agrupar una primera lectura de secuencias de moléculas QCT y una segunda lectura de secuencias de moléculas QCT en un grupo de lecturas de secuencias QCT, del conjunto de grupos de lecturas de secuencias QCT, basándose en una similitud de secuencia de las regiones de variación (por ejemplo, entre una primera región de variación de la primera molécula QCT y una segunda región de variación de la segunda molécula QCT, etc.) que satisface una primera condición (por ejemplo, menos de un número umbral de bases de disimilitud, etc.), y para cada grupo de lecturas de secuencias QCT del conjunto de grupos de lecturas de secuencias QCT, determinar una asignación del grupo de lecturas de secuencias QCT a un identificador de muestra (por ejemplo, a una muestra, a un compartimento de muestras asociado con la preparación de una biblioteca de secuenciación y/o la secuenciación, etc.) de un conjunto de identificadores de muestras que identifican el conjunto de muestras, como cuando la determinación del parámetro relacionado con la secuenciación (por ejemplo, un parámetro de contaminación, etc.) puede basarse en el conjunto de grupos de lecturas de secuencias QCT y las asignaciones de los grupos de lecturas de secuencias QCT a los identificadores de muestras del conjunto de identificadores de muestras. En un ejemplo específico, la agrupación de la primera y la segunda lecturas de secuencias QCT puede incluir agrupar la primera y la segunda lecturas de secuencias QCT en el grupo de lecturas de secuencias QCT basándose en la similitud de secuencia de la región de variación de menos de tres sustituciones puntuales, y basándose en una profundidad de lectura asociada con el grupo de lecturas de secuencias QCT que satisface una segunda condición (por ejemplo, profundidad de lectura mayor de 20 por grupo de lecturas de secuencias QCT; profundidad de lectura mayor de 30; profundidad de lectura mayor de cualquier valor adecuado, etc.). En un ejemplo específico, se puede añadir una lectura de secuencias de moléculas QCT (por ejemplo, una lectura de secuencias que incluye una secuencia de una región EMI) si se observa otra lectura de secuencias de moléculas QCT con 2 o menos sustituciones puntuales en el mismo pocillo a una mayor profundidad de lectura. En un ejemplo específico, cada EMI se asigna a una muestra particular y al correspondiente pocillo e índice o pares de índices.

En las variaciones, la determinación de los grupos de lecturas de secuencias QCT puede incluir determinar y/o descartar (por ejemplo, filtrar, etc.) los grupos de lecturas de secuencias QCT no válidos (por ejemplo, grupos de EMI no válidos, etc.). En un ejemplo, como se muestra en la FIGURA 10, los grupos de lecturas de secuencias QCT no válidos pueden incluir grupos de lecturas de secuencias QCT con una profundidad de lectura por debajo y/o en un umbral (por ejemplo, 20 o menos lecturas; 30 o menos lecturas; un umbral de cualquier profundidad de lectura adecuada, etc.), y/o satisfacer cualquier condición adecuada (por ejemplo, un número de lecturas que coincidan con las condiciones de profundidad de lectura predeterminadas, etc.), tal como cuando los grupos de lecturas de

secuencias QCT no válidos se pueden descartar para el recuento de moléculas. En un ejemplo específico, se pueden usar grupos de lecturas de secuencias QCT válidos (por ejemplo, los grupos de lecturas de secuencias QCT restantes después de descartar los grupos de lecturas de secuencias QCT no válidos, etc.) para determinar la relación entre el número de plantilla de control de calidad y el recuento de lecturas de secuencias para cada muestra (por ejemplo, cuando la relación se puede utilizar como factor de corrección para cuantificar el número de moléculas diana, etc.). En un ejemplo específico, como se muestra en la FIGURA 10, a una profundidad de lectura de EMI promedio >30, los grupos de lecturas de secuencias QCT válidos frente a los no válidos (por ejemplo, grupos de EMI, etc.) pueden identificarse claramente mediante una marcada disminución en la profundidad de secuenciación, y a profundidades de lectura promedio más bajas, se pueden usar enfoques adaptativos (por ejemplo, la determinación del umbral de profundidad de lectura adaptativa, etc.) para identificar los EMI válidos frente a los no válidos. En un ejemplo específico, la determinación de un conjunto de grupos de lecturas de secuencias QCT puede incluir determinar un subconjunto filtrado de grupos de lecturas de secuencias QCT (por ejemplo, grupos de lecturas de secuencias QCT válidos, etc.) basándose en las profundidades de lectura (por ejemplo, que satisfacen unas condiciones de umbral de profundidad de lectura y/u otras condiciones adecuadas, etc.) que corresponden al subconjunto filtrado de grupos de lecturas de secuencias QCT, como cuando se determina un parámetro relacionado con la secuenciación (por ejemplo, el recuento de moléculas diana, como el número de moléculas diana presentes en la muestra original, etc.) puede incluir la determinación de un recuento de moléculas QCT basándose en el subconjunto filtrado de grupos de lecturas de secuencias QCT (por ejemplo, donde el número de grupos de lecturas de secuencias QCT en el subconjunto filtrado de grupos de lecturas de secuencias QCT puede corresponder al recuento de moléculas QCT; etc.); determinar una relación de factor de corrección basándose en el recuento de moléculas QCT y las lecturas de secuencias de moléculas QCT (por ejemplo, dividiendo el recuento de moléculas QCT por las lecturas de secuencias de moléculas QCT, etc.); y determinar el recuento de moléculas diana basándose en la relación del factor de corrección y las lecturas de la secuencia de la molécula diana derivadas de la secuenciación (por ejemplo, multiplicando el número de lecturas de la secuencia de la molécula diana por la relación del factor de corrección, etc.), las lecturas de la secuencia de la molécula diana asociadas con la diana biológica (por ejemplo, que incluye la región de secuencia diana de la molécula diana, etc.). En un ejemplo específico, el método 100 puede incluir determinar de forma adaptativa un umbral de profundidad de lectura basándose en las características de distribución de la profundidad de lectura para las lecturas de secuencias de moléculas QCT, y donde determinar el subconjunto filtrado de grupos de lecturas de secuencias QCT puede incluir determinar el subconjunto filtrado basándose en la satisfacción del umbral de profundidad de lectura determinado adaptativamente por las profundidades de lectura. En un ejemplo específico, cada profundidad de lectura de las profundidades de lectura puede corresponder a más de veinte lecturas (y/u otro número adecuado de lecturas, etc.) para el grupo de lecturas de secuencias QCT correspondiente del subconjunto filtrado de grupos de lecturas de secuencias QCT. En los ejemplos, debido a errores de secuenciación y de PCR, los grupos de lecturas de secuencias QCT no válidos pueden no ser válidos debido a aspectos distintos a la contaminación. Adicional o alternativamente, la determinación de los grupos de lecturas de secuencias QCT válidos o no válidos se puede realizar de cualquier manera adecuada. Sin embargo, la determinación de los grupos S120 de lecturas de secuencias QCT se puede realizar de cualquier manera adecuada.

2.3 Determinación de un parámetro relacionado con la secuenciación.

Las realizaciones del método 100 pueden incluir la determinación de uno o más parámetros S130 relacionados con la secuenciación.

Los parámetros relacionados con la secuenciación pueden incluir uno o más de: parámetros de contaminación (por ejemplo, que describen la contaminación asociada con la preparación y/o secuenciación de la biblioteca de secuenciación, tal como entre diferentes usuarios, muestras, experimentos, etc.); parámetros de recuento de moléculas (por ejemplo, que describen un número de moléculas, tales como las moléculas diana y/o las moléculas QCT, inicialmente presentes en una muestra y/o mezcla determinada, etc.); parámetros de seguimiento de muestras (por ejemplo, asociados con la pérdida de muestra, etc.); parámetros de error en el procesamiento de muestras (por ejemplo, que describen el ruido; operaciones erróneas en el procesamiento de muestras tales como errores de pipeta; errores sistemáticos, etc.); parámetros de error de cuantificación (por ejemplo, que describen los errores de cuantificación, etc.); parámetros de error de análisis (por ejemplo, que describen los errores de análisis computacional, etc.); y/o cualquier parámetro adecuado asociado con la preparación de una biblioteca de secuenciación, la secuenciación, el análisis de la asociación y/u otros aspectos adecuados. En un ejemplo, como se muestra en la FIGURA 11, se pueden usar los números de moléculas QCT determinados en una pluralidad de muestras para determinar los errores de procesamiento de muestras que describen el ruido y/o el procesamiento erróneo de muestras; donde se puede añadir el mismo volumen de moléculas QCT, que corresponde aproximadamente a ~200 moléculas QCT únicas, a cada muestra antes de la PCR, y se pueden determinar los grupos de lecturas de secuencias QCT válidos (por ejemplo, grupos de EMI, etc.) a partir de los datos de secuenciación posteriores a la PCR y la secuenciación; donde el coeficiente de variación (CV) esperado para ~200 moléculas QCT es $\sqrt{200}/200 \sim 7\%$, lo cual es coherente con los datos observados que se muestran en la FIGURA 11 en las 12 muestras; donde si alguna muestra cae por debajo de un cierto umbral (por ejemplo, 3 sigmas, $200 \cdot 3 \cdot \sqrt{200} \sim 150$ o un umbral menos estricto de $\sim 200/2 \sim 100$), el resultado se puede utilizar para identificar el error de procesamiento de la muestra para esa muestra en particular; y donde el número de moléculas QCT también se puede aumentar para determinar los parámetros de error del procesamiento de muestras adicionales que corresponden a menos del 7% de CV en un proceso. En un ejemplo, la determinación del parámetro relacionado con la secuenciación puede incluir identificar las

lecturas de secuencias QCT que no están asignadas a un grupo de lecturas de secuencias QCT del conjunto de grupos de lecturas de secuencias QCT; y determinar al menos uno de una tasa de error de secuenciación y una tasa de error de la polimerasa (por ejemplo, tasas de error de secuenciación de extremo a extremo y de la polimerasa, etc.) a partir de un número de lecturas de secuencias QCT que no están asignadas y un número total de lecturas de secuencias QCT. En los ejemplos específicos, cualquier secuencia que tenga regiones variables (por ejemplo, regiones de variación de la diana, regiones de variación de referencia, etc.) respecto de las secuencias diana o de referencia pero que no sean idénticas en secuencia a una secuencia del grupo de lecturas de QCT se debe a errores de la secuencia o de la polimerasa. En un ejemplo específico, los recuentos de lecturas de estas secuencias, divididos por los recuentos totales de las lecturas de QCT, son la frecuencia combinada de errores de secuenciación y de la polimerasa. Los primeros, los errores de secuenciación, pueden producirse por un proceso lineal, mientras que los errores de la polimerasa pueden producirse por un proceso exponencial (por ejemplo, a menos que se emplee una PCR lineal), en donde el efecto de un error en un ciclo anterior de la PCR puede amplificarse exponencialmente. Por lo tanto, en un ejemplo específico, al analizar la distribución de los recuentos de lecturas de secuencias que no están asignadas a los grupos de lecturas de QCT, se puede calcular la contribución de los errores de secuenciación frente a la polimerasa. Sin embargo, la determinación de las tasas de error de secuenciación y/o las tasas de error de la polimerasa se puede realizar de cualquier manera adecuada.

En las variaciones, la determinación de los parámetros relacionados con la secuenciación puede basarse en el procesamiento con una pluralidad de conjuntos de moléculas QCT (por ejemplo, diferentes conjuntos de moléculas QCT identificadas por diferentes regiones identificadoras de QCT compartidas; diferentes conjuntos de moléculas QCT desplegadas en diferentes etapas asociadas con la preparación de las bibliotecas de secuenciación y/o la secuenciación, etc.), tales como basados en diferentes subconjuntos de grupos de lecturas de secuencias QCT que corresponden a los diferentes conjuntos de moléculas QCT. En un ejemplo, el método 100 puede incluir generar un conjunto de moléculas QCT, y cada molécula QCT incluye una primera región identificadora de QCT compartida entre el conjunto de moléculas QCT y adaptada para identificar la molécula QCT; generar un conjunto de moléculas QCT adicionales, y cada molécula QCT adicional incluye una segunda región identificadora de QCT compartida entre el conjunto de moléculas QCT adicionales y adaptada para identificar la molécula QCT adicional; determinar el conjunto de grupos de lecturas de secuencias QCT basándose en la primera y segunda regiones identificadoras de QCT; y determinar el parámetro relacionado con la secuenciación basándose en el conjunto de grupos de lecturas de secuencias QCT. En un ejemplo específico, el conjunto de moléculas QCT se puede adaptar para la utilización en una primera etapa de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, donde el conjunto de moléculas QCT adicionales se adapta para la utilización en una segunda etapa de la al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT incluye: determinar un primer subconjunto del conjunto de grupos de lecturas de secuencias QCT (por ejemplo, basándose en la primera región identificadora de QCT y las primeras regiones de variación de las primeras moléculas QCT correspondientes, etc.), donde el primer subconjunto corresponde a la primera región identificadora de QCT y está asociado con la primera etapa; y determinar un segundo subconjunto del conjunto de grupos de lecturas de secuencias QCT (por ejemplo, basándose en la segunda región identificadora de QCT y las segundas regiones de variación de las segundas moléculas QCT correspondientes, etc.), donde el segundo subconjunto corresponde a la segunda región identificadora de QCT y está asociado con la segunda etapa; y donde la determinación del parámetro relacionado con la secuenciación incluye la determinación de un parámetro de seguimiento de muestras asociado con la pérdida de muestra, basándose en el primer y segundo subconjuntos del conjunto de grupos de lecturas de secuencias QCT.

En un ejemplo, la determinación de un parámetro relacionado con la secuenciación puede incluir determinar un primer recuento absoluto y un segundo recuento absoluto correspondientes al conjunto de moléculas QCT y al conjunto de moléculas QCT adicionales, respectivamente, basándose en el conjunto de grupos de lecturas de secuencias QCT, y determinar al menos uno de un parámetro de error de pipeta y un parámetro de error de cuantificación basándose en el primer y segundo recuentos absolutos.

