

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 018 032**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

C12Q 1/6827 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.07.2021** **PCT/US2021/043780**

87 Fecha y número de publicación internacional: **03.02.2022** **WO22026761**

96 Fecha de presentación y número de la solicitud europea: **29.07.2021** **E 21758891 (2)**

97 Fecha y número de publicación de la concesión europea: **19.02.2025** **EP 4189111**

54 Título: **Métodos para aislamiento de ADN libre de células**

30 Prioridad:

30.07.2020 US 202063059029 P

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

14.05.2025

73 Titular/es:

GUARDANT HEALTH, INC. (100.00%)
3100 Hanover Street
Palo Alto, CA 94304, US

72 Inventor/es:

JAIMOVICH, ARIEL;
DUENWALD, SVEN JEFFREY;
GRAUMAN, PETER VILEM;
HE, YUPENG;
EID, CHARBEL SAID y
AXELROD, HALEY DARA

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 3 018 032 T3

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

DESCRIPCIÓN

Métodos para aislamiento de ADN libre de células

ANTECEDENTES

[0001] El cáncer es responsable de millones de muertes por año en todo el mundo. La detección temprana del cáncer puede dar como resultado mejores resultados porque el cáncer en etapa temprana tiende a ser más susceptible al tratamiento.

[0002] El crecimiento celular controlado inadecuadamente es un sello distintivo del cáncer que generalmente resulta de una acumulación de cambios genéticos y epigenéticos, tales como variaciones en el número de copias (CNV), variaciones de un solo nucleótido (SNV), fusiones, inserciones y/o deleciones de genes (indeles), las variaciones epigenéticas incluyen la 5-metilación de la citosina (5-metilcitosina) y la asociación del ADN con proteínas de la cromatina y factores de transcripción.

[0003] Las biopsias representan un método tradicional para detectar o diagnosticar el cáncer en el que se extraen células o tejido de un posible sitio de cáncer y se analizan para determinar características fenotípicas y/o genotípicas relevantes. Las biopsias tienen el inconveniente de ser invasivas. El análisis de biopsias también puede llevar mucho tiempo y requerir atención individual por parte de personal médico altamente capacitado, como patólogos.

[0004] La detección de cáncer basada en el análisis de ADN de células cancerosas es una alternativa interesante. Por ejemplo, se ha observado que el ADN de células cancerosas se libera en fluidos corporales. Como tal, el ADN de células cancerosas se puede detectar y analizar de forma no invasiva (quizás requiriendo, por ejemplo, solo una extracción de sangre ("biopsia líquida")). Alternativamente, el ADN se puede obtener de varios tipos de muestras de tejido. Sin embargo, ha sido un desafío desarrollar métodos precisos y sensibles para analizar el ADN de células cancerosas dada la baja concentración y heterogeneidad de dicho ADN, por ejemplo, en biopsias líquidas y en ciertos tipos de muestras de tejido. Además, es deseable limitar la cantidad de secuenciación requerida para proporcionar datos suficientes para el análisis, de modo que los costos sean limitados y/o se maximice el rendimiento. Aislar las fracciones de ADN y prepararlo para el análisis de secuencia es una parte importante de este proceso. En consecuencia, existe una necesidad de métodos y composiciones mejorados para aislar ADN, por ejemplo, para su uso en métodos de detección de cáncer tales como biopsias líquidas.

RESUMEN

[0005] La presente divulgación proporciona composiciones y métodos para aislar ADN, como ADN libre de células o ADN de una muestra de tejido. La presente divulgación se basa en parte en la siguiente realización. Si bien puede ser beneficioso aislar ADN de modo de capturar dos conjuntos de regiones objetivo (como un conjunto de regiones objetivo de secuencia variable y un conjunto de regiones objetivo epigenéticas), ciertos tipos de regiones objetivo epigenéticas que están hipometiladas pueden ser menos informativos con respecto al estado del cáncer en relación con cuando dichas regiones están hipermetiladas, y/o en relación con las regiones objetivo de secuencia variable en general. Como tal, los métodos descritos en este documento pueden reducir el costo y/o ahorrar ancho de banda de secuenciación con un impacto limitado en el rendimiento del ensayo al reducir la cantidad de secuenciación de regiones objetivo epigenéticas hipometiladas. La invención se establece en las reivindicaciones adjuntas.

[0006] Los diversos pasos de los métodos divulgados en este documento, o los pasos llevados a cabo por los sistemas divulgados en este documento, pueden llevarse a cabo al mismo tiempo o en momentos diferentes, y/o en la misma ubicación geográfica o en ubicaciones geográficas diferentes, por ejemplo, países. Los diversos pasos de los métodos divulgados en este documento pueden ser realizados por la misma persona o por personas diferentes.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0007] Los dibujos adjuntos, que se incorporan y constituyen una parte de esta especificación, ilustran ciertas formas de realización y, junto con la descripción escrita, sirven para explicar ciertos principios de los métodos, medios legibles por computadora y sistemas divulgados en este documento. La descripción proporcionada en este documento se entiende mejor cuando se lee junto con los dibujos adjuntos que se incluyen a modo de ejemplo y no a modo de limitación. Se entenderá que los números de referencia similares identifican componentes similares en todos los dibujos, a menos que el contexto indique lo contrario. También se entenderá que algunas o todas las figuras pueden ser representaciones esquemáticas con fines ilustrativos y no necesariamente representan los tamaños relativos reales o las ubicaciones de los elementos mostrados.

La **FIG. 1A** muestra una descripción general de una metodología ejemplar para particionar, agrupar y capturar regiones objetivo de secuencia variable (SVTR) y regiones objetivo epigenéticas (ETR).

La **FIG. 1B** muestra una descripción general de una metodología ejemplar para particionar, agrupar y capturar conjuntos primero y segundo de regiones objetivo.

La **FIG. 2** es un diagrama esquemático de un ejemplo de un sistema adecuado para su uso con algunas formas de realización de la divulgación.

La **FIG. 3A** es un gráfico del recuento total de lecturas en miles de millones versus la entrada en nanogramos para las muestras indicadas descritas en el Ejemplo 2.

La **FIG. 3B** es un gráfico de diversidad medida como número de moléculas únicas en el panel epigenómico versus la entrada en nanogramos para las muestras indicadas descritas en el Ejemplo 2.

La **FIG. 3C** es un gráfico que muestra la correlación de Theta Exp medida a partir de datos escalados al 20 % de una muestra dividida y Theta Exp medida a partir de datos no escalados de muestras de control como se describe en el Ejemplo 2.

La **FIG. 4A** muestra valores de fracción de alelos mutantes para las muestras divididas indicadas medidas como se describe en el Ejemplo 3.

La **FIG. 4B** muestra los valores de la fracción de alelos mutantes para las muestras de control indicadas medidas como se describe en el Ejemplo 3.

DESCRIPCIÓN DETALLADA

[0008] A continuación se hará referencia en detalle a ciertas formas de realización de la invención. Si bien la invención se describirá en conjunto con dichas formas de realización, se entenderá que no tienen la intención de limitar la invención a esas formas de realización. Por el contrario, la invención tiene la intención de cubrir todas las alternativas y modificaciones que pueden incluirse dentro de la invención como se define en las reivindicaciones adjuntas.

[0009] Antes de describir las presentes enseñanzas en detalle, se debe entender que la divulgación no se limita a composiciones o etapas de proceso específicas, ya que pueden variar. Se debe observar que, tal como se utiliza en esta especificación y en las reivindicaciones adjuntas, la forma singular "un", "una", "el" y "ella" incluyen referencias en plural a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "un ácido nucleico" incluye una pluralidad de ácidos nucleicos, la referencia a "una célula" incluye una pluralidad de células, y similares.

[0010] Los rangos numéricos incluyen los números que definen el rango. Se entiende que los valores medidos y mensurables son aproximados, teniendo en cuenta los dígitos significativos y el error asociado con la medición. Además, el uso de "comprende", "comprenden", "que comprende", "contiene", "contienen", "que contiene", "incluye", "incluyen" e "que incluye" no pretende ser limitativo. Se debe entender que tanto la descripción general anterior como la descripción detallada son solo ejemplares y explicativas y no son restrictivas de las enseñanzas.

[0011] A menos que se indique específicamente en la especificación anterior, las formas de realización en la especificación que recitan "que comprende" varios componentes también se contemplan como "que consisten en" o "que consisten esencialmente en" los componentes recitados; las formas de realización en la especificación que recitan "que consisten en" varios componentes también se contemplan como "que comprenden" o "que consisten esencialmente en" los componentes recitados; y las formas de realización en la especificación que recitan "que consisten esencialmente en" varios componentes también se contemplan como "que consisten en" o "que comprenden" los componentes recitados (esta intercambiabilidad no se aplica al uso de estos términos en las reivindicaciones).

[0012] Los encabezados de sección utilizados en este documento son para fines organizativos y no deben interpretarse como limitantes del tema divulgado de ninguna manera. En el caso de que cualquier documento u otro material contradiga cualquier contenido explícito de esta especificación, incluidas las definiciones, prevalece esta especificación.

I. DEFINICIONES

[0013] El término "hipermetilación" se refiere a un nivel o grado aumentado de metilación de una o más moléculas de ácido nucleico en relación con las otras moléculas de ácido nucleico dentro de una población (por ejemplo, muestra) de moléculas de ácido nucleico. En algunas formas de realización, el ADN hipermetilado puede incluir moléculas de ADN que comprenden al menos 1 residuo metilado, al menos 2 residuos metilados, al menos 3 residuos metilados, al menos 5 residuos metilados, al menos 10 residuos metilados, al menos 20 residuos metilados, al menos 25 residuos metilados o al menos 30 residuos metilados. Los residuos metilados incluyen cualquier residuo que contenga un metilo adicional (por ejemplo, metilo sustituido o no sustituido) con respecto a la base no modificada, incluyendo, por ejemplo, 5-metilcitosina y 5-hidroximetilcitosina.

[0014] El término "hipometilación" se refiere a un nivel o grado de metilación disminuido de una o más moléculas de ácido nucleico en relación con las otras moléculas de ácido nucleico dentro de una población (por ejemplo, muestra) de moléculas de ácido nucleico. En algunas formas de realización, el ADN hipometilado incluye moléculas de ADN no metiladas. En algunas formas de realización, el ADN hipometilado puede incluir moléculas de ADN que comprenden 0 residuos metilados, como máximo 1 residuo metilado, como máximo 2 residuos metilados, como máximo 3 residuos metilados, como máximo 4 residuos metilados o como máximo 5 residuos metilados.

[0015] Una "partición hipermetilada" es una partición de ADN en la que una porción mayor del ADN es ADN hipermetilado que en la fuente no particionada de la que se derivó.

[0016] Una "partición hipometilada" es una partición de ADN en la que una porción mayor del ADN es ADN hipometilado que en la fuente no particionada de la que se derivó.

[0017] "ADN libre de células", "moléculas de ADNlc" o simplemente "ADNlc" incluyen moléculas de ADN que se encuentran en un sujeto en forma extracelular (por ejemplo, en sangre, suero, plasma u otros fluidos corporales como linfa, líquido cefalorraquídeo, orina o esputo) e incluye ADN que no está contenido dentro de una célula o de otro modo unido a ella. Si bien el ADN existía originalmente en una célula o células en un organismo biológico complejo grande, por ejemplo, un mamífero, el ADN ha experimentado una liberación desde la(s) célula(s) hacia un fluido que se encuentra en el organismo. Normalmente, el ADNlc se puede obtener mediante la obtención de una muestra del fluido sin la necesidad de realizar un paso de lisis celular in vitro y también incluye la eliminación de células presentes en el fluido (por ejemplo, centrifugación de sangre para eliminar células).

[0018] El "rendimiento de captura" de una colección de sondas para un conjunto de regiones objetivo determinado se refiere a la cantidad (por ejemplo, cantidad relativa a otro conjunto de regiones objetivo o una cantidad absoluta) de ácido nucleico correspondiente al conjunto de regiones objetivo que la colección captura en condiciones típicas. Las condiciones de captura típicas ejemplares son una incubación del ácido nucleico de muestra y las sondas a 65 °C durante 10-18 horas en un pequeño volumen de reacción (aproximadamente 20 µL) que contiene un tampón de hibridación riguroso. El rendimiento de captura puede expresarse en términos absolutos o, para una pluralidad de colecciones de sondas, en términos relativos. Cuando se comparan los rendimientos de captura para una pluralidad de conjuntos de regiones objetivo, se normalizan para el tamaño de la huella del conjunto de regiones objetivo (por ejemplo, sobre una base por kilobase). Así, por ejemplo, si los tamaños de huella de las regiones objetivo primera y segunda son 50 kb y 500 kb, respectivamente (lo que da un factor de normalización de 0,1), entonces el ADN correspondiente al primer conjunto de regiones objetivo se captura con un rendimiento mayor que el ADN correspondiente al segundo conjunto de regiones objetivo cuando la concentración de masa por volumen del ADN capturado correspondiente al primer conjunto de regiones objetivo es más de 0,1 veces la concentración de masa por volumen del ADN capturado correspondiente al segundo conjunto de regiones objetivo. Como ejemplo adicional, utilizando los mismos tamaños de huella, si el ADN capturado correspondiente al primer conjunto de regiones objetivo tiene una concentración de masa por volumen de 0,2 veces la concentración de masa por volumen del ADN capturado correspondiente al segundo conjunto de regiones objetivo, entonces el ADN correspondiente al primer conjunto de regiones objetivo se capturó con un rendimiento de captura dos veces mayor que el ADN correspondiente al segundo conjunto de regiones objetivo.

[0019] "Capturar" o "enriquecer" uno o más ácidos nucleicos objetivo se refiere a aislar o separar preferentemente uno o más ácidos nucleicos objetivo de ácidos nucleicos no objetivo.

[0020] Un "conjunto capturado" de ácidos nucleicos se refiere a ácidos nucleicos que han sufrido captura.

[0021] Un "conjunto de regiones objetivo" o "conjunto de regiones objetivo" o "regiones objetivo" se refiere a una pluralidad de loci genómicos o una pluralidad de regiones genómicas objetivo de captura y/o objetivo de un conjunto de sondas (por ejemplo, mediante complementariedad de secuencias).

[0022] "Correspondiente a un conjunto de regiones objetivo" significa que un ácido nucleico, tal como ADNlc, se originó a partir de un locus en el conjunto de regiones objetivo o se une específicamente a una o más sondas para el conjunto de regiones objetivo.