En un ejemplo específico, como se muestra en la FIGURA 12, el uso de las moléculas QCT en diferentes etapas puede permitir la comparación de los diferentes enfoques de preparación de muestras; tales como cuando los enfoques de purificación de ADN pueden evaluarse mediante la adición de 200 moléculas QCT1 (y/o cualquier número adecuado de moléculas QCT) a cada muestra de plasma antes de la purificación del ADN; donde el ADN se purificó del plasma mediante el método de purificación n.º 1 o el método de purificación n.º 2, y las muestras de ADN resultantes se amplificaron mediante PCR y se secuenciaron; donde se añadieron 200 moléculas QCT2 (y/o cualquier número adecuado de moléculas QCT) después de la purificación del ADN pero antes de la amplificación mediante PCR; donde el número de grupos de lecturas de secuencias QCT válidos que corresponden a las moléculas QCT2 fue similar en dos muestras (dentro del ~25%), lo que indica que el procesamiento después de la purificación no fue diferente para estas dos muestras; y donde hubo ~3 veces menos grupos de lecturas de secuencias QCT válidos para QCT1 para el método de purificación n.º 1, lo que indica que el método de purificación n.º 1 da como resultado una pérdida significativa de muestra (p. ej., de cfADN).

Sin embargo, la determinación de los parámetros S130 relacionados con la secuenciación se puede realizar de cualquier manera adecuada.

2.3.A Determinación de un parámetro de contaminación.

La determinación de los parámetros S130 relacionados con la secuenciación puede incluir adicional o alternativamente la determinación de uno o más parámetros de contaminación S132. Los parámetros de contaminación pueden incluir uno o más de un parámetro de contaminación cruzada (por ejemplo, que describe la contaminación cruzada entre muestras y/o compartimentos de muestras asociados con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación; contaminación cruzada entre diferentes usuarios, etc.), un parámetro de contaminación por arrastre (por ejemplo, que describe la contaminación por arrastre a través de una pluralidad de casos de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, etc.), un parámetro de contaminación por saltos de índices (por ejemplo, que describe la contaminación por saltos de índices asociada con cebadores con saltos de índices, etc.). Los parámetros de contaminación pueden describir un grado de asignación incorrecta de índices (por ejemplo, asociado con la secuenciación de alto rendimiento, etc.), como cuando un parámetro de contaminación puede describir (por ejemplo, un efecto acumulativo de) la contaminación cruzada (y/u otra contaminación adecuada) y la asignación incorrecta de índices, y/o cualquier otra característica adecuada asociada con la preparación de una biblioteca de secuenciación y/o la secuenciación.

En un ejemplo, la determinación de un parámetro de contaminación puede incluir determinar un porcentaje o fracción contaminante total para una muestra particular basándose en sumar las profundidades de lecturas para las secuencias contaminantes (por ejemplo, que se encuentra que está asociado con la muestra particular; que se encuentra que está en un compartimento de muestras correspondiente a la muestra, etc.) y dividirlo por el número total de lecturas (o el número total de lecturas de secuencias de moléculas QCT asociadas con los grupos de lecturas de secuencias QCT válidos). En un ejemplo específico, como se muestra en la FIGURA 10, se pueden determinar los parámetros de contaminación, donde si la secuencia de un grupo de EMI no válido para la secuenciación de la Muestra A se encuentra como un grupo de EMI válido en otra muestra (Muestra B), indica que esta lectura en la Muestra A se debe a la contaminación de la Muestra B; donde, al encontrar y sumar las profundidades de lecturas para todas esas secuencias contaminantes y dividir por el número total de lecturas (o el número total de lecturas que se asignan a los grupos de EMI válidos), se puede determinar un porcentaje o fracción total de contaminación para una muestra en particular; y donde el porcentaje o fracción contaminante total se puede utilizar en el análisis del nivel máximo de sensibilidad y especificidad analítica que el ensayo clínico puede informar, y/o como umbral para informar un ensayo fallido y/o un resultado ininterpretable en lugar de un falso positivo; como cuando, si un ensayo en particular requiere la detección de fracciones de alelos del 0,1%, se puede utilizar una fracción de contaminación total igual o superior al 0,1% para esa muestra para identificar un resultado ininterpretable; y donde, alternativamente, el conocimiento de las fracciones alélicas de las muestras contaminantes se puede utilizar para adaptar este umbral (es decir, para la medición de un alelo particular en una muestra determinada, un 1% de contaminación de otra muestra que tiene un 10% para el mismo alelo tiene el mismo efecto que una contaminación del 10% de una muestra que tiene ese alelo al 1%).

En un ejemplo específico, como se muestra en las FIGURAS 4A-4D, la contaminación se puede medir identificando el origen y el destino de las lecturas de secuencias de moléculas QCT (por ejemplo, las lecturas de las secuencias EMI, etc.) en cada compartimento de muestras (por ejemplo, un pocillo, etc.). En un ejemplo específico, si se observa la misma lectura de secuencias de moléculas QCT (por ejemplo, la misma lectura de secuencias EMI) en una pluralidad de compartimentos de muestras (por ejemplo, una pluralidad de pocillos, etc.), la lectura de secuencias de moléculas QCT puede marcarse como originaria del compartimento de muestras, de la pluralidad de compartimentos de muestras, con la mayor profundidad de lectura y puede considerarse un contaminante en los otros compartimentos de muestras de la pluralidad de compartimentos de muestras (por ejemplo, los otros pocillos, etc.). En un ejemplo específico, la determinación de un parámetro de contaminación puede incluir la identificación de un primer y un segundo grupo de lecturas de secuencias QCT que corresponden a una secuencia de región de variación compartida, donde las asignaciones del primer y segundo grupo de lecturas de secuencias QCT son para identificadores de muestras distintos (p. ej., que identifican distintos compartimentos de muestras; distintas muestras, etc.) del conjunto de identificadores de muestras; generar una comparación de profundidades de lectura entre una primera profundidad de lectura asociada con el primer grupo de lecturas de secuencias QCT y una segunda profundidad de lectura asociada con el segundo grupo de lecturas de secuencias QCT; y basándose en la comparación de la profundidad de lectura, determinar el parámetro de contaminación asociado con una muestra identificada por un identificador de muestra distinto de los identificadores de muestras distintos.

En un ejemplo, la determinación de un parámetro de contaminación puede incluir determinar una primera huella molecular asociada con la primera amplificación en un primer caso de la preparación de una biblioteca de secuenciación, basándose en un conjunto de grupos de lecturas de secuencias QCT; determinar una segunda huella molecular asociada con la segunda amplificación en un segundo caso de la preparación de una biblioteca de secuenciación, basándose en un conjunto adicional de grupos de lecturas de secuencias QCT; y basándose en una comparación entre la primera y la segunda huellas moleculares, determinar un parámetro de contaminación por arrastre que describe la contaminación por arrastre del primer caso al segundo caso.

En una variación, la determinación de un parámetro de contaminación puede incluir determinar parámetros de contaminación por saltos de índices. En un ejemplo específico, como se muestra en las FIGURAS 13A-13B, se pueden usar moléculas QCT para facilitar la identificación y eliminación de cebadores contaminantes y/o con saltos de índices; donde, como se muestra en la FIGURA 13A, cada muestra recibió un código mediante un cebador de indexación D7xx correspondiente y se analizó en el mismo carril de celda de flujo de secuenciación para un experimento de validación;

donde se encontró que D701 y D707 tenían altas fracciones contaminantes que se originan entre sí, posiblemente debido a que los oligos de indexación D701 y D707 se sintetizaron en la misma columna de síntesis de oligos, error de síntesis o saltos de índices, y donde el nivel es significativo al 5% y puede afectar a los resultados clínicos; y donde, como se muestra en la FIGURA 13B, en los análisis posteriores con muestras clínicas, no se usaron los cebadores de indexación, lo que disminuyó el nivel máximo de contaminación por debajo del 1%.

En un ejemplo específico, como se muestra en la FIGURA 14, se pueden usar moléculas QCT para facilitar la medición de los niveles de contaminación reales asociados con el uso de cebadores de índice dual único; donde los cebadores de indexación dual estándar pueden dar como resultado una contaminación del 0,1% (como se muestra en las Muestras 1 a 9), debido a una combinación de contaminación verdadera de muestra a muestra, saltos de índices y/o contaminación por oligos de indexación; donde se espera que la indexación dual única reduzca el efecto de los saltos de índices y la contaminación por oligos de indexación a $0,001 \times 0,001 \sim 1e-6$; pero donde las mediciones indican hasta un 0,03% ($3e-5$) de fracciones contaminantes en las reacciones con indexación dual única (como se muestra en las Muestras 10 a 29), que es mayor que la contaminación esperada $1e-6$, lo que puede indicar la detección de niveles de contaminación verdadera en las condiciones de laboratorio para el ensayo concreto.

Sin embargo, la determinación de los parámetros de contaminación S132 se puede realizar de cualquier manera adecuada.

2.3.B Determinación de un parámetro de recuento de moléculas.

La determinación de los parámetros S130 relacionados con la secuenciación puede incluir adicional o alternativamente la determinación de uno o más parámetros S134 de recuento de moléculas. Los parámetros de recuento de moléculas pueden incluir uno o más recuentos de moléculas diana (por ejemplo, el recuento absoluto de moléculas diana, como en la muestra original; el recuento absoluto de moléculas diana endógenas, como en la muestra original; etc.); recuentos de moléculas de referencia (por ejemplo, el recuento absoluto de moléculas de referencia endógenas, como en la muestra original, etc.); recuentos de moléculas QCT (por ejemplo, que corresponden a una serie de grupos de lecturas de secuencias QCT válidos; que corresponden a una serie de moléculas QCT distintas añadidas a los componentes de la muestra; etc.); relaciones asociadas (por ejemplo, factores de corrección; relaciones entre un recuento de moléculas y un número asociado de lecturas de secuencia; etc.); y/o cualquier otro parámetro adecuado asociado con los recuentos de moléculas.

Los parámetros de recuentos de moléculas se usan preferiblemente para facilitar uno o más diagnósticos, pero se pueden usar adicional o alternativamente (por ejemplo, como entradas) para cualquier parte adecuada de las realizaciones del método 100.

En las variaciones, la determinación de un parámetro de recuento de moléculas (por ejemplo, recuento de moléculas diana, etc.) puede basarse en una relación de factor de corrección determinada basándose en un recuento de moléculas QCT (por ejemplo, correspondiente a una serie de grupos de lecturas de secuencias QCT, tales como un número de grupos de lecturas de secuencias QCT válidos, etc.) y lecturas de secuencias de moléculas QCT (por ejemplo, un número de lecturas de secuencias de moléculas QCT correspondiente a los grupos de lecturas de secuencias QCT, etc.), tal como multiplicando el número de lecturas de secuencias de moléculas diana por la relación del factor de corrección. En un ejemplo específico, el número de grupos de lecturas de secuencias QCT no contaminantes válidos (p. ej., grupos de lecturas de secuencias QCT restantes después de descartar los grupos de lecturas de secuencias QCT con 2 o menos lecturas, y/o con cualquier número adecuado o menos de lecturas; etc.) puede indicar el recuento de moléculas QCT (por ejemplo, el número de moléculas QCT para un compartimento de muestras particular; para una muestra particular; para un identificador de muestra particular, etc.). En un ejemplo específico, dividiendo el recuento de moléculas QCT por las lecturas de secuenciación resultantes de las correspondientes moléculas QCT, se puede encontrar el factor de corrección, como por ejemplo donde el factor de corrección se multiplica por las lecturas de secuenciación que pertenecen a las moléculas diana (por ejemplo, en el compartimento de muestras particular; de la muestra particular; asociado con el identificador de muestra particular; etc.) daría como resultado un recuento de moléculas diana (por ejemplo, un número absoluto de moléculas diana biológicas iniciales a las que se podía acceder mediante el ensayo para la amplificación, etc.). En un ejemplo, la profundidad de la secuenciación de QCT promedio utilizada para determinar el recuento absoluto de las moléculas diana endógenas y el recuento absoluto de las moléculas de referencia endógenas se determina por separado de sus QCT correspondientes.

Alternativamente, en una variación de una realización, el umbral de profundidad de lectura para descartar grupos de lecturas de secuencias QCT (por ejemplo, para determinar los parámetros de recuento de moléculas y/o los parámetros adecuados relacionados con la secuenciación, etc.) se puede determinar de forma adaptativa basándose en las características de la distribución de la profundidad de lectura de secuencias de moléculas QCT (por ejemplo, la lectura de secuencias EMI). Por ejemplo, se puede establecer un umbral para cada muestra indexada calculando la profundidad de lectura media de EMI dentro de cada muestra, calculando la raíz cuadrada de esta profundidad de lectura media y descartando los grupos de lecturas de secuencias QCT con una profundidad de lectura por debajo de la raíz cuadrada de la profundidad de lectura media. Adicional o alternativamente, los umbrales de profundidad de lectura para descartar grupos de lecturas de secuencias QCT se pueden calcular de cualquier manera adecuada.

Sin embargo, la determinación de los parámetros S134 de recuento de moléculas se puede realizar de cualquier manera adecuada.

2.4 Facilitación del diagnóstico.

- 5 Las realizaciones del método 100 pueden incluir adicional o alternativamente la facilitación del diagnóstico S140, que puede funcionar para ayudar, determinar, proporcionar y/o facilitar de otro modo uno o más diagnósticos para una o más afecciones.

10 La facilitación de uno o más diagnósticos puede incluir uno o más de determinar uno o más diagnósticos (por ejemplo, basándose en uno o más parámetros relacionados con la secuenciación, etc.); proporcionar uno o más diagnósticos (por ejemplo, a uno o más usuarios; a uno o más proveedores de atención médica, tal como para el uso por parte de uno o más proveedores de atención médica para proporcionar diagnósticos médicos a pacientes, etc.); ayudar en uno o más diagnósticos (por ejemplo, proporcionar uno o más parámetros relacionados con la secuenciación y/u otros parámetros adecuados a uno o más proveedores de atención médica y/u otras entidades adecuadas, para el uso en la determinación de un diagnóstico, tal como en combinación con otros datos, etc.); y/o cualquier proceso adecuado asociado con el diagnóstico. Por ejemplo, ayudar en el diagnóstico puede incluir proporcionar un parámetro de contaminación (por ejemplo, a un usuario; a un proveedor de atención médica; etc.) adaptado para su uso en la determinación de un resultado de diagnóstico para ensayos asociados con al menos una de pruebas prenatales no invasivas y biopsias líquidas. En un ejemplo, la determinación del recuento de moléculas diana (y/o parámetros adecuados relacionados con la secuenciación, etc.) puede incluir determinar el recuento de moléculas diana (y/o los parámetros adecuados relacionados con la secuenciación, etc.) para facilitar el diagnóstico asociado con al menos una de pruebas prenatales no invasivas y biopsias líquidas.