[0023] "Se une específicamente" en el contexto de una sonda u otro oligonucleótido y una secuencia objetivo significa que, en condiciones de hibridación apropiadas, el oligonucleótido o la sonda se hibridan con su secuencia objetivo, o réplicas de la misma, para formar un híbrido sonda:objetivo estable, mientras que al mismo tiempo se minimiza la formación de híbridos sonda:no objetivo estables. Por lo tanto, una sonda se hibrida con una secuencia objetivo o réplica de la misma en un grado suficientemente mayor que con una secuencia no objetivo, para permitir la captura o detección de la secuencia objetivo. Las condiciones de hibridación apropiadas son bien conocidas en la técnica, se pueden predecir en base a la composición de la secuencia o se pueden determinar utilizando métodos de prueba de rutina (véase, por ejemplo, Sambrook et al., Molecular Cloning, A Laboratory Manual, 2.^a ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) en §§ 1,90-1,91, 7,37-7,57, 9,47-9,51 y 11,47-11,57, particularmente §§ 9,50-9,51, 11,12-11,13, 11,45-11,47 y 11,55-11,57).

[0024] "Conjunto de regiones objetivo de secuencia variable" se refiere a un conjunto de regiones objetivo que pueden exhibir cambios en la secuencia tales como sustituciones de nucleótidos, inserciones, eliminaciones o fusiones o transposiciones de genes en células neoplásicas (por ejemplo, células tumorales y células cancerosas).

[0025] "Conjunto de regiones objetivo epigenéticas" se refiere a un conjunto de regiones objetivo que pueden manifestar modificaciones no secuenciales en células neoplásicas (por ejemplo, células tumorales y células cancerosas) y células no tumorales (por ejemplo, células inmunitarias, células del microambiente tumoral). Estas modificaciones no cambian la secuencia del ADN. Los ejemplos de cambios no secuenciales incluyen, entre otros, cambios en la metilación (aumentos o disminuciones), distribución de nucleosomas, unión de CTCF, sitios de inicio de transcripción, regiones de unión de proteínas reguladoras y cualquier otra proteína que pueda unirse al ADN. Para los propósitos presentes, los loci susceptibles a amplificaciones focales y/o fusiones génicas asociadas a neoplasia, tumor o cáncer también pueden

incluirse en un conjunto de regiones objetivo epigenéticas porque la detección de un cambio en el número de copias mediante secuenciación o una secuencia fusionada que se asigna a más de un locus en un genoma de referencia tiende a ser más similar a la detección de cambios epigenéticos ejemplares discutidos anteriormente que la detección de sustituciones, inserciones o deleciones de nucleótidos, por ejemplo, en que las amplificaciones focales y/o fusiones génicas pueden detectarse a una profundidad de secuenciación relativamente superficial porque su detección no depende de la precisión de las llamadas de bases en una o unas pocas posiciones individuales. Por ejemplo, el conjunto de regiones objetivo epigenéticas puede comprender un conjunto de regiones objetivo para analizar la distribución de la longitud del fragmento o la ubicación del punto final del fragmento. Los términos "epigenético" y "epigenómico" se usan indistintamente en el presente documento.

[0026] Un ADN tumoral circulante o ADNct es un componente de ADNlc que se originó a partir de una célula tumoral o célula cancerosa. En algunas formas de realización, ADNlc comprende ADN que se originó a partir de células normales y ADN que se originó a partir de células tumorales (es decir, ADNct). Las células tumorales son células neoplásicas que se originaron a partir de un tumor, independientemente de si permanecen en el tumor o se separan del tumor (como en los casos, por ejemplo, de células cancerosas metastásicas y células tumorales circulantes).

[0027] Los términos "o una combinación de los mismos" y "o combinaciones de los mismos" como se utilizan en el presente documento se refieren a todas y cada una de las permutaciones y combinaciones de los términos enumerados que preceden al término. Por ejemplo, "A, B, C o combinaciones de los mismos" pretende incluir al menos uno de: A, B, C, AB, AC, BC o ABC, y si el orden es importante en un contexto particular, también BA, CA, CB, ACB, CBA, BCA, BAC o CAB. Continuando con este ejemplo, se incluyen expresamente combinaciones que contienen repeticiones de uno o más elementos o términos, como BB, AAA, AAB, BBC, AAABCCCC, CBBAAA, CABABB, etc. El experto en la materia entenderá que normalmente no hay límite en la cantidad de elementos o términos en cualquier combinación, a menos que sea evidente de otro modo a partir del contexto.

[0028] "O" se utiliza en sentido inclusivo, es decir, equivalente a "y/o", a menos que el contexto requiera lo contrario.

II. MÉTODOS EJEMPLARES

[0029] En el presente documento se proporcionan métodos para aislar ADN y/o identificar la presencia o ausencia de ADN producido por un tumor (o células neoplásicas o células cancerosas). En algunas formas de realización, el ADN comprende ADN libre de células. En algunas formas de realización, el ADN comprende ADN de una muestra de tejido.

[0030] En algunas formas de realización, los métodos comprenden dividir el ADN de una muestra en una pluralidad de particiones, comprendiendo la pluralidad al menos una partición hipermetilada y una partición hipometilada. En algunas formas de realización, los métodos comprenden etiquetar de manera diferencial el ADN de la partición hipermetilada y el ADN de la partición hipometilada. En algunas formas de realización, los métodos comprenden preparar un primer grupo que comprende al menos una porción del ADN de la partición hipometilada. En algunas formas de realización, los métodos comprenden preparar un segundo grupo que comprende al menos una porción del ADN de la partición hipermetilada.

[0031] En algunas formas de realización, el primer grupo comprende además una porción del ADN de la partición hipermetilada. En algunas formas de realización, el segundo grupo comprende además una porción del ADN de la partición hipometilada. En algunas formas de realización, el primer grupo comprende una mayoría del ADN de la partición hipometilada y una minoría del ADN de la partición hipermetilada. En algunas formas de realización, el segundo grupo comprende una mayoría del ADN de la partición hipermetilada y una minoría del ADN de la partición hipometilada.

[0032] En algunas formas de realización, los métodos comprenden capturar al menos un primer conjunto de regiones objetivo del primer grupo. En algunas formas de realización, el primer conjunto comprende regiones objetivo variables de secuencia. En algunas formas de realización, el primer conjunto comprende regiones objetivo variables de hipometilación y/o regiones objetivo variables de fragmentación. En algunas formas de realización, el primer conjunto comprende regiones objetivo variables de secuencia y regiones objetivo variables de fragmentación. En algunas formas de realización, el primer conjunto comprende regiones objetivo variables de secuencia, regiones objetivo variables de hipometilación y regiones objetivo variables de fragmentación. Se puede realizar un paso de amplificación de ADN en el primer grupo antes de este paso de captura. En algunas formas de realización, capturar el primer conjunto de regiones objetivo del primer grupo comprende poner en contacto el ADN del primer grupo con un primer conjunto de sondas específicas de objetivo. En algunas formas de realización, el primer conjunto de sondas específicas de objetivo comprende sondas de unión a objetivo específicas para las regiones objetivo variables de secuencia. En algunas formas de realización, el primer conjunto de sondas específicas al objetivo comprende sondas de unión a la objetivo específicas para las regiones objetivo variables de secuencia, regiones objetivo variables de hipometilación y/o regiones objetivo variables de fragmentación. En algunas formas de realización, los métodos comprenden capturar un segundo conjunto de regiones objetivo o una pluralidad de conjuntos de regiones objetivo del segundo grupo. En algunas formas de realización, la segunda pluralidad comprende regiones objetivo epigenéticas, tales como regiones objetivo variables de hipermetilación y/o regiones objetivo variables de fragmentación. En algunas formas de realización, la segunda pluralidad comprende regiones objetivo variables de secuencia y regiones objetivo epigenéticas, tales como regiones objetivo variables de hipermetilación y/o regiones objetivo variables de fragmentación. Se puede realizar un paso de amplificación del ADN en el segundo grupo antes de este paso de captura. En algunas formas de realización, capturar la segunda pluralidad de conjuntos de regiones

objetivo del segundo grupo comprende poner en contacto el ADN del primer grupo con un segundo conjunto de sondas específicas del objetivo, en donde el segundo conjunto de sondas específicas del objetivo comprende sondas de unión al objetivo específicas para las regiones objetivo de secuencia variable y sondas de unión al objetivo específicas para las regiones objetivo epigenéticas. En algunas formas de realización, el primer conjunto de regiones objetivo y el segundo conjunto de regiones objetivo no son idénticos. Por ejemplo, el primer conjunto de regiones objetivo puede comprender una o más regiones objetivo no presentes en el segundo conjunto de regiones objetivo. Alternativamente o además, el segundo conjunto de regiones objetivo puede comprender una o más regiones objetivo no presentes en el primer conjunto de regiones objetivo. En algunas formas de realización, al menos una región objetivo variable de hipermetilación se captura del segundo grupo, pero no del primer grupo. En algunas formas de realización, una pluralidad de regiones objetivo variables de hipermetilación se capturan del segundo grupo pero no del primer grupo.

[0033] Las regiones objetivo epigenéticas pueden mostrar diferencias en los niveles de metilación y/o patrones de fragmentación dependiendo de si se originaron a partir de un tumor o de células sanas. Las regiones objetivo de secuencia variable pueden mostrar diferencias en la secuencia dependiendo de si se originaron a partir de un tumor o de células sanas. Los pasos de captura producen conjuntos de moléculas capturadas.

[0034] En algunas formas de realización, el análisis de regiones objetivo epigenéticas de la partición hipometilada puede ser menos informativo que el análisis de regiones objetivo de secuencia variable de las particiones hipermetilada e hipometilada y regiones objetivo epigenéticas de la partición hipermetilada. Como tal, la captura de regiones objetivo epigenéticas de la partición hipometilada no está necesariamente incluida en los métodos descritos en el presente documento, o dichas regiones pueden capturarse en menor medida que una o más de las regiones objetivo de secuencia variable de las particiones hipermetilada e hipometilada y regiones objetivo epigenéticas de la partición hipermetilada. Por ejemplo, el primer grupo, del que se capturan regiones objetivo de secuencia variable, puede prepararse con algo (por ejemplo, una mayoría) del ADN de la partición hipometilada; y/o el segundo grupo, del que se capturan regiones objetivo epigenéticas y variables de secuencia, se puede preparar con una parte (por ejemplo, una mayoría, sustancialmente la totalidad o la totalidad) del ADN de la partición hipermetilada y nada o una parte (por ejemplo, una minoría) del ADN de la partición hipometilada. Dichos enfoques pueden reducir o eliminar la secuenciación de regiones objetivo epigenéticas de la partición hipometilada, reduciendo así la cantidad de datos de secuenciación que son suficientes para un análisis posterior.

[0035] En algunas formas de realización, incluir una minoría del ADN de la partición hipometilada en el segundo grupo facilita la cuantificación de una o más características epigenéticas (por ejemplo, metilación u otras características epigenéticas analizadas en detalle en otra parte del presente documento), por ejemplo, sobre una base relativa.

[0036] En algunas formas de realización, el primer grupo comprende al menos aproximadamente el 50 % del ADN de la partición hipometilada. Por ejemplo, la primera porción puede comprender al menos aproximadamente el 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 % o 95 % del ADN de la partición hipometilada. En algunas formas de realización, el 50-55 %, 55-60 %, 60-65 %, 65-70 %, 70-75 %, 75-80 %, 80-85 %, 85-90 % o 90-95 % del ADN de la partición hipometilada. En algunas formas de realización, el primer grupo comprende aproximadamente el 70-90 % del ADN de la partición hipometilada. En algunas formas de realización, el primer grupo comprende aproximadamente el 75-85 % del ADN de la partición hipometilada. En algunas formas de realización, la primera porción comprende aproximadamente el 80 % del ADN de la partición hipometilada. En algunas formas de realización, la primera porción comprende al menos aproximadamente el 80 % del ADN de la partición hipometilada. En algunas formas de realización, el primer grupo comprende sustancialmente todo el ADN de la partición hipometilada.

[0037] En algunas formas de realización, el segundo grupo comprende una segunda porción del ADN de la partición hipometilada. La segunda porción puede ser una minoría del ADN de la partición hipometilada, por ejemplo, menos de aproximadamente el 50 % del ADN de la partición hipometilada, tal como menos de o igual a aproximadamente el 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o 5 % del ADN de la partición hipometilada. En algunas formas de realización, la segunda porción es aproximadamente el 5 %-25 % del ADN de la partición hipometilada. En algunas formas de realización, la segunda porción es aproximadamente el 10 %-20 % del ADN de la partición hipometilada. En algunas formas de realización, la segunda porción es aproximadamente el 10 % del ADN de la partición hipometilada. En algunas formas de realización, la segunda porción es aproximadamente el 15 % del ADN de la partición hipometilada. En algunas formas de realización, la segunda porción es aproximadamente el 20 % del ADN de la partición hipometilada.

[0038] En algunas formas de realización, el primer grupo comprende sustancialmente todo o la totalidad del ADN de la partición hipometilada (por ejemplo, en donde el segundo grupo no comprende ADN de la partición hipometilada). En algunas formas de realización, el segundo grupo no comprende ADN de la partición hipometilada (por ejemplo, en donde el primer grupo comprende sustancialmente todo o la totalidad del ADN de la partición hipometilada).

[0039] En algunas formas de realización, el segundo grupo comprende una parte de la partición hipermetilada, que puede ser cualquiera de los valores y rangos establecidos anteriormente con respecto a la partición hipometilada. En algunas formas de realización, el segundo grupo comprende la totalidad o sustancialmente la totalidad de la partición hipermetilada.

[0040] En algunas formas de realización, la segunda pluralidad de conjuntos de regiones objetivo o segundo conjunto de regiones objetivo comprende un mayor número de regiones objetivo epigenéticas que la primera pluralidad de conjuntos de regiones objetivo o primer conjunto de regiones objetivo. En este contexto, "mayor número" significa que hay más regiones objetivo epigenéticas distintas (en el sentido de una mayor diversidad de loci) en la primera pluralidad que en la segunda pluralidad. En algunas formas de realización, la segunda pluralidad de conjuntos de regiones objetivo o segundo conjunto de regiones objetivo comprende una mayor cantidad de regiones objetivo epigenéticas que la primera pluralidad de conjuntos de regiones objetivo o primer conjunto de regiones objetivo (por ejemplo, una mayor masa o cantidad molar de regiones objetivo epigenéticas). En algunas formas de realización, la primera pluralidad de conjuntos de regiones objetivo o primer conjunto de regiones objetivo no comprende regiones objetivo epigenéticas. En algunas formas de realización, la primera pluralidad de conjuntos de regiones objetivo o primer conjunto de regiones objetivo no comprende regiones objetivo variables de hipermetilación.

[0041] En algunas formas de realización, las moléculas de ADN correspondientes al conjunto de regiones objetivo de secuencia variable se capturan con un rendimiento de captura mayor en el conjunto capturado de moléculas de ADN que las moléculas de ADN correspondientes al conjunto de regiones objetivo epigenéticas.

[0042] En algunas formas de realización, los métodos comprenden poner en contacto ADN obtenido de un sujeto de prueba con un conjunto de sondas específicas del objetivo, en donde el conjunto de sondas específicas del objetivo está configurado para capturar ADN correspondiente al conjunto de regiones objetivo de secuencia variable con un rendimiento de captura mayor que el ADN correspondiente al conjunto de regiones objetivo epigenéticas.

[0043] Puede ser beneficioso capturar ADN correspondiente al conjunto de regiones objetivo de secuencia variable con un mayor rendimiento de captura que el ADN correspondiente al conjunto de regiones objetivo epigenéticas porque puede ser necesaria una mayor profundidad de secuenciación para analizar las regiones objetivo de secuencia variable con suficiente confianza o precisión que la que puede ser necesaria para analizar las regiones objetivo epigenéticas. La mayor profundidad de secuenciación puede dar como resultado más lecturas por molécula de ADN y puede facilitarse mediante la captura de más moléculas únicas por región. El volumen de datos necesario para determinar patrones de fragmentación (por ejemplo, para probar la perturbación de los sitios de inicio de la transcripción o sitios de unión de CTCF) o la abundancia de fragmentos (por ejemplo, en particiones hipermetiladas e hipometiladas) es generalmente menor que el volumen de datos necesario para determinar la presencia o ausencia de mutaciones de secuencia relacionadas con el cáncer. La captura de los conjuntos de regiones objetivo con diferentes rendimientos puede facilitar la secuenciación de las regiones objetivo a diferentes profundidades de secuenciación en la misma ejecución de secuenciación (por ejemplo, utilizando una mezcla agrupada y/o en la misma célula de secuenciación).

[0044] En diversas formas de realización, los métodos comprenden además la secuenciación del ADN capturado, por ejemplo, a diferentes grados de profundidad de secuenciación para los conjuntos de regiones objetivo epigenéticas y de secuencia variable, de acuerdo con la discusión anterior.

1. Etapa de captura; amplificación; adaptadores; códigos de barras

[0045] En algunas formas de realización, los métodos divulgados en el presente documento comprenden una o más etapas de captura de uno o más conjuntos de regiones objetivo de ADN, como ADNlc. La captura se puede realizar utilizando cualquier enfoque adecuado conocido en la técnica.

[0046] En algunas formas de realización, la captura comprende poner en contacto el ADN que se va a capturar con un conjunto de sondas específicas del objetivo. El conjunto de sondas específicas del objetivo puede tener cualquiera de las características descritas en el presente documento para conjuntos de sondas específicas del objetivo, incluidas, entre otras, las formas de realización expuestas anteriormente y las secciones relacionadas con las sondas a continuación.

[0047] La etapa de captura se puede realizar utilizando condiciones adecuadas para la hibridación de ácidos nucleicos específicos, que generalmente dependen en cierta medida de las características de las sondas, como la longitud, la composición de bases, etc. Los expertos en la materia estarán familiarizados con las condiciones adecuadas dado el conocimiento general en la técnica con respecto a la hibridación de ácidos nucleicos. En algunas formas de realización, se forman complejos de sondas específicas al objetivo y ADN.

[0048] En algunas formas de realización, los complejos de sondas específicas al objetivo y ADN se separan del ADN no unido a las sondas específicas al objetivo. Por ejemplo, cuando las sondas específicas al objetivo están unidas covalentemente o no covalentemente a un soporte sólido, se puede utilizar una etapa de lavado o aspiración para separar el material no unido. Alternativamente, cuando los complejos tienen propiedades cromatográficas distintas del material no unido (por ejemplo, cuando las sondas comprenden un ligando que se une a una resina cromatográfica), se puede utilizar la cromatografía.

[0049] Como se analiza en detalle en otra parte del presente documento, el conjunto de sondas específicas al objetivo puede comprender una pluralidad de conjuntos, como sondas para un conjunto de regiones objetivo de secuencia variable y sondas para un conjunto de regiones objetivo epigenéticas. En algunas de dichas formas de realización, la etapa de captura se realiza con las sondas para el conjunto de regiones objetivo de secuencia variable y las sondas para el conjunto

de regiones objetivo epigenéticas en el mismo recipiente al mismo tiempo, por ejemplo, las sondas para los conjuntos de regiones objetivo de secuencia variable y epigenéticas están en la misma composición. Este enfoque proporciona un flujo de trabajo relativamente optimizado. En algunas formas de realización, la concentración de las sondas para el conjunto de regiones objetivo de secuencia variable es mayor que la concentración de las sondas para el conjunto de regiones objetivo epigenéticas.

[0050] Como alternativa, la etapa de captura se realiza con el conjunto de sondas de la región objetivo de secuencia variable en un primer recipiente y con el conjunto de sondas de la región objetivo epigenética en un segundo recipiente, o la etapa de contacto se realiza con el conjunto de sondas de la región objetivo de secuencia variable en un primer momento y un primer recipiente y el conjunto de sondas de la región objetivo epigenética en un segundo momento antes o después del primer momento. Este enfoque permite la preparación de primera y segunda composiciones separadas que comprenden ADN capturado correspondiente al conjunto de la región objetivo de secuencia variable y ADN capturado correspondiente al conjunto de la región objetivo epigenética. Las composiciones se pueden procesar por separado según se desee (por ejemplo, para fraccionar en función de la metilación como se describe en otra parte del presente documento) y recombinarse en proporciones apropiadas para proporcionar material para un procesamiento y análisis adicionales, como la secuenciación.

[0051] En algunas formas de realización, se amplifica el ADN. En algunas formas de realización, la amplificación se realiza antes del paso de captura. En algunas formas de realización, la amplificación se realiza después del paso de captura. En la técnica se conocen métodos para la amplificación no específica del ADN, por ejemplo, la amplificación por PCR no específica. Véase, por ejemplo, Smallwood et al., Nat. Methods 11: 817-820 (2014). Por ejemplo, se pueden utilizar cebadores aleatorios que tienen secuencias adaptadoras en sus extremos 5' y bases aleatorias en el extremo 3'. Normalmente hay 6 bases aleatorias, pero pueden tener entre 4 y 9 bases de longitud. Este enfoque es adecuado para la amplificación de células individuales/con entrada baja y/o la secuenciación con bisulfito.

[0052] En algunas formas de realización, se incluyen adaptadores en el ADN. Esto se puede hacer simultáneamente con un procedimiento de amplificación, por ejemplo, proporcionando los adaptadores en una porción 5' de un cebador, por ejemplo, como se describió anteriormente. Alternativamente, se pueden agregar adaptadores mediante otros enfoques, como la ligación.

[0053] En algunas formas de realización, se incluyen en el ADN etiquetas, que pueden ser o incluir códigos de barras. Las etiquetas pueden facilitar la identificación del origen de un ácido nucleico. Por ejemplo, se pueden utilizar códigos de barras para permitir que se identifique el origen (por ejemplo, el sujeto) de donde proviene el ADN después de la agrupación de una pluralidad de muestras para la secuenciación paralela. Esto se puede hacer simultáneamente con un procedimiento de amplificación, por ejemplo, proporcionando los códigos de barras en una porción 5' de un cebador, por ejemplo, como se describió anteriormente. En algunas formas de realización, los adaptadores y las etiquetas/códigos de barras se proporcionan mediante el mismo cebador o conjunto de cebadores. Por ejemplo, el código de barras puede estar ubicado 3' del adaptador y 5' de la porción de hibridación con el objetivo del cebador. Alternativamente, se pueden agregar códigos de barras mediante otros enfoques, como la ligadura, opcionalmente junto con adaptadores en el mismo sustrato de ligadura.

[0054] Los detalles adicionales con respecto a la amplificación, las etiquetas y los códigos de barras se analizan en la sección "Características generales de los métodos" a continuación, que se pueden combinar en la medida de lo posible con cualquiera de las formas de realización anteriores y las formas de realización expuestas en la sección de introducción y resumen.

2. Conjunto capturado

[0055] En algunas formas de realización, se proporcionan conjuntos capturados de ADN (por ejemplo, ADNlc). En algunas formas de realización, se captura un primer conjunto del primer grupo y se captura un segundo conjunto del segundo grupo. Con respecto a los métodos descritos, dichos conjuntos capturados de ADN se pueden proporcionar, por ejemplo, después de los pasos de captura y/o separación como se describe en el presente documento. Un conjunto capturado, como el segundo conjunto capturado, puede comprender ADN correspondiente a un conjunto de región objetivo de secuencia variable y un conjunto de región objetivo epigenética. En algunas formas de realización, la cantidad de ADN de región objetivo de secuencia variable capturado es mayor que la cantidad de ADN de región objetivo epigenética capturado, cuando se normaliza para la diferencia en el tamaño de las regiones objetivo (tamaño de la huella). En algunas formas de realización, el primer conjunto capturado comprende ADN correspondiente a un conjunto de regiones objetivo de secuencia variable. En algunas formas de realización, el primer conjunto capturado no comprende ADN correspondiente a un conjunto de regiones objetivo epigenéticas. En algunas formas de realización, el primer conjunto capturado comprende menos ADN correspondiente a un conjunto de regiones objetivo epigenéticas que el segundo conjunto capturado. En algunas formas de realización, las regiones objetivo epigenéticas presentes en el primer conjunto capturado comprenden regiones objetivo variables de fragmentación, sitios de unión de CTCF, sitios de inicio de transcripción, regiones objetivo variables de hipermetilación, sitios de empalme, elementos de control reguladores no codificantes (por ejemplo, potenciadores) y/o regiones que pueden mostrar variaciones en el número de copias, tales como amplificaciones focales en el cáncer. En algunas formas de realización, cuando el primer conjunto capturado comprende regiones objetivo epigenéticas, el segundo conjunto capturado comprende regiones objetivo epigenéticas que

no están presentes en el primer conjunto capturado. Por ejemplo, el segundo conjunto capturado puede comprender regiones objetivo variables de hipermetilación que no están presentes en el primer conjunto capturado. En otro ejemplo, el segundo conjunto capturado puede comprender regiones objetivo variables de hipermetilación y el primer conjunto capturado puede carecer de regiones objetivo variables de hipermetilación.

[0056] En un conjunto capturado que comprende ADN correspondiente al conjunto de regiones objetivo de secuencia variable y al conjunto de regiones objetivo epigenéticas, el ADN correspondiente al conjunto de regiones objetivo de secuencia variable puede estar presente en una concentración mayor que el ADN correspondiente al conjunto de regiones objetivo epigenéticas, por ejemplo, una concentración de 1,1 a 1,2 veces mayor, una concentración de 1,2 a 1,4 veces mayor, una concentración de 1,4 a 1,6 veces mayor, una concentración de 1,6 a 1,8 veces mayor, una concentración de 1,8 a 2,0 veces mayor, una concentración de 2,0 a 2,2 veces mayor, una concentración de 2,2 a 2,4 veces mayor, una concentración de 2,4 a 2,6 veces mayor, una concentración de 2,6 a 2,8 veces mayor, una concentración de 2,8 a 3,0 veces mayor, una concentración de 3,0 a 3,5, 3,5 a 4,0, una concentración de 4,0 a 4,5 veces mayor, una concentración de 4,5 a 5,0 veces mayor, una concentración de 5,0 a 5,5 veces mayor, una concentración de 5,5 a 6,0 veces mayor, una concentración de 6,0 a 6,5 veces mayor, una concentración de 6,5 a 7,0 veces mayor, una concentración de 7,0 a 7,5 veces mayor, una concentración de 7,5 a 8,0 veces mayor, una concentración de 8,0 a 8,5 veces mayor, una concentración de 8,5 a 9,0 veces mayor, una concentración de 9,0 a 9,5 veces mayor, una concentración de 9,5 a 10,0 veces mayor, una concentración de 10 a 11 veces mayor, una concentración de 11 a 12 veces mayor, una concentración de 12 a 13 veces mayor, una concentración de 13 a 14 veces mayor, una concentración de 14 a 15 veces mayor, una concentración de 15 a 16 veces mayor, una concentración de 16 a 17 veces mayor, una concentración de 17 a 18 veces mayor, una concentración de 18 a 19 veces mayor o una concentración de 19 a 20 veces mayor. El grado de diferencia en las concentraciones explica la normalización de los tamaños de huella de las regiones objetivo, como se analiza en la sección de definición.

a. Conjunto de regiones objetivo epigenéticas

[0057] El conjunto de regiones objetivo epigenéticas puede comprender uno o más tipos de regiones objetivo que probablemente diferencien el ADN de células neoplásicas (por ejemplo, tumorales o cancerosas) y de células sanas, por ejemplo, células circulantes no neoplásicas. En el presente documento se analizan en detalle tipos ejemplares de dichas regiones. En algunas formas de realización, los métodos de acuerdo con la divulgación comprenden determinar si las moléculas de ADNlc correspondientes al conjunto de regiones objetivo epigenéticas comprenden o indican modificaciones epigenéticas asociadas con el cáncer (por ejemplo, hipermetilación en una o más regiones objetivo variables de hipermetilación; una o más perturbaciones de la unión de CTCF; y/o una o más perturbaciones de los sitios de inicio de la transcripción) y/o variaciones en el número de copias (por ejemplo, amplificaciones focales). El conjunto de regiones objetivo epigenéticas también puede comprender una o más regiones de control, por ejemplo, como se describe en el presente documento.

[0058] En algunas formas de realización, el conjunto de regiones objetivo epigenéticas tiene una huella de al menos 100 kb, por ejemplo, al menos 200 kb, al menos 300 kb o al menos 400 kb. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas tiene una huella en el intervalo de 100-1.000 kb, por ejemplo, 100-200 kb, 200-300 kb, 300-400 kb, 400-500 kb, 500-600 kb, 600-700 kb, 700-800 kb, 800-900 kb y 900-1.000 kb.

i. Regiones objetivo variables de hipermetilación

[0059] En algunas formas de realización, el conjunto de regiones objetivo epigenéticas comprende una o más regiones objetivo variables de hipermetilación. En general, las regiones objetivo variables de hipermetilación se refieren a regiones en las que un aumento en el nivel de metilación observado indica una mayor probabilidad de que una muestra (por ejemplo, de ADNlc) contenga ADN producido por células neoplásicas, como células tumorales o cancerosas. Por ejemplo, se ha observado repetidamente la hipermetilación de promotores de genes supresores de tumores. Véase, por ejemplo, Kang et al., *Genome Biol.* 18:53 (2017) y las referencias citadas allí.

[0060] En Lam et al., *Biochim Biophys Acta.* 1866:106-20 (2016) se proporciona una discusión extensa de las regiones objetivo variables de metilación en el cáncer colorrectal. Estas incluyen VIM, SEPT9, ITGA4, OSM4, GATA4 y NDRG4. En la Tabla 1 se proporciona un conjunto ejemplar de regiones objetivo variables de hipermetilación que comprenden los genes o partes de los mismos basados en los estudios de cáncer colorrectal (CCR). Muchos de estos genes probablemente tengan relevancia para cánceres más allá del cáncer colorrectal; por ejemplo, TP53 es ampliamente reconocido como un supresor tumoral de importancia crítica y la inactivación basada en hipermetilación de este gen puede ser un mecanismo oncogénico común.