25 En las variaciones, la facilitación del diagnóstico puede incluir facilitar el diagnóstico prenatal (por ejemplo, asociado con pruebas prenatales no invasivas; para trastornos genéticos asociados y/o afecciones adecuadas, etc.). En un ejemplo, la facilitación del diagnóstico puede incluir facilitar el diagnóstico prenatal de uno o más trastornos genéticos (por ejemplo, trastornos de un solo gen, anomalías cromosómicas, etc.) basándose en el parámetro de recuento de moléculas diana y un parámetro de recuento de moléculas de referencia (por ejemplo, basándose en una comparación entre el recuento absoluto de las secuencias diana endógenas y el recuento absoluto de las secuencias de referencia endógenas, etc.).

30 En las variaciones, la facilitación del diagnóstico puede incluir facilitar el diagnóstico de uno o más trastornos de un solo gen (y/o trastornos genéticos adecuados). Por ejemplo, la determinación del recuento absoluto de moléculas diana endógenas puede incluir determinar el recuento absoluto de las moléculas diana endógenas que incluyen una mutación asociada con el trastorno de un solo gen (por ejemplo, basándose en dividir el recuento total de lecturas para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, tal como se obtiene dividiendo un número de lecturas de secuencias de moléculas QCT por el número único de moléculas QCT, etc.), donde la determinación del recuento absoluto de las moléculas de referencia endógenas puede incluir la determinación del recuento absoluto de las moléculas de referencia endógenas que carecen de la mutación (por ejemplo, basándose en dividir el recuento total de lecturas para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, etc.); y facilitar el diagnóstico prenatal del trastorno genético del trastorno de un solo gen basándose en (por ejemplo, una comparación entre, etc.) el recuento absoluto de las secuencias diana endógenas y el recuento absoluto de las secuencias de referencia endógenas. En los ejemplos específicos, como se muestra en las FIGURAS 15A-15D, el número de alelos de enfermedad y no enfermedad se puede medir y comparar en una muestra para determinar el genotipo de un feto en desarrollo a partir de la sangre materna; donde la FIGURA 15A incluye el número de moléculas de HbS (hemoglobina mutada) y HbA (hemoglobina normal) tal como se mide mediante las moléculas QCT en muestras con rasgos de células falciformes (SCT) donde se espera que los alelos de HbS y HbA tengan la misma frecuencia, lo que representa los casos en los que tanto la madre embarazada como el feto en desarrollo son heterocigotos para el trastorno; donde la FIGURA 15B incluye el número de moléculas de HbS y HbA medidas por las QCT en muestras de anemia falciforme (SCD) con SCT+10%, que representan los casos en los que la madre embarazada es portadora del trastorno y el feto en desarrollo ha heredado el alelo de la enfermedad de ambos padres y por lo tanto tiene la enfermedad; donde la FIGURA 15C incluye una cantidad de moléculas y una medición de la fracción fetal (por ejemplo, medición en hasta 9 loci donde la madre y el feto difieren en el genotipo) para el uso en el cálculo de la probabilidad posterior de que el feto haya heredado el trastorno mediante un análisis de la dosis de mutación relativa (RMD); y donde la FIGURA 15D incluye la media y los intervalos de confianza del 95% para una adición de SCD del 0% frente al 10% a la muestra de SCT (por ejemplo, que representa el feto portador frente a la enfermedad de una madre embarazada que es portadora del trastorno). Sin embargo, la facilitación del diagnóstico de trastornos de un solo gen se puede realizar de cualquier manera adecuada.

55 En las variaciones, la facilitación del diagnóstico puede incluir facilitar el diagnóstico de una o más anomalías cromosómicas (y/o trastornos genéticos adecuados). Por ejemplo, la determinación del recuento absoluto de las moléculas diana endógenas puede incluir determinar el recuento absoluto de las moléculas diana endógenas asociadas con un primer cromosoma (por ejemplo, basándose en dividir el recuento total de lecturas para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, etc.), donde la determinación del recuento absoluto de las moléculas de referencia endógenas puede incluir la determinación del recuento absoluto de las moléculas de referencia endógenas asociadas con un segundo cromosoma (por ejemplo, basándose en dividir el

recuento total de lecturas para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, etc.); y facilitar el diagnóstico prenatal de la anomalía cromosómica basándose en (por ejemplo, una comparación entre, etc.) el recuento absoluto de secuencias diana endógenas y el recuento absoluto de secuencias de referencia endógenas. En los ejemplos específicos, como se muestra en las FIGURAS 16A-16B, el número de Chr21 y otro cromosoma se puede contar de manera similar mediante el uso de las moléculas QCT para determinar si hay un número excesivo de Chr21 (por ejemplo, en comparación con otro cromosoma) para indicar que el feto tiene síndrome de Down; donde, para una diferencia de 3 frente a 2 cromosomas que se cuentan, la señal puede ser la mitad de un trastorno recesivo hereditario (p. ej., HbSS frente a HbAS es una señal de 2 frente a 1; un aumento del 100% frente a un aumento del 50%), lo que puede indicar el requisito de contar más de un locus en cada cromosoma para mejorar la precisión en la medición del síndrome de Down en el feto en desarrollo a partir del ADN circulante en la sangre materna; y donde pueden usarse adicional o alternativamente partes de las realizaciones del método 100 para facilitar el diagnóstico de otras mutaciones y/o anomalías cromosómicas *de novo* tales como la trisomía 18 y/o el síndrome de DiGeorge.

En las variaciones, la facilitación del diagnóstico puede incluir facilitar el diagnóstico de una o más microdeleciones cromosómicas. Por ejemplo, la determinación del recuento absoluto de las moléculas diana endógenas puede incluir determinar el recuento absoluto de las moléculas diana endógenas asociadas con una región de microdelección, basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, donde la determinación del recuento absoluto de las moléculas de referencia endógenas puede incluir determinar el recuento absoluto de las moléculas de referencia endógenas asociadas con una segunda región cromosómica que no se espera que tenga una microdelección, basándose en dividir el recuento total de lecturas para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, y donde la facilitación del diagnóstico (por ejemplo, el diagnóstico prenatal, etc.) del trastorno genético puede incluir facilitar el diagnóstico (por ejemplo, el diagnóstico prenatal, etc.) de la microdelección cromosómica basándose en la comparación.

En las variaciones, la facilitación del diagnóstico puede incluir facilitar el diagnóstico de una o más variaciones del número de copias. Por ejemplo, la determinación del recuento absoluto de las moléculas diana endógenas puede incluir determinar el recuento absoluto de las moléculas diana endógenas asociadas con una región que puede tener una variación del número de copias, basándose en dividir el recuento total de lecturas para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, donde la determinación del recuento absoluto de las moléculas de referencia endógenas puede incluir la determinación del recuento absoluto de las moléculas de referencia endógenas asociadas con una región que no se espera que tenga una variación del número de copias, basándose en dividir el recuento total de lecturas para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, y donde la facilitación del diagnóstico (por ejemplo, diagnóstico prenatal) del trastorno genético puede incluir facilitar el diagnóstico (por ejemplo, diagnóstico prenatal) de la variación del número de copias basándose en la comparación.

Adicional o alternativamente, la facilitación del diagnóstico puede ser para cualquier afección adecuada.

Como se muestra en las FIGURAS 15C y 16B, la facilitación del diagnóstico puede basarse en una o más mediciones de la fracción fetal. Por ejemplo, la facilitación del diagnóstico prenatal puede incluir facilitar el diagnóstico prenatal del trastorno genético basándose en una medición de la fracción fetal, el recuento absoluto de secuencias diana endógenas y el recuento absoluto de secuencias de referencia endógenas. Sin embargo, el uso de las mediciones de la fracción fetal se puede realizar de cualquier manera adecuada para cualquier proceso adecuado de las realizaciones del método 100, y la facilitación del diagnóstico S140 se puede realizar de cualquier manera adecuada.

Sin embargo, las realizaciones del método 100 se pueden realizar de cualquier manera adecuada.

Las realizaciones del método 100 y/o del sistema 200 pueden incluir cada combinación y permutación de los diversos componentes del sistema y los diversos procesos del método, que incluyen cualquier variante (por ejemplo, realizaciones, variaciones, ejemplos, ejemplos específicos, figuras, etc.), donde las partes de las realizaciones del método 100 y/o los procesos descritos en la presente memoria se pueden realizar de forma asincrónica (por ejemplo, secuencialmente), concurrente (por ejemplo, en paralelo), o en cualquier otro orden adecuado mediante y/o usando uno o más casos, elementos, componentes y/u otros aspectos del sistema 200 y/u otras entidades descritas en la presente memoria.

Las partes de las realizaciones del método 100 y/o del sistema 200 pueden realizarse y/o implementarse al menos en parte como una máquina configurada para recibir un medio legible por computadora que almacena instrucciones legibles por computadora. Las instrucciones pueden ejecutarse mediante componentes ejecutables por computadora que pueden estar integrados con el sistema. El medio legible por computadora puede almacenarse en cualquier medio legible por computadora adecuado tal como memorias RAM, ROM, memoria flash, EEPROM, dispositivos ópticos (CD o DVD), discos duros, unidades de disquete o cualquier dispositivo adecuado. El componente ejecutable por computadora puede ser un procesador general o específico de la aplicación, pero cualquier hardware dedicado adecuado o dispositivo con combinación de hardware/firmware puede ejecutar las instrucciones de forma alternativa o adicional.

REIVINDICACIONES

1. Un método para analizar una muestra materna obtenida de una mujer embarazada para facilitar la determinación prenatal de la presencia de un trastorno genético, y el método comprende:

• añadir a la muestra materna un conjunto de moléculas plantilla de control de calidad (QCT) asociadas con el trastorno genético, y el conjunto de moléculas QCT comprende:

• regiones asociadas a la diana con similitud de secuencia respecto de una región de secuencia diana de las moléculas diana endógenas, y

• regiones de variación con disimilitud de secuencia respecto de una región de secuencia de las moléculas diana endógenas, en donde las regiones de variación comprenden una región identificadora molecular incrustada que tiene una secuencia de nucleótidos única;

• generar una mezcla coamplificada basándose en la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que comprenden la región de secuencia diana de las moléculas diana endógenas;

• secuenciar la mezcla coamplificada;

• determinar computacionalmente un número único del conjunto de moléculas QCT, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT que se detectan a partir de las lecturas de secuencias de moléculas QCT identificadas durante la secuenciación, en donde las lecturas de secuencias de moléculas QCT corresponden al conjunto de moléculas QCT;

• calcular la profundidad de la secuenciación de QCT promedio basándose en dividir el número de lecturas de secuencias de moléculas QCT por el número único de moléculas QCT;

• determinar el recuento absoluto de las moléculas diana endógenas basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio;

• determinar el recuento absoluto de las moléculas de referencia endógenas basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio; y

• facilitar la determinación prenatal de la presencia del trastorno genético basándose en una comparación entre el recuento absoluto de las secuencias diana endógenas y el recuento absoluto de las secuencias de referencia endógenas.

2. El método de la reivindicación 1, en donde:

i) el trastorno genético comprende un trastorno de un solo gen,

en donde la determinación del recuento absoluto de las moléculas diana endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas diana endógenas que tienen una mutación asociada con el trastorno de un solo gen, y

en donde la determinación del recuento absoluto de las moléculas de referencia endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas de referencia endógenas que carecen de la mutación; y en donde la facilitación del diagnóstico prenatal del trastorno genético comprende facilitar el diagnóstico prenatal del trastorno de un solo gen basándose en la comparación;

ii) el trastorno genético comprende una anomalía cromosómica,

en donde la determinación del recuento absoluto de las moléculas diana endógenas comprende determinar, basándose en dividir el recuento total de lecturas para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas diana endógenas que están asociadas con un primer cromosoma, en donde la determinación del recuento absoluto de las moléculas de referencia endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas de referencia endógenas que están asociadas con un segundo cromosoma; y

en donde la facilitación del diagnóstico prenatal del trastorno genético comprende facilitar el diagnóstico prenatal de la anomalía cromosómica basándose en la comparación;

iii) el trastorno genético comprende una microdelección cromosómica,

en donde la determinación del recuento absoluto de las moléculas diana endógenas comprende determinar, basándose en dividir el recuento total de lecturas para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas diana endógenas que tienen la microdelección cromosómica en una primera región cromosómica,

en donde la determinación del recuento absoluto de las moléculas de referencia endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas de referencia endógenas que no se espera que tengan la microdelección cromosómica en una segunda región cromosómica; y

en donde la facilitación del diagnóstico prenatal del trastorno genético comprende facilitar el diagnóstico prenatal de la microdelección cromosómica basándose en la comparación; o

iv) el trastorno genético comprende una variación del número de copias,

en donde la determinación del recuento absoluto de las moléculas diana endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas diana endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas diana endógenas asociadas con una región que tiene una variación del número de copias,

en donde la determinación del recuento absoluto de las moléculas de referencia endógenas comprende determinar, basándose en dividir el recuento de lecturas total para las moléculas de referencia endógenas por la profundidad de la secuenciación de QCT promedio, el número de moléculas de referencia endógenas asociadas con una región que no se espera que tenga una variación del número de copias, y

en donde la facilitación del diagnóstico prenatal del trastorno genético comprende facilitar el diagnóstico prenatal de la variación del número de copias basándose en la comparación.

3. El método de la reivindicación 1, en donde la profundidad de la secuenciación de QCT promedio utilizada para determinar el recuento absoluto de las moléculas diana endógenas y el recuento absoluto de las moléculas de referencia endógenas se determina por separado a partir de sus QCT correspondientes.

4. El método de la reivindicación 1, en donde la facilitación de la determinación prenatal de la presencia del trastorno genético se basa además en una medición de la fracción fetal.

5. Un método para identificar la contaminación asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación de alto rendimiento, y el método comprende:

- generar un conjunto de moléculas plantilla de control de calidad (QCT), y cada molécula QCT comprende:

- una región asociada a una diana con similitud de secuencia respecto de una región de secuencia diana de una diana biológica, y

- una región de variación con disimilitud de secuencia respecto de una región de secuencia de la diana biológica en donde la región de variación comprende una región identificadora molecular incrustada que tiene una secuencia de nucleótidos única; y

- determinar computacionalmente un conjunto de grupos de lecturas de secuencias QCT basándose en las regiones de variación del conjunto de moléculas QCT, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT que se detectan a partir de las lecturas de secuencias de moléculas QCT identificadas durante la secuenciación, en donde las lecturas de secuencias de moléculas QCT corresponden al conjunto de moléculas QCT,

- en donde el conjunto de grupos de lecturas de secuencias QCT comprende las lecturas de secuencias de moléculas QCT derivadas de la secuenciación de alto rendimiento que corresponden a un conjunto de mezclas de QCT generadas basándose en el conjunto de moléculas QCT y un conjunto de muestras que comprenden la diana biológica,

- en donde la preparación de una biblioteca de secuenciación comprende la coamplificación del conjunto de moléculas QCT y moléculas de ácido nucleico que comprenden la diana biológica, basándose en la similitud de secuencia de la región asociada a la diana y la región de secuencia diana de la diana biológica en la que el conjunto de moléculas QCT se añade a una concentración de menos de 0,00001 nanogramos de moléculas QCT amplificables por muestra a más de una muestra en la biblioteca de secuenciación, y

- en donde el conjunto de grupos de lecturas de secuencias QCT permite la determinación del recuento de moléculas QCT que estaban en la mezcla antes de la coamplificación; y

- basándose en el conjunto de grupos de lecturas de secuencias QCT, determinar un parámetro de contaminación que describe la contaminación asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación de alto rendimiento.