Tabla 1. Regiones objetivo de hipermetilación ejemplares (genes o partes de las mismas) basadas en estudios de CCR.

Nombre del gen	Nombre adicional del gen	Cromosoma
VIM		chr10
SEPT9		chr17
CYCD2	CCND2	chr12
TFPI2		chr7
GATA4		chr8
RARB2	RARB	chr3
P16INK4a	CDKN2A	chr9
MGMT	MGMT	chr10
APC		chr5
NDRG4		chr16
HLTF		chr3
HPP1	TMEFF2	chr2
hMLH1	MLH1	chr3
RASSF1A	RASSF1	chr3
CDH13		chr16
IGFBP3		chr7
ITGA4		chr2

[0061] En algunas formas de realización, las regiones objetivo variables de hipermetilación comprenden una pluralidad de genes o porciones de los mismos enumerados en la Tabla 1, p. ej., al menos 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, o 100 % de los genes o porciones de los mismos enumerados en la Tabla 1. Por ejemplo, para cada locus incluido como región objetivo, puede haber una o más sondas con un sitio de hibridación que se une entre el sitio de inicio de la transcripción y el codón de terminación (el último codón de terminación para genes que se empalman alternativamente) del gen. En algunas formas de realización, las una o más sondas se unen dentro de los 300 pb corriente arriba y/o corriente abajo de los genes o porciones de los mismos enumerados en la Tabla 1, por ejemplo, dentro de los 200 o 100 pb.

[0062] Las regiones objetivo variables de metilación en varios tipos de cáncer de pulmón se analizan en detalle, por ejemplo, en Ooki et al., Clin. Cancer Res. 23:7141-52 (2017); Belinsky, Annu. Rev. Physiol. 77:453-74 (2015); Hulbert et al., Clin. Cancer Res. 23:1998-2005 (2017); Shi et al., BMC Genomics 18:901 (2017); Schneider et al., BMC Cancer. 11:102 (2011); Lissa et al., Transl Lung Cancer Res 5(5):492-504 (2016); Skvortsova et al., Br. Cancer. 94(10):1492-1495 (2006); Kim et al., Cancer Res. 61:3419-3424 (2001); Furonaka y col., Pathology International 55:303-309 (2005); Gomes y col., Rev. Port. Pneumol. 20:20-30 (2014); Kim y col., Oncogene. 20:1765-70 (2001); Hopkins-Donaldson y col., Cell Death Differ. 10:356-64 (2003); Kikuchi y col., Clin. Cancer Res. 11:2954-61 (2005); Heller y col., Oncogene 25:959-968 (2006); Licchesi y col., Carcinogenesis. 29:895-904 (2008); Guo y col., Clin. Cancer Res. 10:7917-24 (2004); Palmisano y col., Cancer Res. 63:4620-4625 (2003); y Toyooka et al., Cancer Res. 61:4556-4560, (2001).

[0063] En la Tabla 2 se proporciona un conjunto ejemplar de regiones objetivo variables de hipermetilación que comprenden genes o porciones de los mismos basados en los estudios de cáncer de pulmón. Muchos de estos genes probablemente tengan relevancia para cánceres más allá del cáncer de pulmón; por ejemplo, Casp8 (Caspasa 8) es una enzima clave en la muerte celular programada y la inactivación basada en hipermetilación de este gen puede ser un mecanismo oncogénico común no limitado al cáncer de pulmón. Además, varios genes aparecen en las Tablas 1 y 2, lo que indica generalidad.

Tabla 2. Regiones objetivo de hipermetilación ejemplares (genes o porciones de los mismos) basadas en estudios de cáncer de pulmón

Nombre del gen	Cromosoma
MARCH11	chr5
TAC1	chr7
TCF21	chr6
SHOX2	chr3
P16	chr3
Casp8	chr2
CDH13	chr16
MGMT	chr10
MLH1	chr3
MSH2	chr2
TSLC1	chr11
APC	chr5
DKK1	chr10
DKK3	chr11
LKBI	chr11
WIFI	Chr12
RUNX3	chr11
GATA4	chr8
GATA5	chr20
PAX5	chr9
E-Cadherina	chr16
H-Cadherina	chr16

[0064] Cualquiera de las formas de realización anteriores relativas a las regiones objetivo identificadas en la Tabla 2 se puede combinar con cualquiera de las formas de realización descritas anteriormente relativas a las regiones objetivo identificadas en la Tabla 1. En algunas formas de realización, se pueden combinar las formas de realización descritas anteriormente relativas a las regiones objetivo identificadas en la Tabla 1. En algunas formas de realización, las regiones objetivo variables de hipermetilación comprenden una pluralidad de genes o porciones de los mismos enumerados en la Tabla 1 o la Tabla 2, por ejemplo, al menos el 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los genes o porciones de los mismos enumerados en la Tabla 1 o la Tabla 2.

[0065] Se pueden obtener regiones objetivo de hipermetilación adicionales, por ejemplo, del Atlas del Genoma del Cáncer. Kang et al., Genome Biology 18:53 (2017), describe la construcción de un método probabilístico llamado Cancer Locator usando regiones objetivo de hipermetilación de mama, colon, riñón, hígado y pulmón. En algunas formas de realización, las regiones objetivo de hipermetilación pueden ser específicas para uno o más tipos de cáncer. En consecuencia, en algunas formas de realización, las regiones objetivo de hipermetilación incluyen uno, dos, tres, cuatro o cinco subconjuntos de regiones objetivo de hipermetilación que colectivamente muestran hipermetilación en uno, dos, tres, cuatro o cinco cánceres de mama, colon, riñón, hígado y pulmón.

ii. Regiones objetivo variables de hipometilación

[0066] La hipometilación global es un fenómeno observado comúnmente en varios cánceres. Véase, por ejemplo, Hon et al., Genome Res. 22:246-258 (2012) (cáncer de mama); Ehrlich, Epigenomics 1:239-259 (2009) (artículo de revisión que señala observaciones de hipometilación en cánceres de colon, ovario, próstata, leucemia, hepatocelular y de cuello uterino). Por ejemplo, regiones como elementos repetidos, por ejemplo, elementos LINE1, elementos Alu, repeticiones en tándem centroméricas, repeticiones en tándem pericentroméricas y ADN satélite, y regiones intergénicas que normalmente están metiladas en células sanas pueden mostrar una metilación reducida en células tumorales. En consecuencia, en algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye regiones objetivo variables de hipometilación, donde una disminución en el nivel de metilación observada indica una mayor probabilidad de

que una muestra (por ejemplo, de ADNlc) contenga ADN producido por células neoplásicas, como células tumorales o cancerosas.

[0067] En algunas formas de realización, las regiones objetivo variables de hipometilación incluyen elementos repetidos y/o regiones intergénicas. En algunas formas de realización, los elementos repetidos incluyen uno, dos, tres, cuatro o cinco de los elementos LINE1, elementos Alu, repeticiones en tándem centroméricas, repeticiones en tándem pericentroméricas y/o ADN satélite.

[0068] Las regiones genómicas específicas ejemplares que muestran hipometilación asociada al cáncer incluyen los nucleótidos 8403565-8953708 y 151104701-151106035 del cromosoma humano 1, por ejemplo, de acuerdo con la construcción del genoma humano hgl 9 o hg38. En algunas formas de realización, las regiones objetivo variables de hipometilación se superponen o comprenden una o ambas de estas regiones.

iii. Regiones de unión de CTCF

[0069] CTCF es una proteína de unión al ADN que contribuye a la organización de la cromatina y a menudo se colocaliza con la cohesina. Se ha informado de la perturbación de los sitios de unión de CTCF en una variedad de cánceres diferentes. Véase, por ejemplo, Katainen et al., *Nature Genetics*, doi:10.1038/ng.3335, publicado en línea el 8 de junio de 2015; Guo et al., *Nat. Commun.* 9:1520 (2018). La unión de CTCF da como resultado patrones reconocibles en ADNlc que se pueden detectar mediante secuenciación, por ejemplo, a través del análisis de la longitud de fragmentos. Por ejemplo, se proporcionan detalles sobre el análisis de la longitud de fragmentos basado en secuenciación en Snyder et al., *Cell* 164:57-68 (2016); WO 2018/009723; y US20170211143A1.

[0070] Por lo tanto, las perturbaciones de la unión de CTCF dan como resultado una variación en los patrones de fragmentación de ADNlc. Como tal, los sitios de unión de CTCF representan un tipo de regiones objetivo variables de fragmentación.

[0071] Existen muchos sitios de unión de CTCF conocidos. Véase, por ejemplo, la base de datos CTCFBSDB (CTCF Binding Site Database), disponible en Internet en insulatordb.uthsc.edu/; Cuddapah et al., *Genome Res.* 19:24-32 (2009); Martin et al., *Nat. Struct. Mol. Biol.* 18:708-14 (2011); Rhee et al., *Cell.* 147:1408-19 (2011). Los sitios de unión de CTCF ejemplares se encuentran en los nucleótidos 56014955-56016161 en el cromosoma 8 y los nucleótidos 95359169-95360473 en el cromosoma 13, por ejemplo, de acuerdo con la construcción del genoma humano hg19 o hg38.

[0072] Por consiguiente, en algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye regiones de unión a CTCF. En algunas formas de realización, las regiones de unión a CTCF comprenden al menos 10, 20, 50, 100, 200 o 500 regiones de unión a CTCF, o 10-20, 20-50, 50-100, 100-200, 200-500 o 500-1.000 regiones de unión a CTCF, por ejemplo, como las regiones de unión a CTCF descritas anteriormente o en uno o más de los artículos de CTCFBSDB o Cuddapah et al., Martin et al. o Rhee et al. citados anteriormente.

[0073] En algunas formas de realización, al menos algunos de los sitios CTCF pueden estar metilados o no metilados, en donde el estado de metilación se correlaciona con si la célula es o no una célula cancerosa. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas comprende al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 750 pb, al menos 1.000 pb de regiones corriente arriba y/o corriente abajo de los sitios de unión CTCF.

iv. Sitios de inicio de la transcripción

[0074] Los sitios de inicio de la transcripción también pueden mostrar perturbaciones en células neoplásicas. Por ejemplo, la organización de nucleosomas en varios sitios de inicio de la transcripción en células sanas del linaje hematopoyético - que contribuye sustancialmente al ADNlc en individuos sanos- puede diferir de la organización de nucleosomas en esos sitios de inicio de la transcripción en células neoplásicas. Esto da como resultado diferentes patrones de ADNlc que pueden detectarse mediante secuenciación, por ejemplo, como se analiza en general en Snyder et al., *Cell* 164:57-68 (2016); WO 2018/009723; y US20170211143A1.

[0075] Por lo tanto, las perturbaciones de los sitios de inicio de la transcripción también dan como resultado una variación en los patrones de fragmentación del ADNlc. Como tal, los sitios de inicio de la transcripción también representan un tipo de regiones objetivo variables de fragmentación.

[0076] Los sitios de inicio de la transcripción humana están disponibles en DBTSS (Base de datos de sitios de inicio de la transcripción humana), disponible en Internet en dbtss.hgc.jp y descrito en Yamashita et al., *Nucleic Acids Res.* 34 (número de la base de datos): D86-D89 (2006).

[0077] Por consiguiente, en algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye sitios de inicio de la transcripción. En algunas formas de realización, los sitios de inicio de la transcripción comprenden al menos 10, 20, 50, 100, 200 o 500 sitios de inicio de la transcripción, o 10-20, 20-50, 50-100, 100-200, 200-500 o 500-1.000 sitios de inicio de la transcripción, por ejemplo, como los sitios de inicio de la transcripción enumerados en DBTSS. En algunas

formas de realización, al menos algunos de los sitios de inicio de la transcripción pueden estar metilados o no metilados, en donde el estado de metilación se correlaciona con el hecho de que la célula sea o no una célula cancerosa. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas comprende al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 750 pb, al menos 1.000 pb de regiones ascendentes y/o descendentes de los sitios de inicio de la transcripción.

v. Variaciones en el número de copias; amplificaciones focales

[0078] Aunque las variaciones en el número de copias, como las amplificaciones focales, son mutaciones somáticas, se pueden detectar mediante secuenciación basada en la frecuencia de lectura de una manera análoga a los enfoques para detectar ciertos cambios epigenéticos, como los cambios en la metilación. Como tal, las regiones que pueden mostrar variaciones en el número de copias, tales como amplificaciones focales en el cáncer, se pueden incluir en el conjunto de regiones objetivo epigenéticas y pueden comprender una o más de AR, BRAF, CCND1, CCND2, CCNE1, CDK4, CDK6, EGFR, ERBB2, FGFR1, FGFR2, KIT, KRAS, MET, MYC, PDGFRA, PIK3CA y RAF1. Por ejemplo, en algunas formas de realización, el conjunto de regiones objetivo epigenéticas comprende al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 o 18 de las dianas anteriores.

vi. Regiones de control de metilación

[0079] Puede ser útil incluir regiones de control para facilitar la validación de datos. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye regiones de control que se espera que estén metiladas o no metiladas en esencialmente todas las muestras, independientemente de si el ADN se deriva de una célula cancerosa o de una célula normal. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye regiones hipometiladas de control que se espera que estén hipometiladas en esencialmente todas las muestras. En algunas formas de realización, el conjunto de regiones objetivo epigenéticas incluye regiones hipermetiladas de control que se espera que estén hipermetiladas en esencialmente todas las muestras.

b. Conjunto de regiones objetivo de secuencia variable

[0080] En algunas formas de realización, el conjunto de regiones objetivo de secuencia variable comprende una pluralidad de regiones que se sabe que sufren mutaciones somáticas en el cáncer (denominadas en el presente documento mutaciones asociadas al cáncer). En consecuencia, los métodos pueden comprender determinar si las moléculas de ADNlc correspondientes al conjunto de regiones objetivo de secuencia variable comprenden mutaciones asociadas al cáncer.