6. El método de la reivindicación 5, en donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT comprende:

- agrupar una primera lectura de secuencias de moléculas QCT y una segunda lectura de secuencias de moléculas QCT en un grupo de lecturas de secuencias QCT, del conjunto de grupos de lecturas de secuencias QCT, basándose en una similitud de secuencia de las regiones de variación entre una primera región de variación de la primera molécula QCT, y una segunda región de variación de la segunda molécula QCT, que satisface una primera condición, y

- para cada grupo de lecturas de secuencias QCT del conjunto de grupos de lecturas de secuencias QCT, determinar una asignación del grupo de lecturas de secuencias QCT a un identificador de muestra de un conjunto de identificadores de muestras que identifican el conjunto de muestras,

- en donde la determinación del parámetro de contaminación se basa además en las asignaciones de los grupos de lecturas de secuencias QCT a los identificadores de muestras del conjunto de identificadores de muestras.

7. El método de la reivindicación 5 o 6,

donde:

i) la determinación del parámetro de contaminación comprende:

- identificar un primer y un segundo grupo de lecturas de secuencias QCT que corresponden a una secuencia de región de variación compartida, en donde las asignaciones del primer y segundo grupos de lecturas de secuencias QCT son a identificadores de muestras distintos del conjunto de identificadores de muestras;

- generar una comparación de profundidades de lectura entre una primera profundidad de lectura asociada con el primer grupo de lecturas de secuencias QCT y una segunda profundidad de lectura asociada con el segundo grupo de lecturas de secuencias QCT; y

- basándose en la comparación de profundidades de lectura, determinar el parámetro de contaminación asociado con una muestra identificada por un identificador de muestra distinto de los identificadores de muestras distintos; o

ii) la determinación del parámetro de contaminación comprende:

- determinar una primera huella molecular asociada con la primera amplificación en un primer caso de la preparación de una biblioteca de secuenciación, basándose en el conjunto de grupos de lecturas de secuencias QCT;

- determinar una segunda huella molecular asociada con la segunda amplificación en un segundo caso de la preparación de una biblioteca de secuenciación, basándose en un conjunto adicional de grupos de lecturas de secuencias QCT; y

- basándose en una comparación entre la primera y la segunda huellas moleculares, determinar un parámetro de contaminación por arrastre que describe la contaminación por arrastre desde el primer caso al segundo caso.

8. El método de la reivindicación 6, en donde la agrupación de la primera y segunda lecturas de secuencias QCT en el grupo de lecturas de secuencias QCT se basa en la similitud de secuencia de la región de variación de menos de tres sustituciones puntuales, y se basa además en una profundidad de lectura asociada con el grupo de lecturas de secuencias QCT que satisface una segunda condición.

9. El método de la reivindicación 5,

- en donde cada región de variación del conjunto de moléculas QCT comprende una región de identificador molecular integrado (EMI) que comprende un conjunto de bases "N" variables, en donde cada base "N" se selecciona de una cualquiera de una base "A", una base "G", una base "T" y una base "C",

- en donde cada molécula QCT del conjunto de moléculas QCT comprende además una región EMI adicional que comprende un conjunto adicional de bases "N" variables, en donde la región EMI adicional está separada de la región EMI por una región de secuencia de la molécula QCT, en donde el conjunto de bases "N" variables y el conjunto adicional de bases "N" variables comprenden cada uno más de tres bases "N", y
- 5 • en donde la determinación del parámetro de contaminación se basa en el conjunto de grupos de lecturas de secuencias QCT derivados basándose en las regiones EMI y las regiones EMI adicionales del conjunto de moléculas QCT.
- 10. Un método de caracterización asociado con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, y el método comprende:
 - 10 • generar un conjunto de moléculas plantilla de control de calidad (QCT), y cada molécula QCT comprende:
 - una región asociada a una diana que coincide con una región de secuencia diana de una molécula diana; y
 - una región de variación que no coincide con una región correspondiente de la molécula diana, y la región de variación comprende un identificador molecular incrustado que tiene una secuencia de nucleótidos única;
 - 15 • generar una mezcla que comprende el conjunto de moléculas QCT y moléculas de ácido nucleico de una muestra;
 - coamplificar la mezcla;
 - secuenciar la mezcla coamplificada para generar datos de secuenciación;
 - 20 • determinar computacionalmente un conjunto de grupos de lecturas de secuencias QCT, mediante el uso de los datos de secuenciación, basándose en el número de regiones de variación distintas del conjunto de moléculas QCT identificadas en los datos de secuenciación,
 - determinar computacionalmente el recuento de moléculas QCT que estaban en la mezcla antes de la coamplificación basándose en los grupos de lecturas de secuencias QCT; y
 - 25 • basándose en el recuento de moléculas QCT, determinar la abundancia absoluta de la molécula diana asociada con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación.
 - 11. El método de la reivindicación 10,
 - en donde cada molécula QCT comprende una primera región identificadora de QCT compartida entre el conjunto de moléculas QCT y adaptada para identificar la molécula QCT,
 - 30 • en donde el método comprende además generar un conjunto de moléculas QCT adicionales, y cada molécula QCT adicional comprende una segunda región identificadora de QCT compartida entre el conjunto de moléculas QCT adicionales y adaptada para identificar la molécula QCT adicional; y
 - en donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT comprende determinar el conjunto de grupos de lecturas de secuencias QCT basándose en la primera y la segunda regiones identificadoras de QCT.
 - 35 El método de la reivindicación 11,
 - en donde el conjunto de moléculas QCT está adaptado para su utilización en una primera etapa de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación,
 - en donde el conjunto de moléculas QCT adicionales está adaptado para su utilización en una segunda etapa de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación,
 - 40 • en donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT comprende:
 - determinar un primer subconjunto del conjunto de grupos de lecturas de secuencias QCT, en donde el primer subconjunto corresponde a la primera región identificadora de QCT y está asociado con la primera etapa, y
 - 45 ◦ determinar un segundo subconjunto del conjunto de grupos de lecturas de secuencias QCT, en donde el segundo subconjunto corresponde a la segunda región identificadora de QCT y está asociado con la segunda etapa; y

- en donde la determinación del parámetro relacionado con la secuenciación comprende determinar un parámetro de seguimiento de muestras asociado con la pérdida de muestra, basándose en el primer y segundo subconjuntos del conjunto de grupos de lecturas de secuencias QCT.

13. El método de la reivindicación 11, en donde:

5 i) la determinación del parámetro relacionado con la secuenciación comprende:

- determinar un primer recuento absoluto y un segundo recuento absoluto correspondientes al conjunto de moléculas QCT y al conjunto de moléculas QCT adicionales, respectivamente, basándose en el conjunto de grupos de lecturas de secuencias QCT, y

10 • determinar al menos uno de un parámetro de error de pipeta y un parámetro de error de cuantificación basándose en el primer y segundo recuentos absolutos; o

ii) la determinación del parámetro relacionado con la secuenciación comprende:

- identificar las lecturas de secuencias QCT que no están asignadas a un grupo de lecturas de secuencias QCT del conjunto de grupos de lecturas de secuencias QCT; y

15 • determinar al menos uno de una tasa de error de secuenciación y una tasa de error de la polimerasa a partir de un número de lecturas de secuencias QCT que no están asignadas y un número total de lecturas de secuencias QCT.

14. El método de la reivindicación 11,

20 • en donde la región de variación de cada molécula QCT comprende una primera región identificadora molecular (EMI) incrustada separada de una segunda región EMI por al menos la primera región identificadora de QCT,

• en donde cada molécula QCT adicional comprende una primera región EMI adicional separada de una segunda región EMI adicional por al menos la segunda región identificadora de QCT,

25 • en donde la primera región EMI, la segunda región EMI, la primera región EMI adicional y la segunda región EMI adicional comprenden un conjunto de bases "N" variables, y en donde cada base "N" se selecciona de cualquiera de una base "A", una base "G", una base "T" y una base "C",

• en donde la determinación computacional del conjunto de grupos de lecturas de secuencias QCT comprende determinar el conjunto de grupos de lecturas de secuencias QCT basándose en la primera y segunda regiones identificadoras de QCT, y en la primera región EMI, la segunda región EMI, la primera región EMI adicional y la segunda región EMI adicional,

30 • en donde, opcionalmente, para cada molécula QCT, la secuencia de la molécula QCT correspondiente se caracteriza por una similitud de secuencia completa con una primera plantilla de secuencia de la diana biológica excepto por la primera región identificadora de QCT, la primera región EMI y la segunda región EMI, y

35 • en donde, opcionalmente, para cada molécula QCT adicional, la secuencia de molécula QCT adicional correspondiente se caracteriza por una similitud de secuencia completa con una segunda plantilla de secuencia excepto por la segunda región identificadora de QCT, la primera región EMI adicional y la segunda región EMI adicional.

15. El método de la reivindicación 14,

40 en donde, para cada molécula QCT, la secuencia de la molécula QCT correspondiente se caracteriza por una similitud de secuencia completa con una primera plantilla de secuencia de la diana biológica excepto por la primera región identificadora de QCT, la primera región EMI y la segunda región EMI; en donde, para cada molécula de QCT adicional, la secuencia de molécula QCT adicional correspondiente se caracteriza por una similitud de secuencia completa con una segunda plantilla de secuencia excepto por la segunda región identificadora de QCT, la primera región EMI adicional y la segunda región EMI adicional.

45 16. El método de la reivindicación 10,

en donde la determinación del conjunto de grupos de lecturas de secuencias QCT comprende determinar un subconjunto filtrado de grupos de lecturas de secuencias QCT basándose en las profundidades de lectura que corresponden al subconjunto filtrado de grupos de lecturas de secuencias QCT,

en donde la determinación del recuento de moléculas diana comprende:

determinar el recuento de moléculas QCT basándose en el subconjunto filtrado de grupos de lecturas de secuencias QCT;

determinar una relación de factor de corrección basándose en el recuento de moléculas QCT y las lecturas de secuencias de moléculas QCT; y

5 determinar el recuento de moléculas diana basándose en la relación del factor de corrección y las lecturas de la secuencia de la molécula diana derivadas de la secuenciación, las lecturas de la secuencia de la molécula diana asociadas con la diana biológica.

10 17. El método de la reivindicación 16, que comprende además determinar de forma adaptativa un umbral de profundidad de lectura basándose en las características de distribución de la profundidad de lectura para las lecturas de secuencias de moléculas QCT, en donde la determinación del subconjunto filtrado de grupos de lecturas de secuencias QCT comprende determinar el subconjunto filtrado basándose en la satisfacción del umbral de profundidad de lectura determinado adaptativamente por las profundidades de lectura.

15 18. El método de la reivindicación 16, en donde cada profundidad de lectura de las profundidades de lectura corresponde a más de veinte lecturas para el grupo de lecturas de secuencias QCT correspondiente del subconjunto filtrado de grupos de lecturas de secuencias QCT.

19. El método de la reivindicación 10, en donde la determinación del recuento de moléculas diana comprende determinar el recuento de moléculas diana para facilitar el diagnóstico asociado con al menos una de las pruebas prenatales no invasivas y biopsias líquidas.

20 20. El método de la reivindicación 10, en donde la determinación del parámetro relacionado con la secuenciación comprende determinar un parámetro de contaminación que comprende al menos uno de: un parámetro de contaminación cruzada que describe la contaminación cruzada entre compartimentos de muestras asociados con al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, un parámetro de contaminación por arrastre que describe la contaminación por arrastre a través de una pluralidad de casos de al menos una de la preparación de una biblioteca de secuenciación y la secuenciación, y un parámetro de contaminación por saltos de índices que describe la contaminación por saltos de índices asociada con cebadores con saltos de índices.

21. El método de la reivindicación 20,
en donde el conjunto de moléculas QCT está adaptado para la secuenciación,
en donde la generación del conjunto de moléculas QCT comprende:
amplificar un primer subconjunto de moléculas QCT del conjunto de moléculas QCT; y
30 amplificar un segundo subconjunto de moléculas QCT del conjunto de moléculas QCT,
en donde las lecturas de secuenciación de las moléculas QCT derivan de la secuenciación correspondiente a:

la mezcla de QCT generada basándose en el primer subconjunto de moléculas QCT y la muestra que comprende la diana biológica, y

35 una mezcla de QCT adicional generada basándose en el segundo subconjunto de moléculas QCT y una muestra adicional que comprende la diana biológica, en donde la muestra y la muestra adicional corresponden respectivamente a un primer compartimento de muestras y un segundo compartimento de muestras de los compartimentos de muestras.

40 22. El método de cualquiera de las reivindicaciones 1, 5 y 10, en donde la secuencia de nucleótidos única es una secuencia de nucleótidos aleatoria.

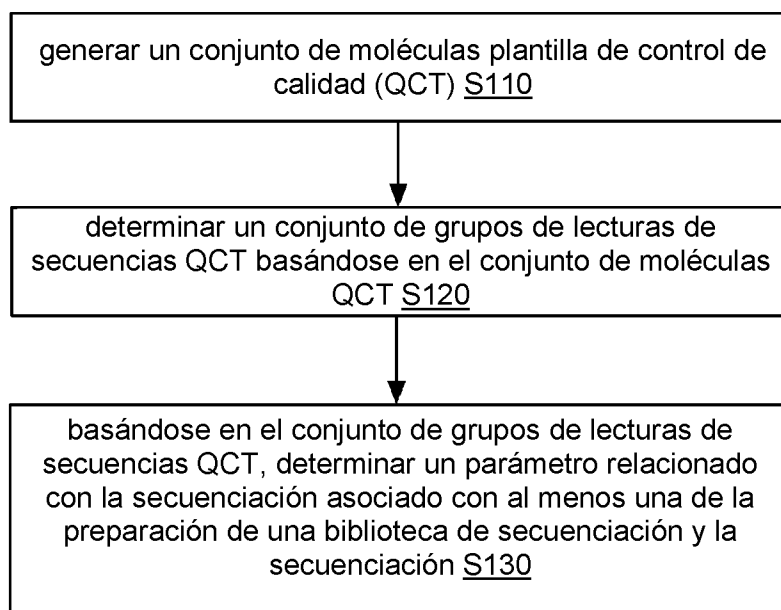


FIGURA 1A

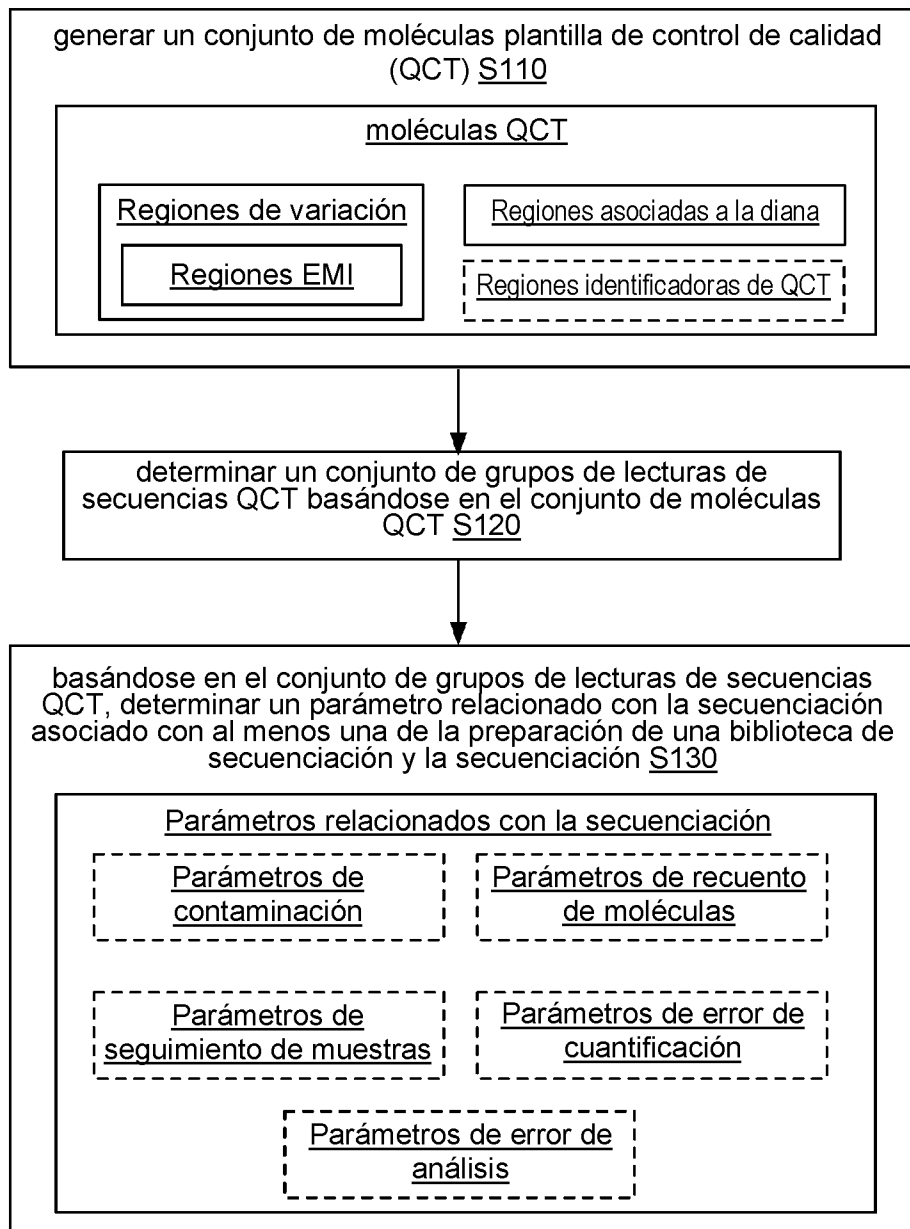


FIGURA 1B

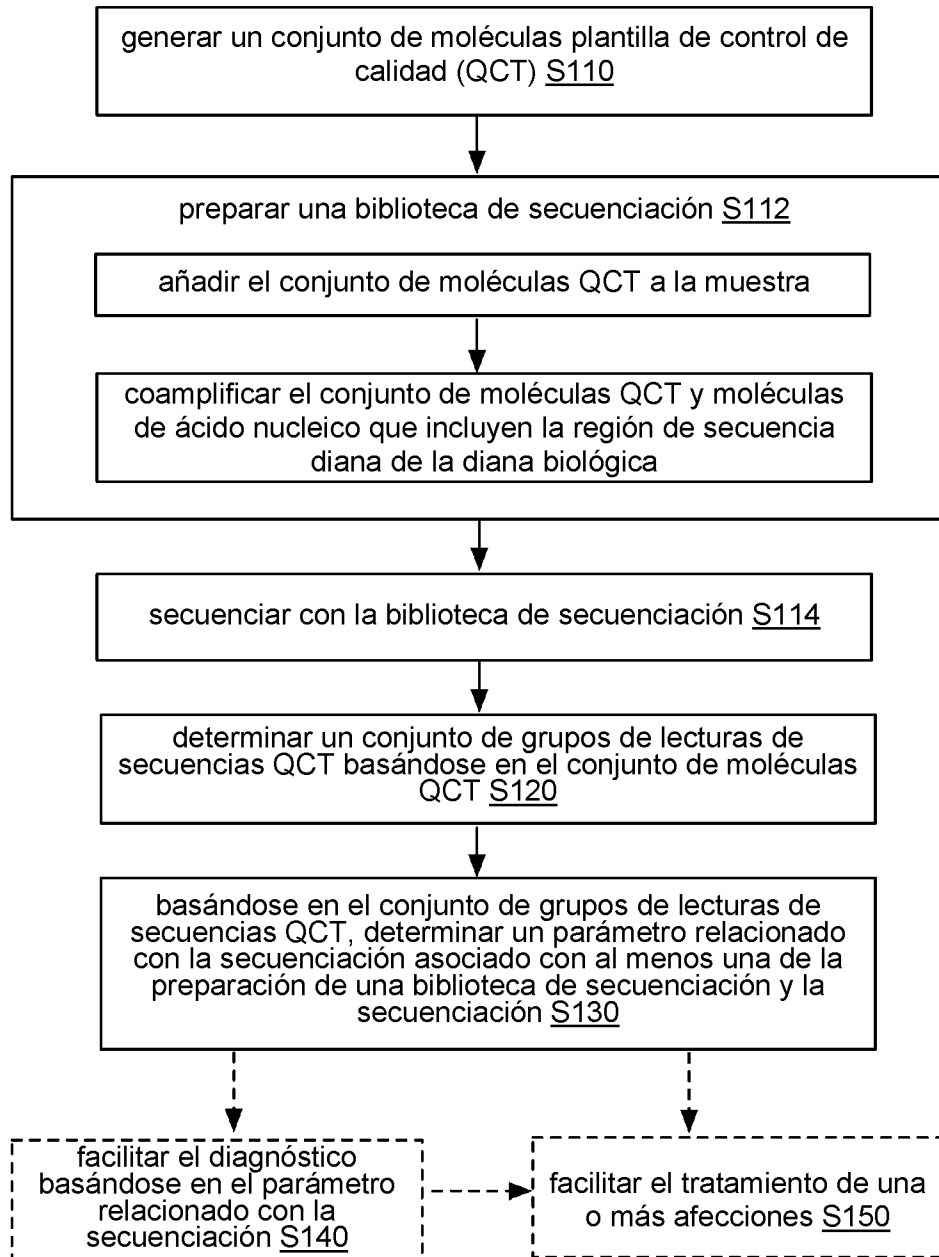


FIGURA 1C

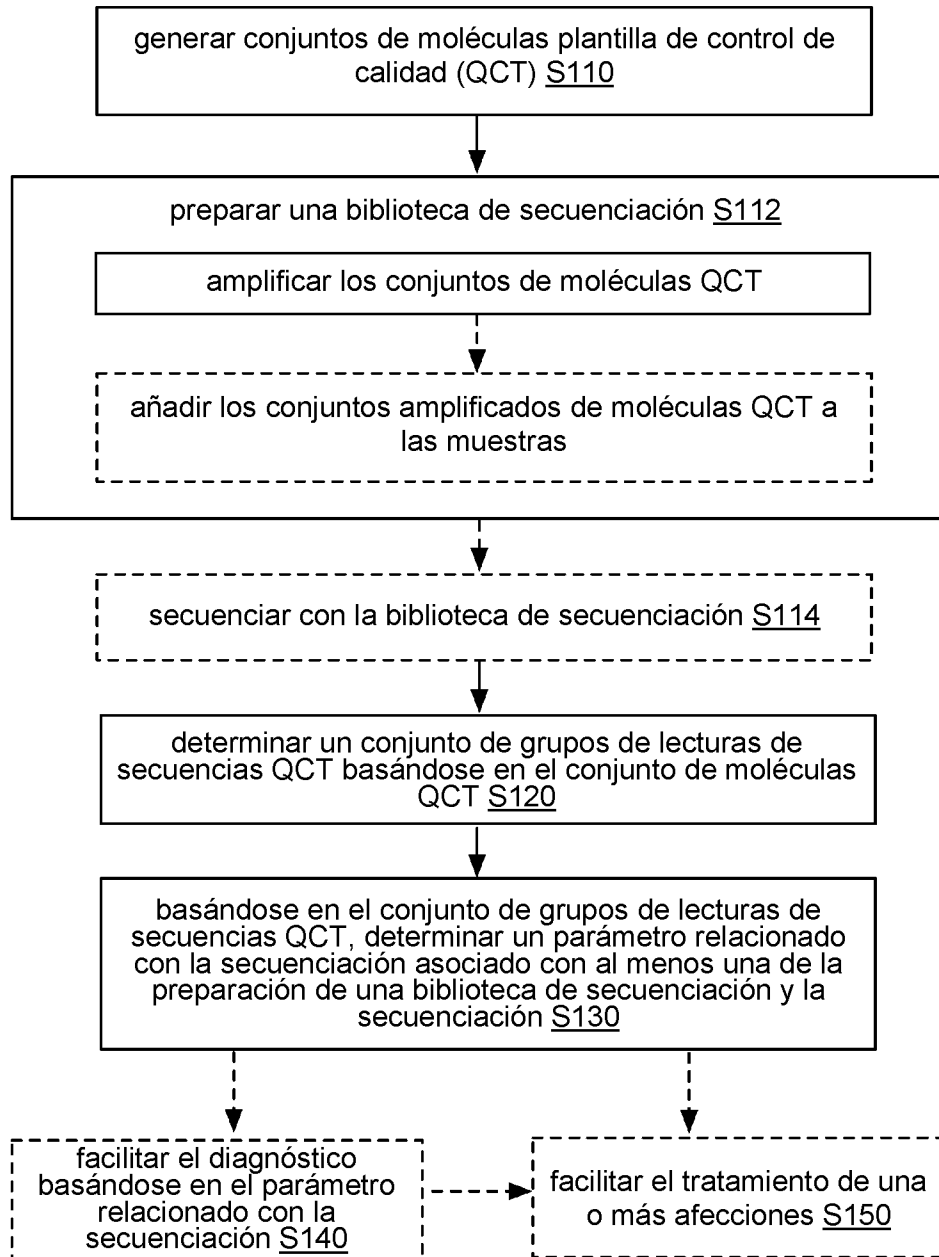


FIGURA 1D

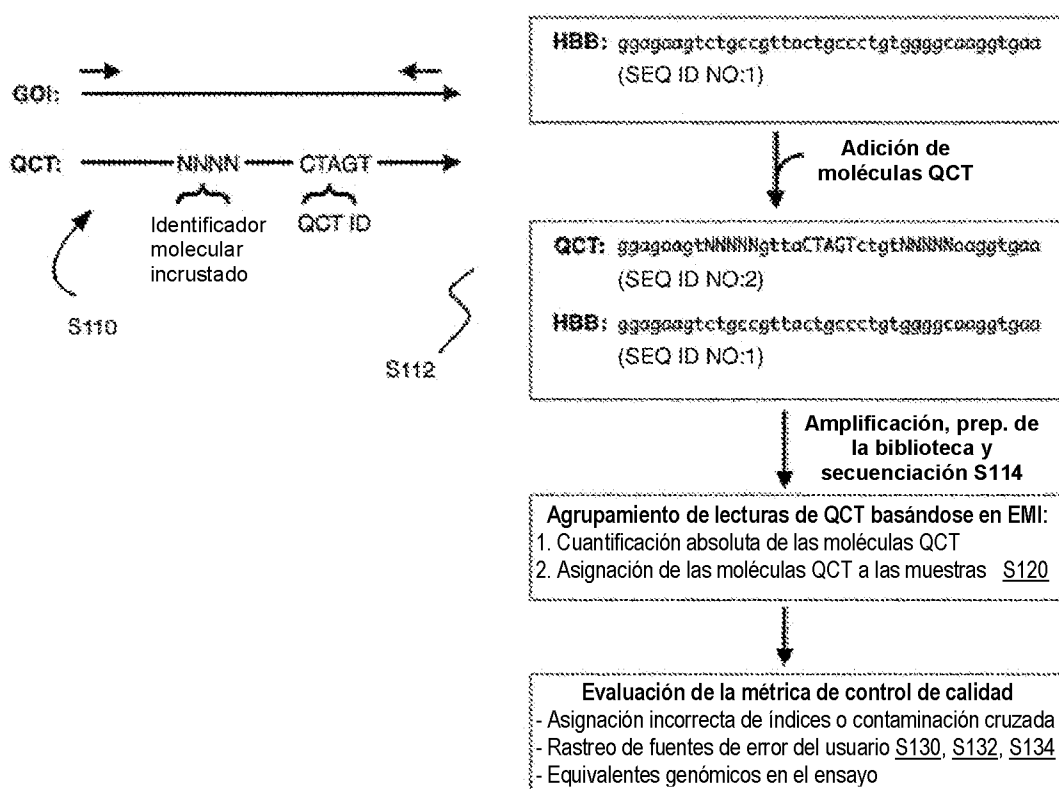


FIGURA 2

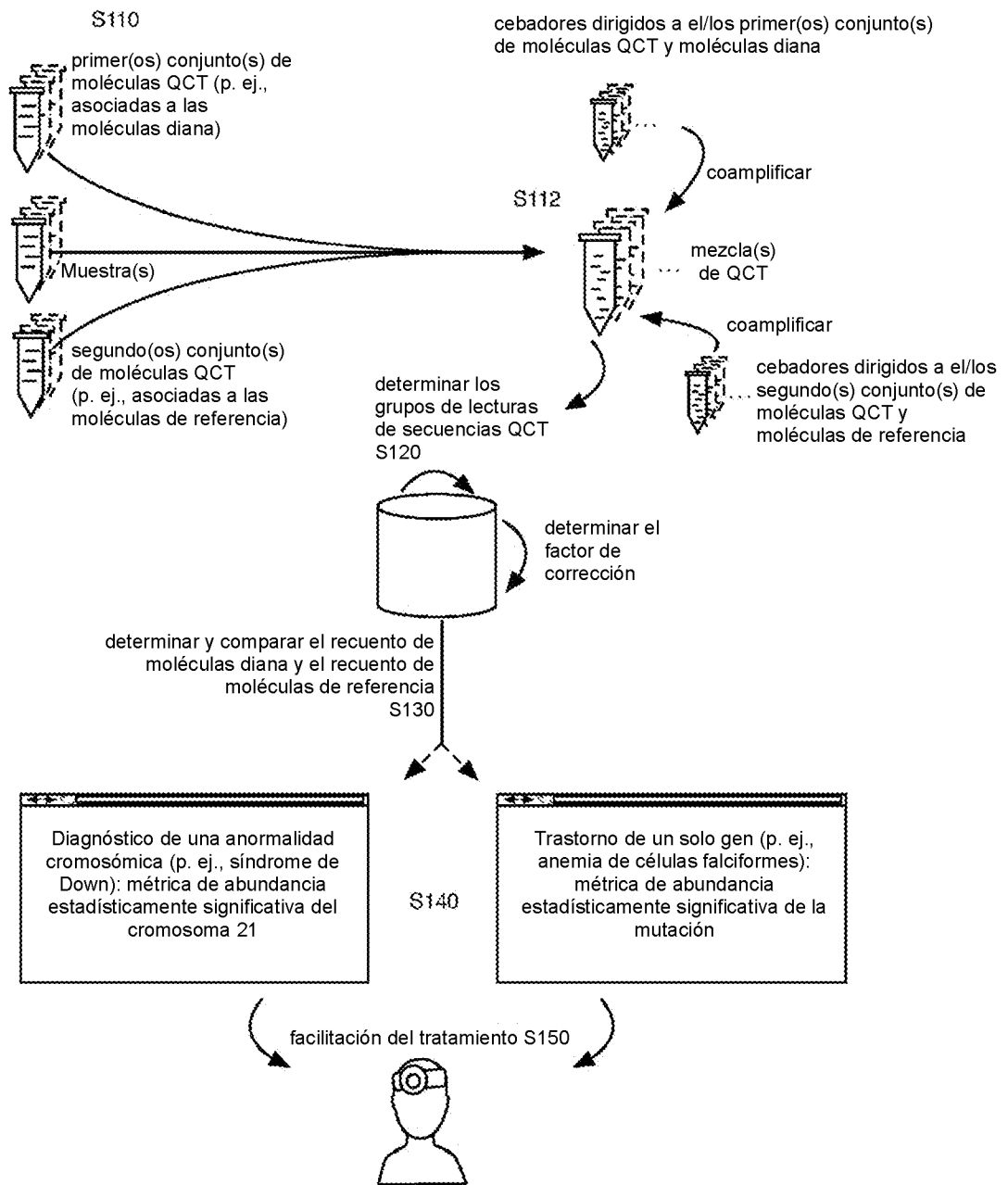


FIGURA 3

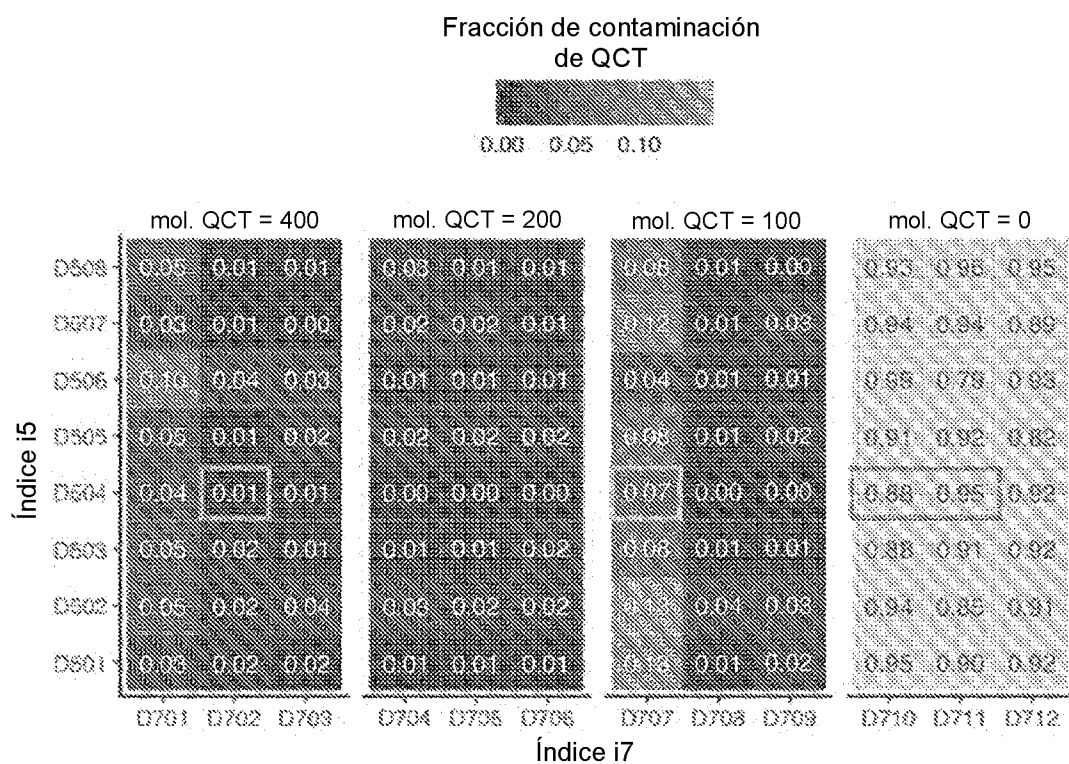


FIGURA 4A

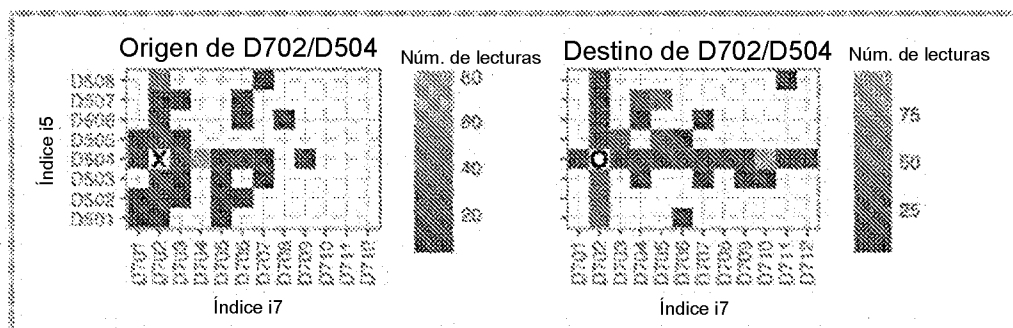


FIGURA 4B

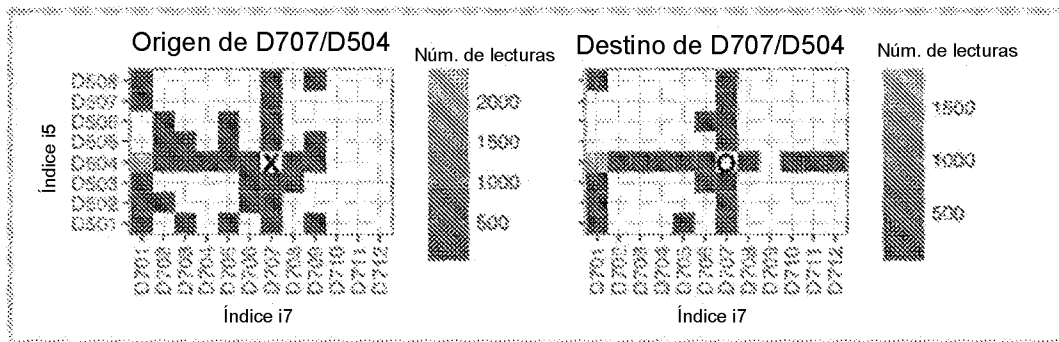


FIGURA 4C

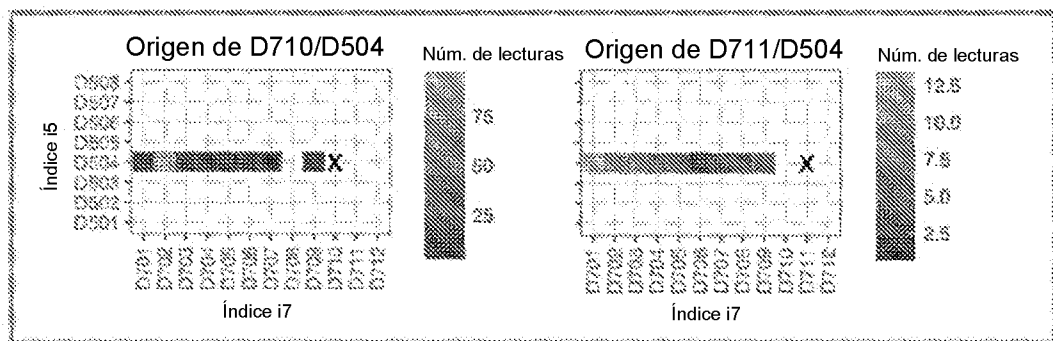


FIGURA 4D

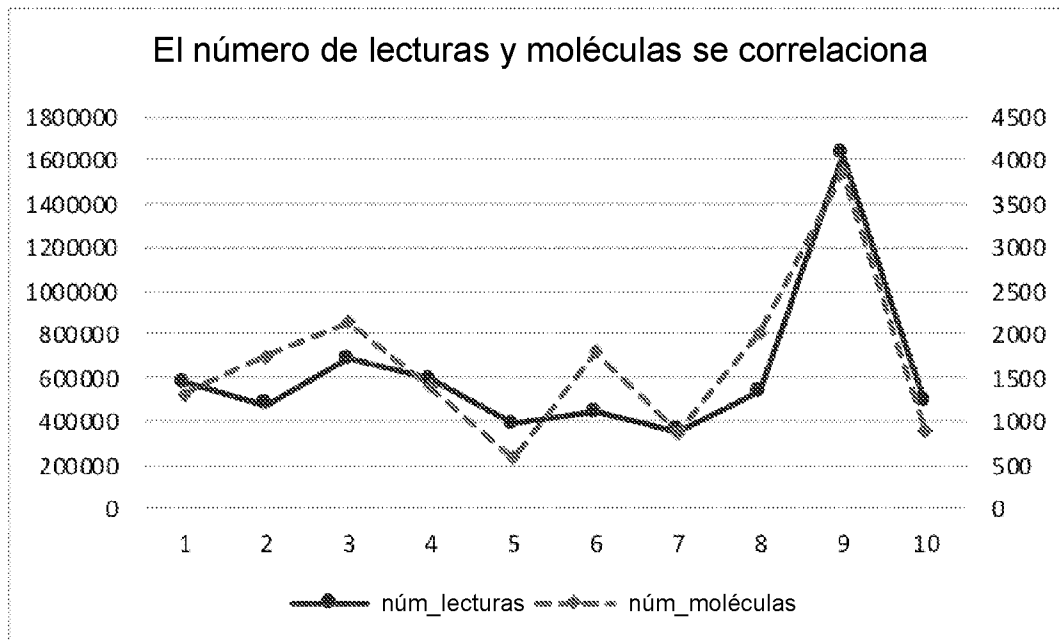


FIGURA 5A

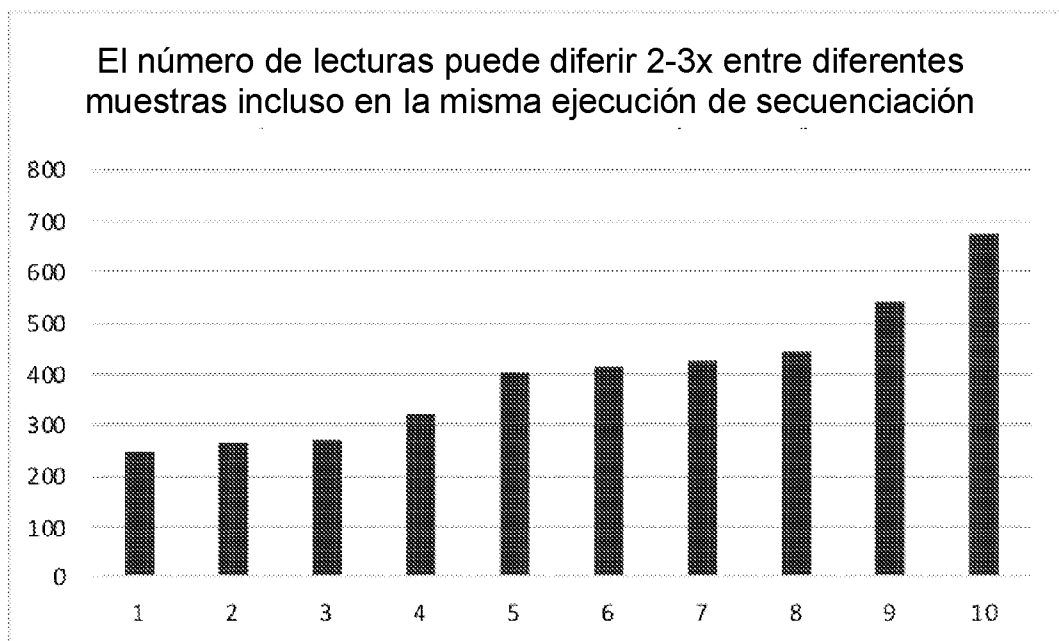


FIGURA 5B