[0081] En algunas formas de realización, el conjunto de regiones objetivo de secuencia variable se dirige a una pluralidad de genes o regiones genómicas diferentes ("panel") seleccionados de modo que una proporción determinada de sujetos que tienen un cáncer exhibe una variante genética o un marcador tumoral en uno o más genes o regiones genómicas diferentes en el panel. El panel puede seleccionarse para limitar una región para secuenciación a un número fijo de pares de bases. El panel puede seleccionarse para secuenciar una cantidad deseada de ADN, por ejemplo, ajustando la afinidad y/o la cantidad de las sondas como se describe en otra parte del presente documento. El panel puede seleccionarse además para lograr una profundidad de lectura de secuencia deseada. El panel puede seleccionarse para lograr una profundidad de lectura de secuencia deseada o una cobertura de lectura de secuencia para una cantidad de pares de bases secuenciados. El panel puede seleccionarse para lograr una sensibilidad teórica, una especificidad teórica y/o una precisión teórica para detectar una o más variantes genéticas en una muestra. En algunas formas de realización, el conjunto de regiones objetivo de secuencia variable se puede utilizar para detectar mutaciones somáticas en uno o más genes o regiones genómicas. En algunas formas de realización, las mutaciones somáticas pueden ser al menos una de estas mutaciones, tales como SNV, indeles, variantes de número de copias, fusiones de genes y/o pérdida de heterocigosidad.

[0082] Las sondas para detectar el panel de regiones pueden incluir aquellas para detectar regiones genómicas de interés (regiones de puntos calientes) así como sondas que reconocen nucleosomas (por ejemplo, codones KRAS 12 y 13) y pueden estar diseñadas para optimizar la captura en función del análisis de la cobertura de ADNlc y la variación del tamaño de fragmentos impactada por los patrones de unión de nucleosomas y la composición de la secuencia de GC. Las regiones utilizadas en el presente documento también pueden incluir regiones que no sean puntos calientes optimizadas en función de las posiciones de nucleosomas y los modelos de GC.

[0083] Se pueden encontrar ejemplos de listados de ubicaciones genómicas de interés en la Tabla 3 y la Tabla 4. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos una porción de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65 o 70 de los genes de la Tabla 3. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, o 70 de los SNV de la Tabla 3. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, o 6 de

las fusiones de la Tabla 3. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos una parte de al menos 1, al menos 2, o 3 de los indeles de la Tabla 3. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos una parte de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70, o 73 de los genes de Tabla 4. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70 o 73 de los SNV de la Tabla 4. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos 1, al menos 2, al menos 3, al menos 4, al menos 5 o 6 de las fusiones de la Tabla 4. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos una parte de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29, al menos 30, al menos 31, al menos 32, al menos 33, al menos 34, al menos 35, al menos 36, al menos 37, al menos 38, al menos 39, al menos 40, al menos 41, al menos 42, al menos 43, al menos 44, al menos 45, al menos 46, al menos 47, al menos 48, al menos 49, al menos 50, al menos 51, al menos al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17 o 18 de los indeles de la Tabla 4. Cada una de estas ubicaciones genómicas de interés puede identificarse como una región de la estructura principal o una región de punto caliente para un panel determinado. En la Tabla 5 se puede encontrar un ejemplo de una lista de ubicaciones genómicas de punto caliente de interés. Las coordenadas de la Tabla 5 se basan en el ensamblaje hg19 del genoma humano, pero un experto en la materia estará familiarizado con otros ensamblajes y puede identificar conjuntos de coordenadas correspondientes a los exones, intrones, codones, etc. indicados en un ensamblaje de su elección. En algunas formas de realización, un conjunto de regiones objetivo de secuencia variable utilizado en los métodos de la presente divulgación comprende al menos una porción de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19 o al menos 20 de los genes de la Tabla 5. Cada región genómica de punto caliente se enumera con varias características, incluido el gen asociado, el cromosoma en el que reside, la posición de inicio y finalización del genoma que representa el locus del gen, la longitud del locus del gen en pares de bases, los exones cubiertos por el gen y la característica crítica (por ejemplo, tipo de mutación) que una región genómica de interés dada puede intentar capturar.

Tabla 3

Mutaciones de punto (SNV) e indeles						Fusiones
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	CDKN2B	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	SRC	STK11	
TERT	TP53	TSC1	VHL			

Tabla 4

Mutaciones de punto (SNV) e indeles						Fusiones
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	DDR2	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	MAPK1	STK11	
TERT	TP53	TSC1	VHL	MAPK3	MTOR	
NTRK3						

Tabla 5

Gen	Cromosoma	Posición de inicio	Posición de parada	Longitud (bp)	Exones/intrones cubiertos	Características
ALK	chr2	29446405	29446655	250	intrón 19	Fusión
ALK	chr2	29446062	29446197	135	intrón 20	Fusión
ALK	chr2	29446198	29446404	206	exón 20	Fusión
ALK	chr2	29447353	29447473	120	intrón 19	Fusión
ALK	chr2	29447614	29448316	702	intrón 19	Fusión
ALK	chr2	29448317	29448441	124	exón 19	Fusión
ALK	chr2	29449366	29449777	411	intrón 18	Fusión
ALK	chr2	29449778	29449950	172	exón 18	Fusión
BRAF	chr7	140453064	140453203	139	exón 15	BRAF V600
CTNNB1	chr3	41266007	41266254	247	exón 3	S37
EGFR	chr7	55240528	55240827	299	exones 18 y 19	G719 y deleciones
EGFR	chr7	55241603	55241746	143	exón 20	Inserciones/T7 90M
EGFR	chr7	55242404	55242523	119	exón 21	L858R
ERBB2	chr17	37880952	37881174	222	exón 20	Inserciones
ESR1	chr6	152419857	152420111	254	exón 10	V534, P535, L536, Y537, D538
FGFR2	chr10	123279482	123279693	211	exón 6	S252
GATA3	chr10	8111426	8111571	145	exón 5	SS / Indeles
GATA3	chr10	8115692	8116002	310	exón 6	SS / Indeles
GNAS	chr20	57484395	57484488	93	exón 8	R844
IDH1	chr2	209113083	209113394	311	exón 4	R132
IDH2	chr15	90631809	90631989	180	exón 4	R140, R172
KIT	chr4	55524171	55524258	87	exón 1	
KIT	chr4	55561667	55561957	290	exón 2	
KIT	chr4	55564439	55564741	302	exón 3	
KIT	chr4	55565785	55565942	157	exón 4	
KIT	chr4	55569879	55570068	189	exón 5	
KIT	chr4	55573253	55573463	210	exón 6	
KIT	chr4	55575579	55575719	140	exón 7	
KIT	chr4	55589739	55589874	135	exón 8	
KIT	chr4	55592012	55592226	214	exón 9	
KIT	chr4	55593373	55593718	345	exones 10 y 11	557, 559, 560, 576
KIT	chr4	55593978	55594297	319	exones 12 y 13	V654
KIT	chr4	55595490	55595661	171	exón 14	T670, S709
KIT	chr4	55597483	55597595	112	exón 15	D716
KIT	chr4	55598026	55598174	148	exón 16	L783
KIT	chr4	55599225	55599368	143	exón 17	C809, R815, D816, L818, D820, S821F, N822, Y823

(Continuación)

Gen	Cromosoma	Posición de inicio	Posición de parada	Longitud (bp)	Exones/intrones cubiertos	Características
KIT	chr4	55602653	55602785	132	exón 18	A829P
KIT	chr4	55602876	55602996	120	exón 19	
KIT	chr4	55603330	55603456	126	exón 20	
KIT	chr4	55604584	55604733	149	exón 21	
KRAS	chr12	25378537	25378717	180	exón 4	A146
KRAS	chr12	25380157	25380356	199	exón 3	Q61
KRAS	chr12	25398197	25398328	131	exón 2	G12/G13
MET	chr7	116411535	116412255	720	exón 13, exón14, intrón 13, intrón 14	MET exón 14 SS
NRAS	chr1	115256410	115256609	199	exón 3	Q61
NRAS	chr1	115258660	115258791	131	exón 2	G12/G13
PIK3CA	chr3	178935987	178936132	145	exón 10	E545K
PIK3CA	chr3	178951871	178952162	291	exón 21	H1047R
PTEN	chr10	89692759	89693018	259	exón 5	R130
SMAD4	chr18	48604616	48604849	233	exón 12	D537
TERT	chr5	1294841	1295512	671	promotor	chr5:1295228
TP53	chr17	7573916	7574043	127	exón 11	Q331, R337, R342
TP53	chr17	7577008	7577165	157	exón 8	R273
TP53	chr17	7577488	7577618	130	exón 7	R248
TP53	chr17	7578127	7578299	172	exón 6	R213/Y220
TP53	chr17	7578360	7578564	204	exón 5	R175 / Deleciones
TP53	chr17	7579301	7579600	299	exón 4	
				12574 (region objetivo total)		
				16330 (cobertura de sonda total)		

[0084] Adicionalmente, o como alternativa, hay conjuntos de regiones objetivo adecuados disponibles en la literatura. Por ejemplo, Gale et al., PLoS One 13: e0194630 (2018) describe un panel de 35 dianas genéticas relacionadas con el cáncer que se pueden usar como parte o la totalidad de un conjunto de regiones objetivo de secuencia variable. Estos 35 objetivos son AKT1, ALK, BRAF, CCND1, CDK2A, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FOXL2, GATA3, GNA11, GNAQ, GNAS, HRAS, IDH1, IDH2, KIT, KRAS, MED12, MET, MYC, NFE2L2, NRAS, PDGFRA, PIK3CA, PPP2RIA, PTEN, RET, STK11, TP53 y U2AF1.

[0085] En algunas formas de realización, el conjunto de regiones objetivo de secuencia variable comprende regiones objetivo de al menos 10, 20, 30 o 35 genes relacionados con el cáncer, tales como los genes relacionados con el cáncer enumerados anteriormente.

3. Partición; Análisis de características epigenéticas

[0086] En ciertas formas de realización descritas en el presente documento, el ADN de una muestra (por ejemplo, que comprende una población de diferentes formas de ácidos nucleicos, tales como ADN hipermetilado e hipometilado) se puede particionar físicamente en función de una o más características de los ácidos nucleicos antes del análisis, por ejemplo, secuenciación o etiquetado y secuenciación. Este enfoque se puede utilizar para determinar, por ejemplo, si las regiones objetivo epigenéticas variables de hipermetilación muestran una hipermetilación característica de las células tumorales o las regiones objetivo epigenéticas variables de hipometilación muestran una hipometilación característica de

las células tumorales. Además, al particionar una población de ácidos nucleicos heterogéneos, se pueden aumentar las señales raras, por ejemplo, al enriquecer las moléculas de ácidos nucleicos raras que son más prevalentes en una fracción (o partición) de la población. Por ejemplo, una variación genética presente en el ADN hipermetilado, pero menos (o no) en el ADN hipometilado se puede detectar más fácilmente dividiendo una muestra en moléculas de ácido nucleico hipermetiladas e hipometiladas. Al analizar múltiples fracciones de una muestra, se puede realizar un análisis multidimensional de un solo locus de un genoma o especie de ácido nucleico y, por lo tanto, se puede lograr una mayor sensibilidad.

[0087] En algunos casos, una muestra de ácido nucleico heterogéneo se divide en dos o más particiones (por ejemplo, al menos 3, 4, 5, 6 o 7 particiones). En algunas formas de realización, cada partición se etiqueta de forma diferencial. Las particiones etiquetadas se pueden agrupar para la preparación y/o secuenciación colectiva de la muestra. Los pasos de partición-etiquetado-agrupación pueden ocurrir más de una vez, y cada ronda de partición se produce en función de diferentes características (se proporcionan ejemplos en el presente documento) y se etiqueta utilizando etiquetas diferenciales que se distinguen de otras particiones y medios de partición. En algunas formas de realización, el ADN de una muestra se divide en al menos una partición hipermetilada y una partición hipometilada, y al menos una de esas particiones se divide además en función de una característica adicional. En algunas formas de realización, el ADN de una muestra se divide en al menos dos particiones en función de una primera característica distinta de la metilación, y al menos una de esas particiones se divide además en función de la metilación. Como tal, la muestra a la que se hace referencia en la etapa de partición de los métodos descritos en el presente documento que genera al menos una partición hipermetilada y una partición hipometilada puede ser o comprender una partición de una etapa de partición anterior que utiliza una característica distinta del nivel de metilación. En algunas formas de realización, la característica adicional o la característica distinta de la metilación comprende la unión a una proteína específica, modificaciones químicas adicionales, tamaño de fragmento o cualquier combinación de las mismas. La partición basada en la unión a una proteína específica se puede realizar utilizando, por ejemplo, un procedimiento de extracción con anticuerpos biotinilados para la proteína específica. Las modificaciones químicas adicionales (por ejemplo, modificaciones que son adicionales en relación con la 5-metilación de la citosina) incluyen, por ejemplo, la 5-hidroximetilación de la citosina o la N6-metilación de la adenina.

[0088] Ejemplos de características que pueden usarse para la partición incluyen longitud de secuencia, nivel de metilación, unión de nucleosomas, desajuste de secuencia, inmunoprecipitación y/o proteínas que se unen al ADN. Las particiones resultantes pueden incluir una o más de las siguientes formas de ácido nucleico: ADN monocatenario (ADNmc), ADN bicatenario (ADNbc), fragmentos de ADN más cortos y fragmentos de ADN más largos. En algunas formas de realización, una población heterogénea de ácidos nucleicos se divide en ácidos nucleicos con una o más modificaciones epigenéticas y sin una o más modificaciones epigenéticas. Ejemplos de modificaciones epigenéticas incluyen presencia o ausencia de metilación; nivel de metilación; tipo de metilación (por ejemplo, 5-metilcitosina frente a otros tipos de metilación, como metilación de adenina y/o hidroximetilación de citosina); y asociación y nivel de asociación con una o más proteínas, como histonas. Como alternativa, o adicionalmente, una población heterogénea de ácidos nucleicos puede dividirse en moléculas de ácidos nucleicos asociadas con nucleosomas y moléculas de ácidos nucleicos desprovistas de nucleosomas. Como alternativa, o adicionalmente, una población heterogénea de ácidos nucleicos puede dividirse en ADN monocatenario (ADNmc) y ADN bicatenario (ADNbc). Como alternativa, o adicionalmente, una población heterogénea de ácidos nucleicos puede dividirse en función de la longitud del ácido nucleico (por ejemplo, moléculas de hasta 160 pb y moléculas que tienen una longitud de más de 160 pb).

[0089] En algunos casos, cada partición (representativa de una forma de ácido nucleico diferente) está etiquetada de manera diferencial, y las particiones se agrupan juntas en parte o para proporcionar al menos un primer y un segundo grupo antes de la captura, como se analiza en otra parte del presente documento.

[0090] La FIG. 1A ilustra una forma de realización de la divulgación. Una muestra que comprende una población de diferentes ácidos nucleicos (101) se divide (102) en dos o más particiones diferentes (103a, b). Cada partición (103a, b) es representativa de una forma de ácido nucleico diferente, por ejemplo, ADN hipometilado e hipermetilado. Cada partición está marcada de forma distinta (104), lo que proporciona particiones marcadas (105a, b). Los ácidos nucleicos marcados se agrupan para proporcionar un primer grupo (106a) que comprende al menos una parte del ADN de la primera partición (por ejemplo, hipometilada) y un segundo grupo (106b) que comprende al menos una parte del ADN de la segunda partición (por ejemplo, hipermetilada) y, opcionalmente, como se indica mediante la flecha discontinua, al menos una parte de la primera partición (por ejemplo, hipometilada). Las regiones objetivo de secuencia variable (SVTR) se capturan del primer grupo, y las SVTR y las regiones objetivo epigenéticas (ETR) se capturan del segundo grupo. El ADN capturado se puede combinar en esta etapa si se desea, y se puede procesar para un análisis posterior, por ejemplo, incluyendo la secuenciación, como se describe en otra parte del presente documento. En algunas formas de realización, las lecturas se analizan in silico y se utilizan etiquetas para clasificar las lecturas de diferentes particiones. El análisis para detectar variantes genéticas se puede realizar a nivel de partición por partición, así como a nivel de población de ácidos nucleicos completos. Por ejemplo, el análisis puede incluir análisis in silico para determinar variantes genéticas, como CNV, SNV, indel, fusión en ácidos nucleicos en cada partición. En otro ejemplo, el análisis puede incluir un análisis in silico para determinar variantes genéticas, como CNV, SNV, indel, fusión en ácidos nucleicos en datos de una partición hipometilada, y se pueden determinar características epigenéticas en datos de una o más particiones (por ejemplo, se pueden analizar regiones objetivo variables de fragmentación en datos de una pluralidad de particiones, como una partición hipermetilada, una partición hipometilada y, opcionalmente, una partición intermedia). En algunos casos, el análisis in silico puede incluir la determinación de la estructura de la cromatina. Por ejemplo, se puede utilizar la cobertura de lecturas de secuencias

para determinar la posición de los nucleosomas en la cromatina. Una cobertura más alta puede correlacionarse con una mayor ocupación de nucleosomas en la región genómica, mientras que una cobertura más baja puede correlacionarse con una menor ocupación de nucleosomas o una región depletada de nucleosomas (NDR).

5 **[0091]** La FIG. 1B ilustra otra forma de realización de la divulgación. Una muestra que comprende una población de diferentes ácidos nucleicos (151) se divide (152) en dos o más particiones diferentes (153a, b). Cada partición (153a, b) es representativa de una forma de ácido nucleico diferente, por ejemplo, ADN hipometilado e hipermetilado. En algunas formas de realización, se obtiene una tercera partición (por ejemplo, partición intermedia) (no se muestra). Opcionalmente, cada partición se etiqueta de forma distinta (154), lo que proporciona particiones etiquetadas (155a, b). Los ácidos nucleicos divididos, opcionalmente etiquetados, se agrupan para proporcionar un primer grupo (156a) que comprende al menos una parte del ADN de la primera partición (por ejemplo, hipometilada) y un segundo grupo (156b) que comprende al menos una parte del ADN de la segunda partición (por ejemplo, hipermetilada). En algunas formas de realización, el segundo grupo también puede comprender una partición intermedia. En algunas formas de realización, el primer grupo puede comprender ADN de la primera partición (por ejemplo, hipometilada) solamente. En algunas formas de realización, el segundo grupo puede comprender ADN de la segunda partición (por ejemplo, hipermetilada) solamente. En algunas formas de realización, el segundo grupo puede comprender ADN de la segunda partición (por ejemplo, hipermetilada) y la tercera partición (por ejemplo, partición intermedia) solamente. Se captura un primer conjunto de regiones objetivo del primer grupo, y se captura un segundo conjunto de regiones objetivo del segundo grupo. El ADN capturado (por ejemplo, si está etiquetado de forma diferencial) se puede combinar en esta etapa si se desea, o los grupos se pueden manejar por separado. En algunas formas de realización, las moléculas en cada partición se etiquetan de forma diferencial de modo que los códigos de barras moleculares utilizados para etiquetar moléculas en una partición sean diferentes de los códigos de barras moleculares utilizados en la(s) otra(s) partición(es). En algunas formas de realización, la partición se puede identificar por la etiqueta de partición, en donde todas las moléculas que pertenecen a una partición particular se pueden identificar a partir de la etiqueta de partición. Las moléculas etiquetadas se pueden agrupar juntas. En algunas formas de realización, las particiones no se agrupan juntas, sino que se manejan por separado. En tales formas de realización, las moléculas pueden etiquetarse con códigos de barras moleculares que identifican moléculas individuales; dichos códigos de barras moleculares no necesitan ser específicos de la partición. En cualquier caso, un análisis adicional puede incluir, por ejemplo, secuenciación, como se describe en otra parte del presente documento. En algunas formas de realización, las lecturas se analizan in silico y las etiquetas se utilizan para clasificar las lecturas de diferentes particiones. El análisis para detectar variantes genéticas se puede realizar a nivel de partición por partición, así como a nivel de población de ácidos nucleicos completos. Por ejemplo, el análisis puede incluir análisis in silico para determinar variantes genéticas, tales como CNV, SNV, indel, fusión en ácidos nucleicos en cada partición. En otro ejemplo, el análisis puede incluir análisis in silico para determinar variantes genéticas, tales como CNV, SNV, indel, fusión en ácidos nucleicos en datos de una partición hipometilada, y las características epigenéticas se pueden determinar en datos de una o más particiones (por ejemplo, las regiones objetivo variables de fragmentación se pueden analizar en datos de una pluralidad de particiones tales como una partición hipermetilada, una partición hipometilada y, opcionalmente, una partición intermedia). En algunos casos, el análisis in silico puede incluir la determinación de la estructura de la cromatina. Por ejemplo, la cobertura de las lecturas de secuencias se puede utilizar para determinar la posición de los nucleosomas en la cromatina. Una cobertura más alta puede correlacionarse con una mayor ocupación de nucleosomas en la región genómica, mientras que una cobertura más baja puede correlacionarse con una menor ocupación de nucleosomas o una región depletada de nucleosomas (NDR).

45 **[0092]** Como ejemplo de un flujo de trabajo en el que los grupos se manejan por separado, una pluralidad de muestras (por ejemplo, 96 muestras) se pueden dividir en particiones hipometiladas e hipermetiladas, y cada muestra recibe una etiqueta que comprende un índice específico de la muestra (o par de índices) y, opcionalmente, un código de barras molecular para preparar bibliotecas separadas. Las bibliotecas hipometiladas se pueden secuenciar juntas, por ejemplo, en una primera célula de flujo, y las bibliotecas hipermetiladas se pueden secuenciar juntas, por ejemplo, en una segunda célula de flujo. Después de adquirir datos de secuencia, las fuentes de muestra se podrían hacer coincidir fácilmente entre las células de flujo utilizando el índice específico de la muestra o el par de índices. En algunas formas de realización, se utilizan diferentes cantidades de carga para las 2 células de flujo (por ejemplo, secuenciar solo el 10 % del material hipometilado, pero todo o sustancialmente todo el material hipermetilado o secuenciar solo el 10 % del material hipermetilado, pero todo o sustancialmente todo el material hipometilado).

55 **[0093]** En general, las muestras pueden incluir ácidos nucleicos que varían en modificaciones que incluyen modificaciones posteriores a la replicación de los nucleótidos y unión, generalmente de forma no covalente, a una o más proteínas.

60 **[0094]** En una forma de realización, la población de ácidos nucleicos es una obtenida de una muestra de suero, plasma o sangre de un sujeto sospechoso de tener neoplasia, un tumor o cáncer o previamente diagnosticado con neoplasia, un tumor o cáncer. La población de ácidos nucleicos incluye ácidos nucleicos que tienen niveles variables de metilación. La metilación puede ocurrir a partir de una o más modificaciones post-replicación o transcripcionales. Las modificaciones post-replicación incluyen modificaciones del nucleótido citosina, particularmente en la posición 5 de la nucleobase, por ejemplo, 5-metilcitosina, 5-hidroximetilcitosina, 5-formilcitosina y 5-carboxilcitosina.

65 **[0095]** En algunas formas de realización, los ácidos nucleicos en la población original pueden ser monocatenarios y/o bicatenarios. La partición basada en la monocatenaria o la bicatenaria de los ácidos nucleicos se puede lograr, por

ejemplo, utilizando sondas de captura marcadas para particionar el ADNmc y utilizando adaptadores bicatenarios para particionar el ADNbc.

[0096] Los agentes de afinidad pueden ser anticuerpos con la especificidad deseada, socios de unión naturales o variantes de los mismos (Bock et al., Nat Biotech 28: 1106-1114 (2010); Song et al., Nat Biotech 29: 68-72 (2011)), o péptidos artificiales seleccionados, por ejemplo, mediante visualización de fagos para tener especificidad para un objetivo determinado.

[0097] Los ejemplos de fracciones de captura contempladas en este documento incluyen dominios de unión de metilo (MBD) y proteínas de unión de metilo (MBP) como se describe en este documento.

[0098] Asimismo, se puede realizar la partición de diferentes formas de ácidos nucleicos utilizando proteínas de unión a histonas que pueden separar los ácidos nucleicos unidos a histonas de los ácidos nucleicos libres o no unidos. Los ejemplos de proteínas de unión a histonas que se pueden utilizar en los métodos descritos en el presente documento incluyen RBBP4 (RbAp48) y péptidos de dominio SANT.

[0099] Aunque para algunos agentes de afinidad y modificaciones, la unión al agente puede ocurrir de una manera esencialmente total o nula dependiendo de si un ácido nucleico tiene una modificación, la separación puede ser de grado. En tales casos, los ácidos nucleicos sobrerrepresentados en una modificación se unen al agente en mayor medida que los ácidos nucleicos subrepresentados en la modificación. Alternativamente, los ácidos nucleicos que tienen modificaciones pueden unirse de una manera total o nula. Pero entonces, varios niveles de modificaciones pueden eluirse secuencialmente del agente de unión.

[0100] Por ejemplo, en algunas formas de realización, la partición puede ser binaria o basarse en el grado/nivel de modificaciones. Por ejemplo, todos los fragmentos metilados pueden ser particionados a partir de fragmentos no metilados utilizando proteínas de dominio de unión a metilo (por ejemplo, MethylMiner Methylated DNA Enrichment Kit (Thermo Fisher Scientific). Posteriormente, la partición adicional puede implicar la elución de fragmentos que tienen diferentes niveles de metilación ajustando la concentración de sal en una solución con el dominio de unión a metilo y los fragmentos unidos. A medida que aumenta la concentración de sal, se eluyen los fragmentos que tienen mayores niveles de metilación.

[0101] En algunos casos, las particiones finales son representantes de ácidos nucleicos que tienen diferentes grados de modificaciones (sobrerrepresentación o subrepresentación de modificaciones). La sobrerrepresentación y la subrepresentación pueden definirse por el número de modificaciones que presenta un ácido nucleico en relación con el número medio de modificaciones por hebra en una población. Por ejemplo, si el número medio de residuos de 5-metilcitosina en el ácido nucleico de una muestra es 2, un ácido nucleico que incluye más de dos residuos de 5-metilcitosina está sobrerrepresentado en esta modificación y un ácido nucleico con 1 o cero residuos de 5-metilcitosina está subrepresentado. El efecto de la separación por afinidad es enriquecer los ácidos nucleicos sobrerrepresentados en una modificación en una fase unida y los ácidos nucleicos subrepresentados en una modificación en una fase no unida (es decir, en solución). Los ácidos nucleicos en la fase unida pueden eluirse antes del procesamiento posterior.

[0102] Cuando se utiliza el kit de enriquecimiento de ADN metilado MethylMiner (Thermo Fisher Scientific), se pueden dividir varios niveles de metilación utilizando eluciones secuenciales. Por ejemplo, una partición hipometilada (por ejemplo, sin metilación) se puede separar de una partición metilada poniendo en contacto la población de ácidos nucleicos con el MBD del kit, que está unido a perlas magnéticas. Las perlas se utilizan para separar los ácidos nucleicos metilados de los ácidos nucleicos no metilados. Posteriormente, se realizan uno o más pasos de elución de forma secuencial para eluir los ácidos nucleicos que tienen diferentes niveles de metilación. Por ejemplo, un primer conjunto de ácidos nucleicos metilados se puede eluir a una concentración de sal de 160 mM o superior, por ejemplo, al menos 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 700 mM, 800 mM, 900 mM, 1.000 mM o 2.000 mM. Después de que se eluyen dichos ácidos nucleicos metilados, se utiliza una vez más la separación magnética para separar un nivel más alto de ácidos nucleicos metilados de aquellos con un nivel más bajo de metilación. Los pasos de elución y separación magnética se pueden repetir para crear varias particiones, como una partición hipometilada (por ejemplo, representativa de la ausencia de metilación), una partición metilada (representativa de un nivel bajo de metilación) y una partición hipermetilada (representativa de un nivel alto de metilación).

[0103] En algunos métodos, los ácidos nucleicos unidos a un agente utilizado para la separación por afinidad se someten a una etapa de lavado. La etapa de lavado elimina los ácidos nucleicos débilmente unidos al agente de afinidad. Dichos ácidos nucleicos pueden enriquecerse en ácidos nucleicos que tienen la modificación en un grado cercano a la media o mediana (es decir, intermedio entre los ácidos nucleicos que permanecen unidos a la fase sólida y los ácidos nucleicos que no se unen a la fase sólida en el contacto inicial de la muestra con el agente).

[0104] La separación por afinidad da como resultado al menos dos, y a veces tres o más particiones de ácidos nucleicos con diferentes grados de una modificación. Mientras las particiones siguen estando separadas, los ácidos nucleicos de al menos una partición, y normalmente dos o tres (o más) particiones están unidos a etiquetas de ácidos nucleicos, normalmente proporcionadas como componentes de adaptadores, y los ácidos nucleicos en diferentes particiones reciben diferentes etiquetas que distinguen a los miembros de una partición de otra. Las etiquetas unidas a moléculas de ácidos

nucleicos de la misma partición pueden ser iguales o diferentes entre sí. Pero si son diferentes entre sí, las etiquetas pueden tener parte de su código en común para identificar las moléculas a las que están unidas como pertenecientes a una partición particular.

5 **[0105]** En algunas formas de realización, los métodos y sistemas utilizados para dividir muestras de ácidos nucleicos en función de características tales como la metilación se pueden encontrar en el documento WO2018/119452. Además, métodos, sistemas y composiciones ejemplares útiles para aislar ácidos nucleicos se pueden encontrar en el documento WO2020/160414.

10 **[0106]** Para obtener más detalles sobre la división en porciones de muestras de ácido nucleico en función de características como la metilación, consulte el documento WO2018/119452.

[0107] En algunas formas de realización, las moléculas de ácido nucleico se pueden fraccionar en diferentes particiones en función de las moléculas de ácido nucleico que están unidas a una proteína específica o un fragmento de la misma y aquellas que no están unidas a esa proteína específica o fragmento de la misma.