	sample_id	num_seqs	contam_frac	collision_frac	contam_collision_frac	ident_frac	reads_per_qctmol
X	A01	X 8	0.633	0.000	X 0.633	X 0.721	X 2.5
X	A02	✓ 151	0.048	0.029	X 0.078	X 0.948	X 21.9
X	A03	X 45	0.033	0.607	X 0.640	✓ 0.968	X 27.2
X	A04	X 45	0.086	0.403	X 0.488	X 0.594	X 9.2
X	A05	✓ 263	0.005	0.012	✓ 0.017	X 0.839	✓ 58.8
✓	A06	✓ 131	0.005	0.000	✓ 0.005	✓ 0.977	✓ 146.2
✓	E31	✓ 221	0.000	0.005	✓ 0.005	✓ 0.984	✓ 872.8
✓	E32	✓ 212	0.001	0.000	✓ 0.001	✓ 0.985	✓ 409.1
✓	E33	✓ 188	0.002	0.000	✓ 0.002	✓ 0.984	✓ 703.9
✓	E34	✓ 193	0.000	0.000	✓ 0.000	✓ 0.985	✓ 1325.8
✓	E35	✓ 208	0.000	0.000	✓ 0.000	✓ 0.985	✓ 884.0
	PASAN EL FILTRO	>100			<0.02	>0.95	>30