15 **[0108]** Las moléculas de ácido nucleico se pueden fraccionar en función de la unión de proteína-ADN. Los complejos proteína-ADN se pueden fraccionar en función de una propiedad específica de una proteína. Los ejemplos de dichas propiedades incluyen varios epítomos, modificaciones (por ejemplo, metilación o acetilación de histonas) o actividad enzimática. Los ejemplos de proteínas que se pueden unir al ADN y servir como base para el fraccionamiento pueden incluir, entre otros, proteína A y proteína G. Se puede utilizar cualquier método adecuado para fraccionar las moléculas de ácido nucleico en función de las regiones unidas a proteínas. Los ejemplos de métodos utilizados para fraccionar moléculas de ácido nucleico en función de las regiones unidas a proteínas incluyen, entre otros, SDS-PAGE, inmunoprecipitación de cromatina (ChIP), cromatografía de heparina y fraccionamiento de flujo de campo asimétrico (AF4).

20 **[0109]** En algunas formas de realización, la partición de los ácidos nucleicos se realiza poniendo en contacto los ácidos nucleicos con un dominio de unión a la metilación ("MBD") de una proteína de unión a la metilación ("MBP"). El MBD se une a la 5-metilcitosina (5mC). El MBD se acopla a perlas paramagnéticas, como Dynabeads® M-280 Streptavidin a través de un enlazador de biotina. La partición en fracciones con diferentes grados de metilación se puede realizar eluyendo las fracciones aumentando la concentración de NaCl.

[0110] Los ejemplos de MBP contemplados en este documento incluyen, pero no se limitan a:

- 35 (a) MeCP2 es una proteína que se une preferentemente a 5-metil-citosina sobre citosina sin modificar.
 (b) RPL26, PRP8 y la proteína de reparación de desajustes de ADN MHS6 se unen preferentemente a 5-hidroximetil-citosina sobre citosina sin modificar.
 (c) FOXK1, FOXK2, FOXP1, FOXP4 y FOXI3 se unen preferentemente a 5-formil citosina sobre citosina sin modificar (Iurlaro et al., Genome Biol. 14: R119 (2013)).
 40 (d) Anticuerpos específicos para una o más bases de nucleótidos metiladas.

[0111] En general, la elución es una función del número de sitios metilados por molécula, y las moléculas que tienen más metilación eluyen con concentraciones de sal mayores. Para eluir el ADN en poblaciones distintas en función del grado de metilación, se puede utilizar una serie de tampones de elución con una concentración de NaCl creciente. La concentración de sal puede variar de aproximadamente 100 mM a aproximadamente 2500 mM de NaCl. En una forma de realización, el proceso da como resultado tres (3) particiones. Las moléculas se ponen en contacto con una solución a una primera concentración de sal y que comprende una molécula que comprende un dominio de unión a metilo, molécula que puede estar unida a una fracción de captura, como la estreptavidina. A la primera concentración de sal, una población de moléculas se unirá al MBD y una población permanecerá no unida. La población no unida puede separarse como una población "hipometilada". Por ejemplo, una primera partición representativa de la forma hipometilada de ADN es aquella que permanece sin unirse a una baja concentración de sal, por ejemplo, 100 mM o 160 mM. Una segunda partición representativa de ADN metilado intermedio se eluye utilizando una concentración de sal intermedia, por ejemplo, entre 100 mM y 2.000 mM de concentración. Esto también se separa de la muestra. Una tercera partición representativa de la forma hipermetilada de ADN se eluye utilizando una alta concentración de sal, por ejemplo, al menos aproximadamente 2.000 mM.

a. Etiquetado de particiones

60 **[0112]** En algunas formas de realización, dos o más particiones, por ejemplo, cada partición, está/n etiquetadas de forma diferencial. Las etiquetas pueden ser moléculas, tales como ácidos nucleicos, que contienen información que indica una característica de la molécula con la que está asociada la etiqueta. Por ejemplo, las moléculas pueden llevar una etiqueta de muestra (que distingue las moléculas en una muestra de las de una muestra diferente), una etiqueta de partición (que distingue las moléculas en una partición de las de una partición diferente) o una etiqueta molecular (que distingue diferentes moléculas entre sí (en escenarios de etiquetado tanto únicos como no únicos). En ciertas formas de realización, una etiqueta puede comprender uno o una combinación de códigos de barras. Como se usa en el presente documento, el término "código de barras" se refiere a una molécula de ácido nucleico que tiene una secuencia de nucleótidos particular,

o a la secuencia de nucleótidos en sí, dependiendo del contexto. Un código de barras puede tener, por ejemplo, entre 10 y 100 nucleótidos. Una colección de códigos de barras puede tener secuencias degeneradas o puede tener secuencias que tienen una cierta distancia de Hamming, según se desee para el propósito específico. Entonces, por ejemplo, un índice de muestra, índice de partición o índice molecular puede estar compuesto por un código de barras o una combinación de dos códigos de barras, cada uno unido a diferentes extremos de una molécula.

[0113] Se pueden utilizar etiquetas para etiquetar las particiones de población de polinucleótidos individuales de modo de correlacionar la etiqueta (o etiquetas) con una partición específica. Alternativamente, se pueden utilizar etiquetas en formas de realización de la invención que no emplean un paso de partición. En algunas formas de realización, se puede utilizar una sola etiqueta para etiquetar una partición específica. En algunas formas de realización, se pueden utilizar múltiples etiquetas diferentes para etiquetar una partición específica. En formas de realización que emplean múltiples etiquetas diferentes para etiquetar una partición específica, el conjunto de etiquetas utilizado para etiquetar una partición se puede diferenciar fácilmente del conjunto de etiquetas utilizado para etiquetar otras particiones. En algunas formas de realización, las etiquetas pueden tener funciones adicionales, por ejemplo, las etiquetas se pueden usar para indexar fuentes de muestra o se pueden usar como identificadores moleculares únicos (que se pueden usar para mejorar la calidad de los datos de secuenciación al diferenciar errores de secuenciación de mutaciones, por ejemplo, como en Kinde et al., Proc Nat'l Acad Sci USA 108: 9530-9535 (2011), Kou et al., PLoS ONE, 11: e0146638 (2016)) o se pueden usar como identificadores de moléculas no únicos, por ejemplo, como se describe en la Patente de EE. UU. N.º 9.598.731. De manera similar, en algunas formas de realización, las etiquetas pueden tener funciones adicionales, por ejemplo, las etiquetas se pueden usar para indexar fuentes de muestra o se pueden usar como identificadores moleculares no únicos (que se pueden usar para mejorar la calidad de los datos de secuenciación al diferenciar errores de secuenciación de mutaciones).

[0114] En una forma de realización, el etiquetado de particiones comprende el etiquetado de moléculas en cada partición con una etiqueta de partición. Después de volver a combinar particiones y secuenciar moléculas, las etiquetas de partición identifican la partición de origen. En otra forma de realización, se etiquetan diferentes particiones con diferentes conjuntos de etiquetas moleculares, por ejemplo, que comprenden un par de códigos de barras. De esta manera, cada código de barras molecular indica la partición de origen y además es útil para distinguir moléculas dentro de una partición. Por ejemplo, se puede utilizar un primer conjunto de 35 códigos de barras para etiquetar moléculas en una primera partición, mientras que se puede utilizar un segundo conjunto de 35 códigos de barras para etiquetar moléculas en una segunda partición.

[0115] En algunas formas de realización, después de la partición y el etiquetado con etiquetas de partición, las moléculas se pueden agrupar para la secuenciación en una sola ejecución. En algunas formas de realización, se añade una etiqueta de muestra a las moléculas, por ejemplo, en un paso posterior a la adición de etiquetas de partición y la agrupación. Las etiquetas de muestra pueden facilitar la agrupación de material generado a partir de múltiples muestras para la secuenciación en una sola ejecución de secuenciación.

[0116] Alternativamente, en algunas formas de realización, las etiquetas de partición pueden estar correlacionadas con la muestra, así como con la partición. Como un ejemplo simple, una primera etiqueta puede indicar una primera partición de una primera muestra; una segunda etiqueta puede indicar una segunda partición de la primera muestra; una tercera etiqueta puede indicar una primera partición de una segunda muestra; y una cuarta etiqueta puede indicar una segunda partición de la segunda muestra.

[0117] Aunque las etiquetas pueden estar unidas a moléculas ya divididas en función de una o más características, las moléculas marcadas finales en la biblioteca pueden ya no poseer esa característica. Por ejemplo, aunque las moléculas de ADN monocatenario pueden dividirse y marcarse, es probable que las moléculas marcadas finales en la biblioteca sean bicatenarias. De manera similar, aunque el ADN puede estar sujeto a división en función de diferentes niveles de metilación, en la biblioteca final, es probable que las moléculas marcadas derivadas de estas moléculas no estén metiladas. En consecuencia, la etiqueta unida a la molécula en la biblioteca indica típicamente la característica de la "molécula madre" de la que se deriva la molécula marcada final, no necesariamente la característica de la molécula marcada en sí.

[0118] Como ejemplo, los códigos de barras 1, 2, 3, 4, etc. se utilizan para marcar y etiquetar moléculas en la primera división; los códigos de barras A, B, C, D, etc. se utilizan para marcar y etiquetar moléculas en la segunda división; y los códigos de barras a, b, c, d, etc. se utilizan para marcar y etiquetar moléculas en la tercera división. Las particiones marcadas de forma diferencial se pueden agrupar antes de la secuenciación. Las particiones marcadas de forma diferencial se pueden secuenciar por separado o juntas al mismo tiempo, por ejemplo, en la misma célula de flujo de un secuenciador Illumina.

[0119] Después de la secuenciación, el análisis de las lecturas para detectar variantes genéticas se puede realizar en un nivel de partición por partición, así como en un nivel de población de ácidos nucleicos completo. Las etiquetas se utilizan para ordenar las lecturas de diferentes particiones. El análisis puede incluir análisis in silico para determinar la variación genética y epigenética (una o más de metilación, estructura de cromatina, etc.) utilizando información de secuencia, longitud de coordenadas genómicas, cobertura y/o número de copias. En algunas formas de realización, una cobertura más alta puede correlacionarse con una mayor ocupación de nucleosomas en la región genómica, mientras que una

cobertura más baja puede correlacionarse con una menor ocupación de nucleosomas o una región agotada de nucleosomas (NDR).

b. Determinación del patrón de 5-metilcitosina de ácidos nucleicos; secuenciación con bisulfito

[0120] La secuenciación basada en bisulfito y sus variantes proporcionan otro medio para determinar el patrón de metilación de un ácido nucleico que no depende de la partición basada en el nivel de metilación antes de la secuenciación. En algunas formas de realización, la determinación del patrón de metilación comprende distinguir la 5-metilcitosina (5mC) de la citosina no metilada. En algunas formas de realización, la determinación del patrón de metilación comprende distinguir la N-metiladenina de la adenina no metilada. En algunas formas de realización, la determinación del patrón de metilación comprende distinguir la 5-hidroximetilcitosina (5hmC), la 5-formilcitosina (5fC) y la 5-carboxilcitosina (5caC) de la citosina no metilada. Los ejemplos de secuenciación con bisulfito incluyen, entre otros, la secuenciación oxidativa con bisulfito (OX-BS-seq), la secuenciación con bisulfito asistida por Tet (TAB-seq) y la secuenciación reducida con bisulfito (redBS-seq). En algunas formas de realización, la determinación del patrón de metilación comprende la secuenciación de bisulfito del genoma completo, por ejemplo, como en MethylC-seq (Urich et al., Nature Protocols 10:475-483 (2015)). En algunas formas de realización, la determinación del patrón de metilación comprende la determinación del patrón de metilación basada en matrices, por ejemplo, como en Methylation EPIC Beadchip o el uso de matrices Illumina Infinium (por ejemplo, matrices HumanMethylation450) (véase The Cancer Genome Atlas Research Network, Nature 507:315-322 (2014)). En algunas formas de realización, la determinación del patrón de metilación comprende PCR con bisulfito. En algunas formas de realización, la determinación del patrón de metilación comprende EM-Seq (US 2013/0244237 A1). En algunas formas de realización, la determinación del patrón de metilación comprende TAPS (WO 2019/136413 A1).

[0121] La secuenciación oxidativa por bisulfito (OX-BS-seq) se utiliza para distinguir entre 5mC y 5hmC, convirtiendo primero 5hmC en 5fC y luego procediendo con la secuenciación por bisulfito. La secuenciación asistida por bisulfito Tet (TAB-seq) también se puede utilizar para distinguir 5mC y 5hmC. En la secuenciación TAB, 5hmC está protegido por glucosilación. A continuación, se utiliza una enzima Tet para convertir 5mC en 5caC antes de proceder con la secuenciación por bisulfito. La secuenciación reducida por bisulfito se utiliza para distinguir 5fC de las citosinas modificadas.

[0122] Generalmente, en la secuenciación con bisulfito, una muestra de ácido nucleico se divide en dos alícuotas y una alícuota se trata con bisulfito. El bisulfito convierte la citosina nativa y ciertos nucleótidos de citosina modificados (por ejemplo, 5-formilcitosina o 5-carboxilcitosina) en uracilo, mientras que otras citosinas modificadas (por ejemplo, 5-metilcitosina, 5-hidroximetilcitosina) no se convierten. La comparación de las secuencias de ácido nucleico de las moléculas de las dos alícuotas indica qué citosinas se convirtieron y cuáles no en uracilos. En consecuencia, se pueden determinar las citosinas que se modificaron y cuáles no. La división inicial de la muestra en dos alícuotas es desventajosa para las muestras que contienen solo pequeñas cantidades de ácidos nucleicos y/o compuestas de orígenes de células/tejidos heterogéneos, tales como fluidos corporales que contienen ADN libre de células.

[0123] Por consiguiente, en algunas formas de realización, la secuenciación con bisulfito se realiza sin dividir inicialmente una muestra en dos alícuotas, por ejemplo, de la siguiente manera. En algunas formas de realización, los ácidos nucleicos en una población están unidos a una fracción de captura, tal como cualquiera de las descritas en este documento, es decir, una etiqueta que puede capturarse o inmovilizarse. Después de la unión de las fracciones de captura a los ácidos nucleicos de la muestra, los ácidos nucleicos de la muestra sirven como plantillas para la amplificación. Después de la amplificación, las plantillas originales permanecen unidas a las fracciones de captura, pero los amplicones no están unidos a las fracciones de captura.

[0124] La fracción de captura se puede unir a ácidos nucleicos de muestra como un componente de un adaptador, que también puede proporcionar sitios de unión de cebadores de amplificación y/o secuenciación. En algunos métodos, los ácidos nucleicos de muestra se unen a adaptadores en ambos extremos, y ambos adaptadores llevan una fracción de captura. Preferiblemente, cualquier residuo de citosina en los adaptadores se modifica, como, por ejemplo, con 5-metilcitosina, para proteger contra la acción del bisulfito. En algunos casos, las fracciones de captura se unen a las plantillas originales mediante un enlace escindible (por ejemplo, destiobiotina-TEG fotoescindible o residuos de uracilo escindibles con la enzima USER™, Chem. Commun. (Camb). 51: 3266-3269 (2015)), en cuyo caso las fracciones de captura se pueden eliminar, si se desea.