FIGURA 6

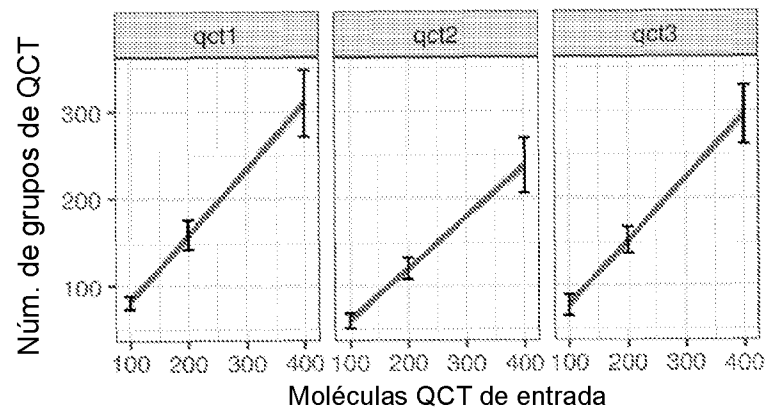


FIGURA 7A

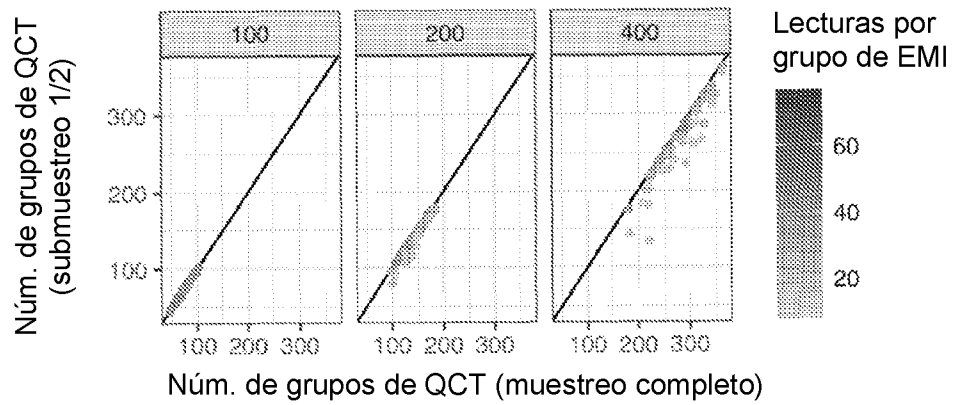


FIGURA 7B

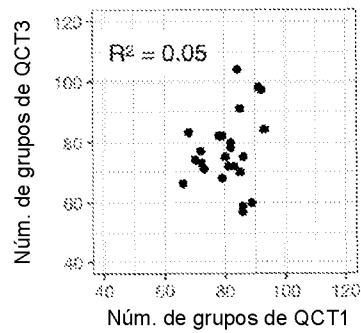


FIGURA 7C

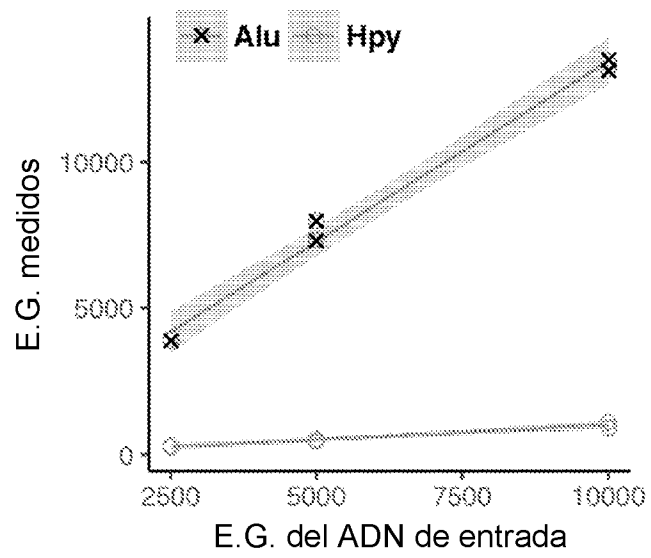


FIGURA 8A

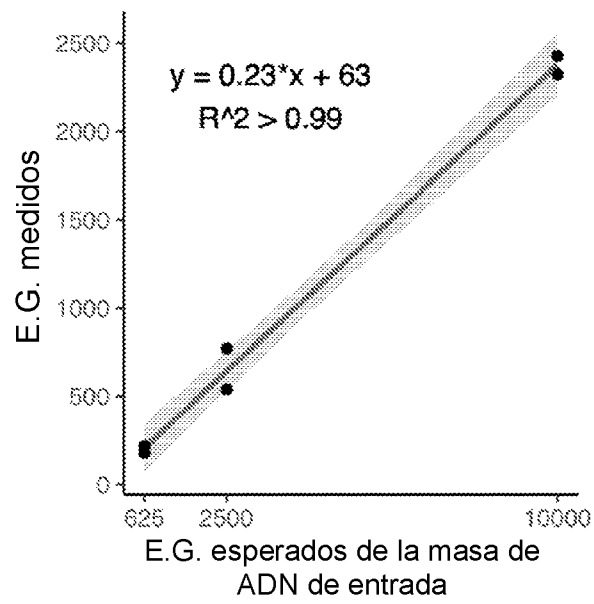


FIGURA 8B

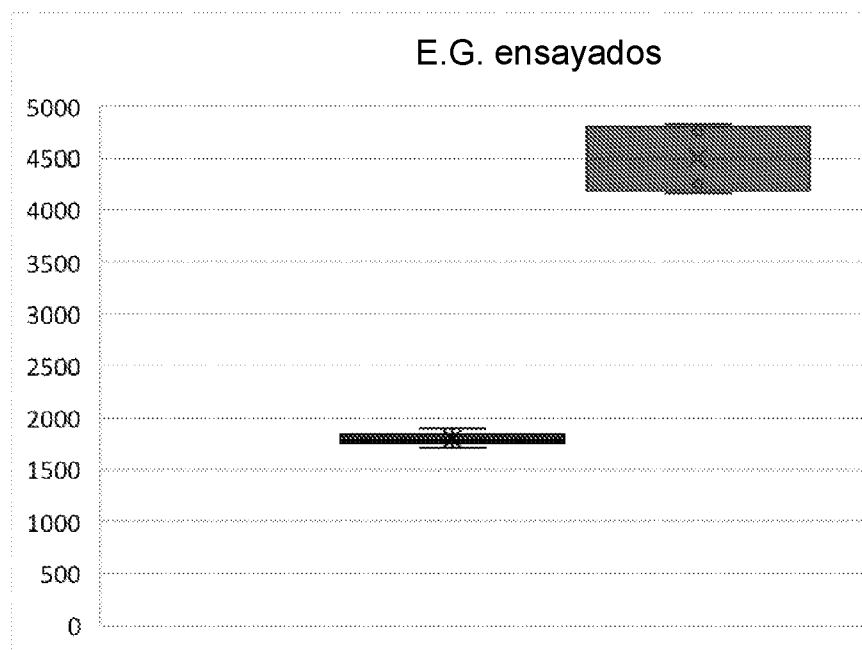


FIGURA 9

Tubo A de PCR - Muestra A				Tubo B de PCR - Muestra B			
núm_lecturas	Grupo de EMI	SEQ	ID NO.	núm_lecturas	Grupo de EMI	SEQ	ID NO.
133	CATTACCGCT	3		109	GAGCTATAGC	33	
122	AGATCCGATT	4		108	CTTATCTTTA	34	
120	GGATCATCGG	5		104	AGATACCGTT	35	
119	TTACGGGTAA	6		103	CAGATGGTAT	36	
112	TGAGTTATGG	7		102	GGAGTCCTGG	37	
112	TAAATAATGT	8		102	GCCTAAAAGA	38	
111	GTCCAACGTA	9		93	TTCCGCAGAA	31	
107	AGAGTCTTAT	10		92	GGGGTGGTGT	39	
106	TACACTCCTT	11		91	TAATCGTCTA	40	
102	TGTCATCTTA	12		91	GTACAGCGCA	41	
102	GTATTGCGAT	13		73	TTTGTGTTAT	42	
96	AATGATTAAT	14		67	GTTTCAACTT	43	
93	TATATGGTCA	15		54	TGTTAAAGAT	44	
81	GGTGCTATAG	16		48	GCTCCTTCAT	45	
69	GTTTAACTGG	17		47	GTGCGCGGTT	46	
69	ACGTGATGTA	18		46	TCTATCTAAG	47	
62	TTTCATCGGT	19		43	TCCTGTAAAC	48	
60	CCTATCCGAT	20		38	AATAGAATGG	49	
56	ATAGTAACAT	21		2	TCATCGTGCG	50	
54	GGACTATGGT	22		2	TCACGAGTAC	51	
52	AATATAGTCA	23		2	TAGACAAGCG	52	
43	CTTAGGTGGC	24		2	CTTGTGTTAT	53	
2	CTGCCCTGGC	25		2	CGTGCTCTCT	54	
2	CTGCCGGCGC	26		2	CGATGGATGA	55	
2	ACTTTCCTGC	27		2	CGAAATTGCG	56	
1	TTTCTCTTTA	28		1	TTTTCCAGCT	57	
1	TTTATGGGCA	29		1	TGGATATTCC	58	
1	TTGTAGGTGC	30		1	TATTATTTAA	59	
1	TTCCGGCAGAA	31		1	TATCCGACTA	60	
1	TTCAGAATAC	32		1	TAGTCTTGTC	61	

FIGURA 10

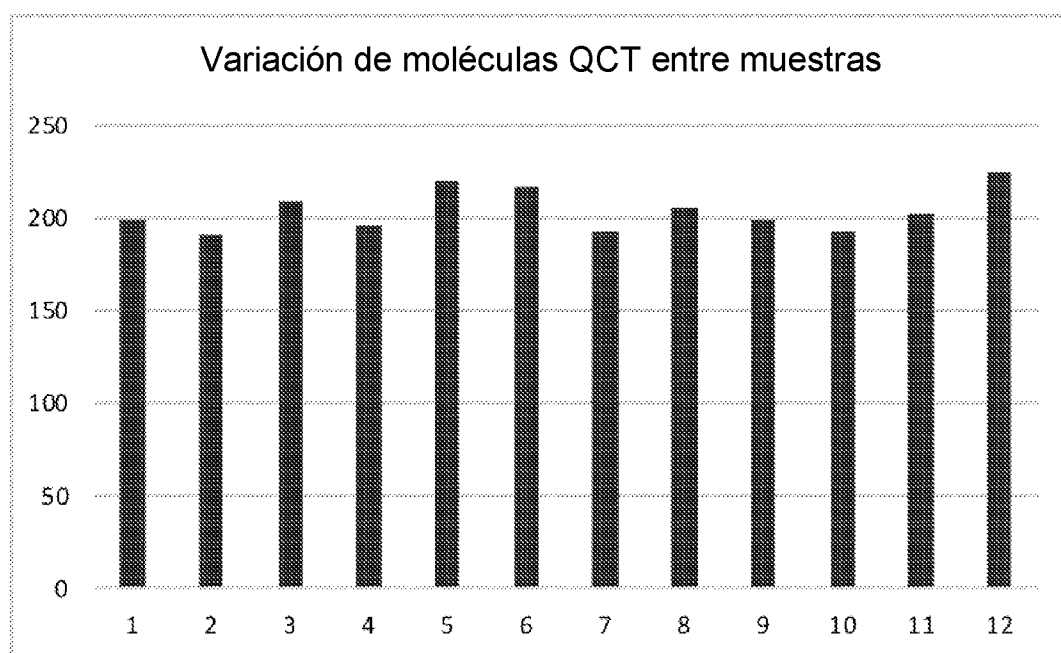


FIGURA 11

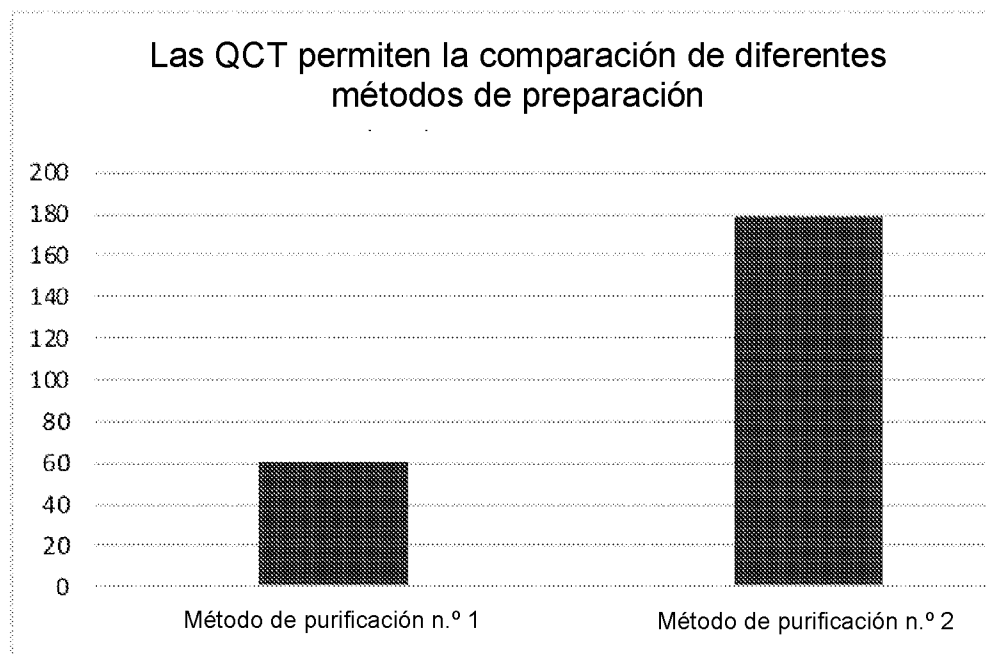


FIGURA 12

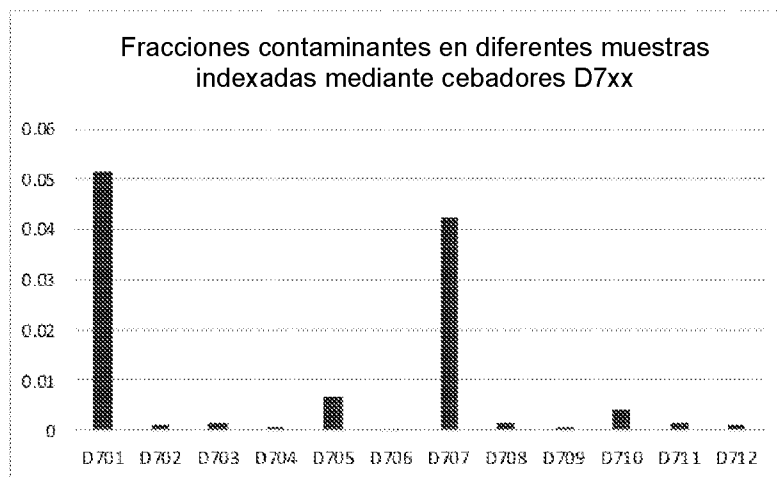


FIGURA 13A

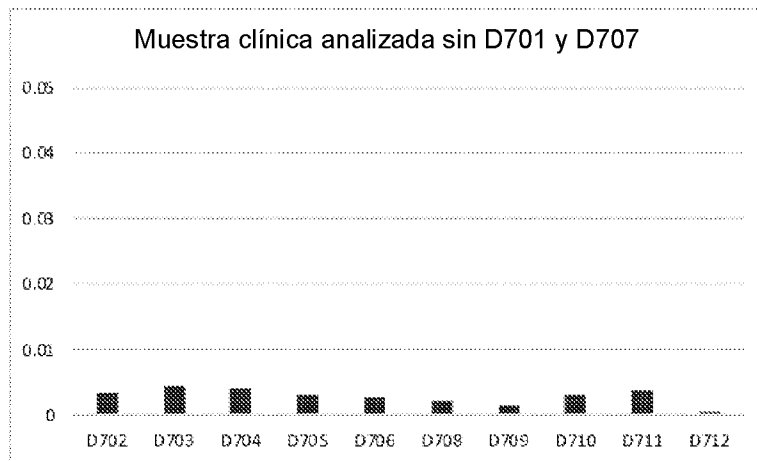


FIGURA 13B

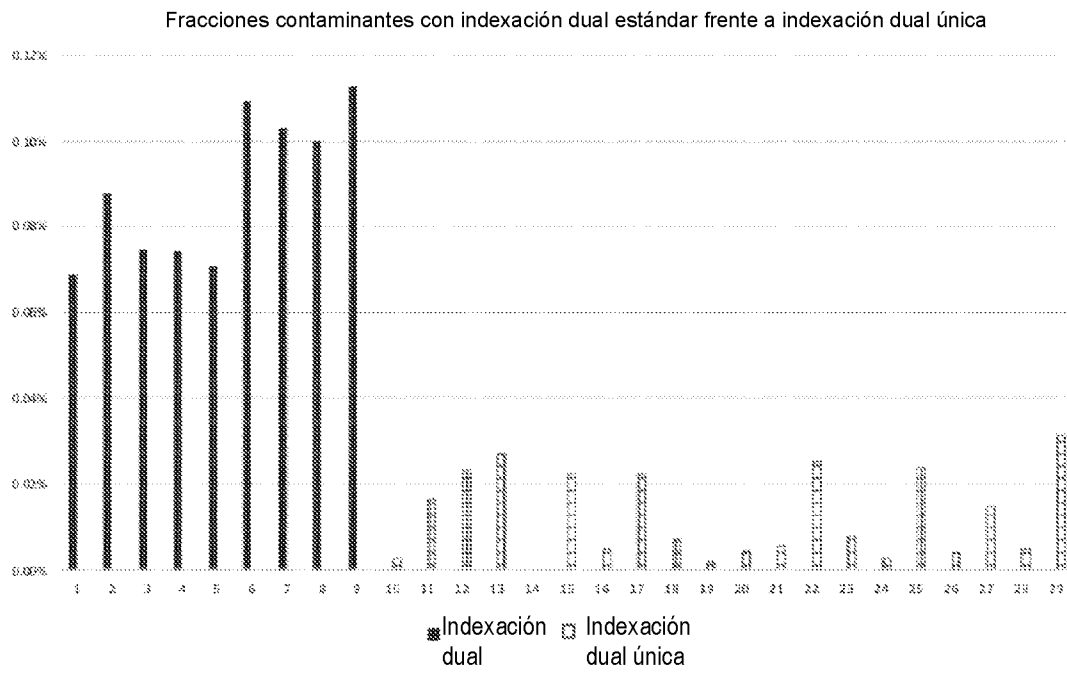


FIGURA 14

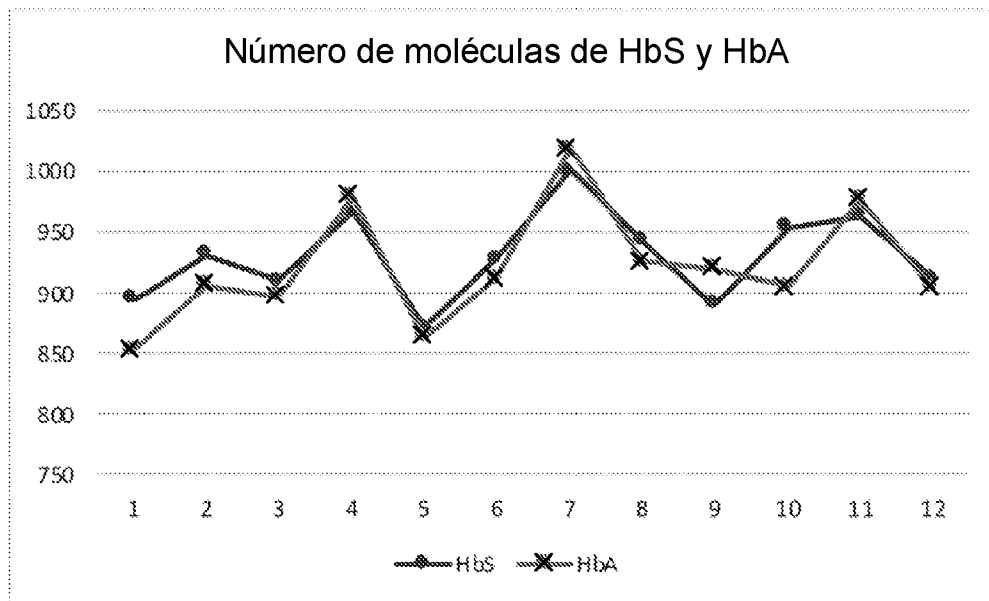


FIGURE 15A

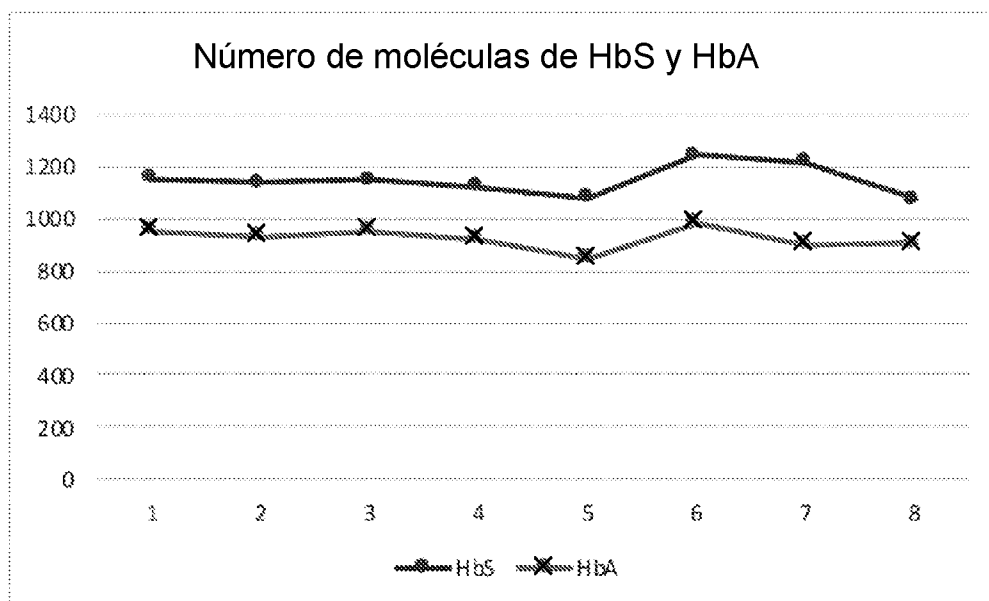


FIGURA 15B

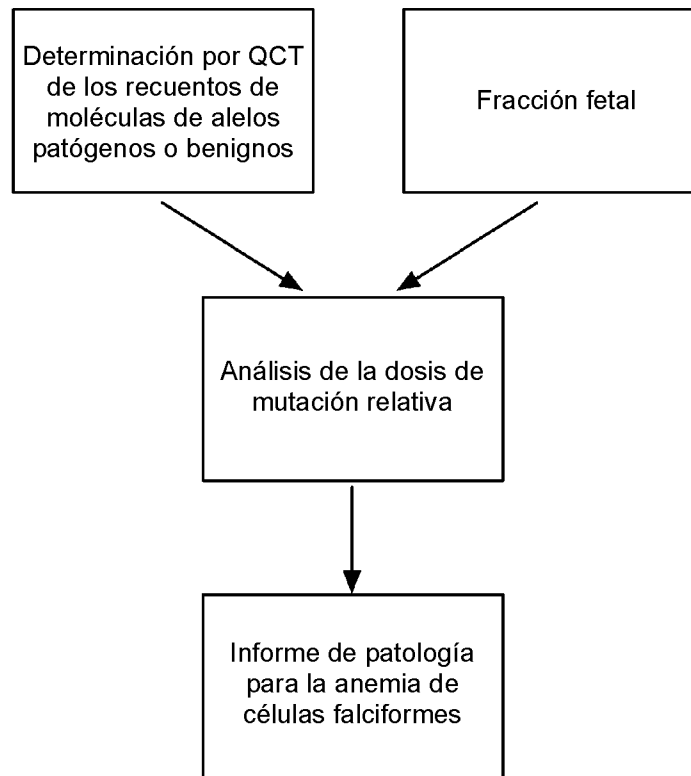


FIGURA 15C

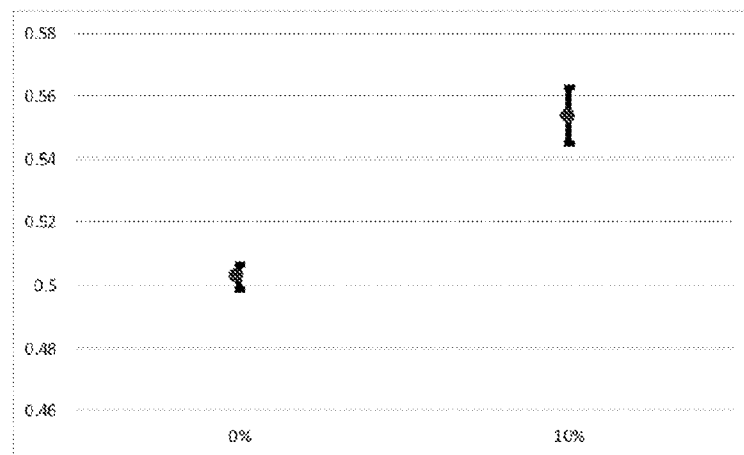


FIGURA 15D

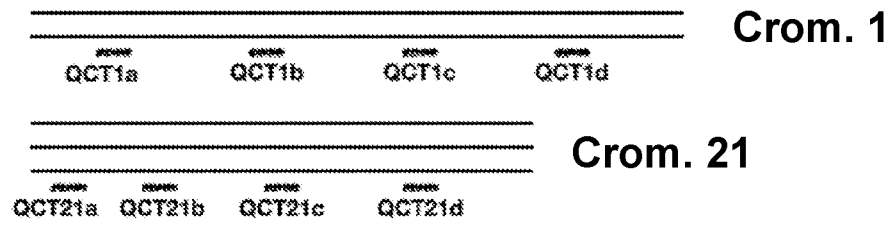


FIGURA 16A

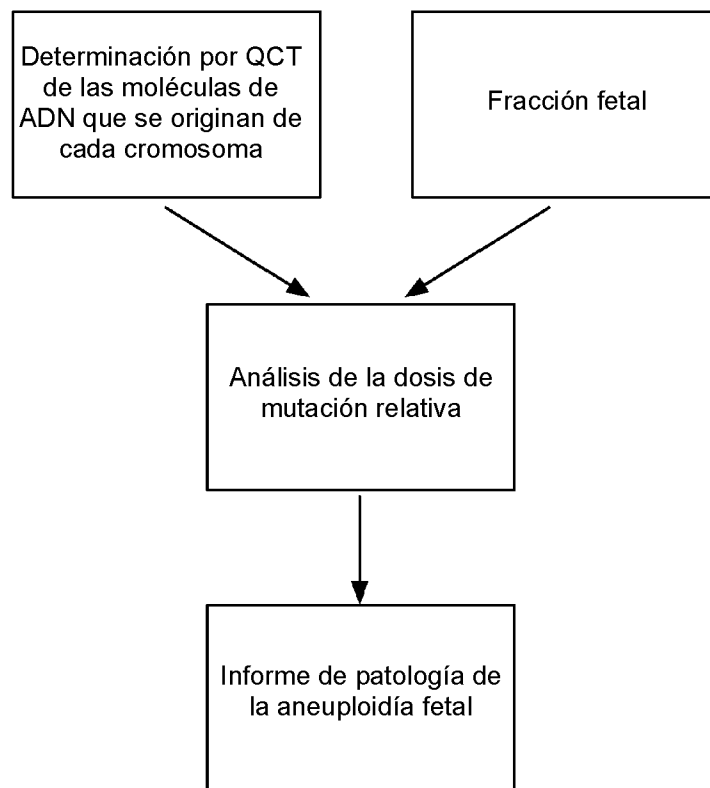


FIGURA 16B