[0125] Los amplicones se desnaturalizan y se ponen en contacto con un reactivo de afinidad para la etiqueta de captura. Las plantillas originales se unen al reactivo de afinidad mientras que las moléculas de ácido nucleico resultantes de la amplificación no lo hacen. Por lo tanto, las plantillas originales se pueden separar de las moléculas de ácido nucleico resultantes de la amplificación.

[0126] Después de la separación de las plantillas originales de las moléculas de ácido nucleico resultantes de la amplificación, las plantillas originales pueden someterse a un tratamiento con bisulfito. Alternativamente, los productos de amplificación pueden someterse a un tratamiento con bisulfito y la población de plantillas originales no. Después de dicho tratamiento, las poblaciones respectivas pueden amplificarse (lo que en el caso de la población de plantillas originales convierte los uracilos en timinas). Las poblaciones también pueden someterse a una hibridación de sonda de biotina para su captura. A continuación, se analizan las poblaciones respectivas y se comparan las secuencias para determinar qué

citiosinas estaban 5-metiladas (o 5-hidroximetiladas) en la muestra original. La detección de un nucleótido T en la población de plantillas (que corresponde a una citosina no metilada convertida en uracilo) y un nucleótido C en la posición correspondiente de la población amplificada indica una C no modificada. La presencia de C en las posiciones correspondientes de la plantilla original y las poblaciones amplificadas indica una C modificada en la muestra original.

[0127] En algunas formas de realización, un método utiliza la preparación de bibliotecas NGS secuenciales de ADN-seq y bisulfito-seq (BIS-seq) de bibliotecas de ADN etiquetadas molecularmente (véase WO 2018/119452, p. ej., en la FIG. 4). Este proceso se realiza mediante el etiquetado de adaptadores (p. ej., biotina), amplificación de ADN-seq de toda la biblioteca, recuperación de la molécula original (p. ej., extracción de perlas de estreptavidina), conversión de bisulfito y BIS-seq. En algunas formas de realización, el método identifica 5-metilcitosina con resolución de base única, a través de la amplificación preparativa secuencial de NGS de moléculas de la biblioteca original con y sin tratamiento con bisulfito. Esto se puede lograr modificando los adaptadores NGS 5-metilados (adaptadores direccionales; en forma de Y/bifurcados con 5-metilcitosina reemplazando) utilizados en BIS-seq con una etiqueta (p. ej., biotina) en una de las dos cadenas del adaptador. Las moléculas de ADN de muestra se ligan con un adaptador y se amplifican (por ejemplo, mediante PCR). Como solo las moléculas progenitoras tendrán un extremo adaptador marcado, se pueden recuperar selectivamente de su progenie amplificada mediante métodos de captura específicos de la etiqueta (por ejemplo, perlas magnéticas de estreptavidina). Como las moléculas progenitoras retienen marcas de 5-metilación, la conversión con bisulfito en la biblioteca capturada producirá un estado de 5-metilación con una resolución de base única en la secuenciación BIS, lo que retendrá la información molecular en la secuenciación de ADN correspondiente. En algunas formas de realización, la biblioteca tratada con bisulfito se puede combinar con una biblioteca no tratada antes de la captura/NGS mediante la adición de una secuencia de ADN de etiqueta de muestra en un flujo de trabajo de NGS multiplexado estándar. Al igual que con los flujos de trabajo de secuenciación BIS, se puede realizar un análisis bioinformático para la alineación genómica y la identificación de bases 5-metiladas. En resumen, este método proporciona la capacidad de recuperar selectivamente las moléculas parentales ligadas, que llevan marcas de 5-metilcitosina, después de la amplificación de la biblioteca, lo que permite el procesamiento paralelo del ADN convertido con bisulfito. Esto supera la naturaleza destructiva del tratamiento con bisulfito en la calidad/sensibilidad de la información de secuenciación de ADN extraída de un flujo de trabajo. Con este método, las moléculas de ADN parental ligadas recuperadas (a través de adaptadores marcados) permiten la amplificación de la biblioteca de ADN completa y la aplicación paralela de tratamientos que provocan modificaciones epigenéticas del ADN. La presente divulgación analiza el uso de métodos BIS-seq para identificar la 5-metilación de citosina (5-metilcitosina), pero el uso de métodos BIS-seq no es necesario en muchas formas de realización. Se han desarrollado variantes de BIS-seq para identificar citosinas hidroximetiladas (5hmC; OX-BS-seq, TAB-seq), formilcitosina (5fC; redBS-seq) y citosinas carboxiladas. Estas metodologías se pueden implementar con la preparación de biblioteca secuencial/paralela descrita en este documento.

c. Métodos alternativos de análisis de ácidos nucleicos modificados

[0128] En algunos de estos métodos, una población de ácidos nucleicos que tienen la modificación en diferentes grados (por ejemplo, 0, 1, 2, 3, 4, 5 o más grupos metilo por molécula de ácido nucleico) se pone en contacto con adaptadores antes del fraccionamiento de la población dependiendo del grado de la modificación. Los adaptadores se unen a un extremo o a ambos extremos de las moléculas de ácido nucleico en la población. Preferiblemente, los adaptadores incluyen diferentes etiquetas en cantidades suficientes para que la cantidad de combinaciones de etiquetas resulte en una probabilidad baja, por ejemplo, 95, 99 o 99,9 % de que dos ácidos nucleicos con los mismos puntos de inicio y detención reciban la misma combinación de etiquetas. Después de la unión de los adaptadores, los ácidos nucleicos se amplifican a partir de cebadores que se unen a los sitios de unión de cebadores dentro de los adaptadores. Los adaptadores, ya sea que tengan las mismas etiquetas o diferentes, pueden incluir los mismos sitios de unión de cebadores o diferentes, pero preferiblemente los adaptadores incluyen el mismo sitio de unión de cebadores. Después de la amplificación, los ácidos nucleicos se ponen en contacto con un agente que se une preferentemente a los ácidos nucleicos que tienen la modificación (como los agentes descritos anteriormente). Los ácidos nucleicos se separan en al menos dos particiones que difieren en el grado en que los ácidos nucleicos que tienen la modificación se unen a los agentes. Por ejemplo, si el agente tiene afinidad por los ácidos nucleicos que tienen la modificación, los ácidos nucleicos sobrerrepresentados en la modificación (en comparación con la representación media en la población) se unen preferentemente al agente, mientras que los ácidos nucleicos subrepresentados para la modificación no se unen o se eluyen más fácilmente del agente. Después de la separación, las diferentes particiones pueden someterse a pasos de procesamiento adicionales, que normalmente incluyen una mayor amplificación y un análisis de secuencia, en paralelo, pero por separado. A continuación, se pueden comparar los datos de secuencia de las diferentes particiones.

[0129] Tal esquema de separación se puede realizar utilizando el siguiente procedimiento ejemplar. Los ácidos nucleicos se unen en ambos extremos a adaptadores en forma de Y que incluyen sitios de unión de cebadores y etiquetas. Las moléculas se amplifican. Las moléculas amplificadas se fraccionan luego por contacto con un anticuerpo que se une preferentemente a 5-metilcitosina para producir dos particiones. Una partición incluye moléculas originales que carecen de metilación y copias de amplificación que han perdido la metilación. La otra partición incluye moléculas de ADN original con metilación. Las dos particiones se procesan y secuencian luego por separado con una amplificación adicional de la partición metilada. Los datos de secuencia de las dos particiones se pueden comparar luego. En este ejemplo, las etiquetas no se utilizan para distinguir entre ADN metilado y no metilado sino más bien para distinguir entre diferentes moléculas dentro de estas particiones de modo que se pueda determinar si las lecturas con los mismos puntos de inicio y detención se basan en las mismas o diferentes moléculas.

[0130] La divulgación proporciona métodos adicionales para analizar una población de ácido nucleico en la que al menos algunos de los ácidos nucleicos incluyen uno o más residuos de citosina modificados, tales como 5-metilcitosina y cualquiera de las otras modificaciones descritas anteriormente. En estos métodos, la población de ácidos nucleicos se pone en contacto con adaptadores que incluyen uno o más residuos de citosina modificados en la posición 5C, tales como 5-metilcitosina. Preferiblemente, todos los residuos de citosina en dichos adaptadores también están modificados, o todas esas citosinas en una región de unión de cebador de los adaptadores están modificadas. Los adaptadores se unen a ambos extremos de las moléculas de ácido nucleico en la población. Preferiblemente, los adaptadores incluyen diferentes etiquetas en cantidades suficientes para que la cantidad de combinaciones de etiquetas resulte en una probabilidad baja, por ejemplo, 95, 99 o 99,9 % de que dos ácidos nucleicos con los mismos puntos de inicio y detención reciban la misma combinación de etiquetas. Los sitios de unión de cebador en dichos adaptadores pueden ser iguales o diferentes, pero preferiblemente son los mismos. Después de la unión de los adaptadores, los ácidos nucleicos se amplifican a partir de cebadores que se unen a los sitios de unión de cebadores de los adaptadores. Los ácidos nucleicos amplificados se dividen en una primera y una segunda alícuota. La primera alícuota se analiza para obtener datos de secuencia con o sin procesamiento adicional. Los datos de secuencia de la primera alícuota se determinan así independientemente del estado de metilación inicial de las moléculas de ácido nucleico. Las moléculas de ácido nucleico en la segunda alícuota se tratan con bisulfito. Este tratamiento convierte las citosinas no modificadas en uracilos. Los ácidos nucleicos tratados con bisulfito se someten a continuación a amplificación cebada por cebadores en los sitios de unión de cebadores originales de los adaptadores unidos al ácido nucleico. Ahora sólo las moléculas de ácido nucleico originalmente unidas a los adaptadores (a diferencia de los productos de amplificación de los mismos) son amplificables porque estos ácidos nucleicos retienen las citosinas en los sitios de unión de cebadores de los adaptadores, mientras que los productos de amplificación han perdido la metilación de estos residuos de citosina, que han sufrido una conversión en uracilos en el tratamiento con bisulfito. De esta forma, sólo las moléculas originales de las poblaciones, de las cuales al menos algunas están metiladas, se amplifican. Después de la amplificación, estos ácidos nucleicos se someten a un análisis de secuencia. La comparación de las secuencias determinadas a partir de la primera y la segunda alícuota puede indicar, entre otras cosas, qué citosinas de la población de ácidos nucleicos se han metilado.

[0131] Tal análisis se puede realizar utilizando el siguiente procedimiento ejemplar. El ADN metilado se une a adaptadores en forma de Y en ambos extremos que incluyen sitios de unión de cebadores y etiquetas. Las citosinas en los adaptadores están 5-metiladas. La metilación de los cebadores sirve para proteger los sitios de unión de cebadores en un paso de bisulfito posterior. Después de la unión de los adaptadores, las moléculas de ADN se amplifican. El producto de amplificación se divide en dos alícuotas para secuenciación con y sin tratamiento con bisulfito. La alícuota no sometida a secuenciación con bisulfito se puede someter a análisis de secuencia con o sin procesamiento adicional. La otra alícuota se trata con bisulfito, que convierte las citosinas no metiladas en uracilos. Solo los sitios de unión de cebadores protegidos por la metilación de citosinas pueden soportar la amplificación cuando se ponen en contacto con cebadores específicos para los sitios de unión de cebadores originales. Por lo tanto, solo las moléculas originales y no las copias de la primera amplificación se someten a una amplificación adicional. Las moléculas amplificadas adicionales se someten luego a análisis de secuencia. Las secuencias pueden entonces compararse a partir de las dos alícuotas. Como en el esquema de separación discutido anteriormente, las etiquetas de ácido nucleico en los adaptadores no se utilizan para distinguir entre ADN metilado y no metilado sino para distinguir moléculas de ácido nucleico dentro de la misma partición.

d. PCR sensible a la metilación

[0132] En algunas formas de realización, la amplificación sensible a la metilación se utiliza para evaluar la metilación en regiones objetivo de variable de hipermetilación y/o variable de hipometilación. Diversos pasos pueden volverse sensibles a la metilación adaptando enfoques conocidos a los métodos descritos en este documento.

[0133] Por ejemplo, una muestra puede dividirse en alícuotas, por ejemplo, antes o después de un paso de captura como se describe en este documento, y una alícuota puede digerirse con una enzima de restricción sensible a la metilación, por ejemplo, como se describe en Moore et al., *Methods Mol Biol.* 325:239-49 (2006). Las secuencias no metiladas se digieren en esta alícuota. Las alícuotas digeridas y no digeridas pueden luego llevarse a cabo a través de pasos apropiados como se describe en este documento (amplificación, opcionalmente etiquetado, secuenciación y similares) y las secuencias analizadas para determinar el grado de digestión en la muestra tratada, que refleja la presencia de citosinas no metiladas. Como alternativa, se puede evitar la división en alícuotas amplificando una muestra, separando el material amplificado de las plantillas originales y luego digiriendo el material original con una enzima de restricción sensible a la metilación antes de realizar una amplificación adicional, por ejemplo, como se discutió anteriormente con respecto a la secuenciación con bisulfito.

[0134] En otro ejemplo, una muestra puede dividirse en alícuotas y una alícuota puede tratarse para convertir citosinas no metiladas en uracilo, por ejemplo, como se describe en US 2003/0082600 antes de la captura. La conversión de citosinas no metiladas en uracilo reducirá la eficiencia de captura de regiones objetivo con baja metilación al alterar la secuencia de las regiones. Las alícuotas tratadas y no tratadas pueden luego llevarse adelante a través de pasos apropiados como se describe en este documento (captura, amplificación, opcionalmente etiquetado, secuenciación y similares) y las secuencias pueden analizarse para determinar el grado de agotamiento de regiones objetivo en la muestra tratada, que refleja la presencia de citosinas no metiladas.

4. Sujetos

[0135] En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto que tiene un cáncer. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto sospechoso de tener un cáncer. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto que tiene un tumor. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto que se sospecha que tiene un tumor. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto que tiene neoplasia. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto que se sospecha que tiene neoplasia. En algunas formas de realización, el ADN (por ejemplo, ADNlc) se obtiene de un sujeto en remisión de un tumor, cáncer o neoplasia (por ejemplo, después de quimioterapia, resección quirúrgica, radiación o una combinación de las mismas). En cualquiera de las formas de realización anteriores, el cáncer, tumor o neoplasia o cáncer, tumor o neoplasia sospechosos pueden ser de pulmón, colon, recto, riñón, mama, próstata o hígado. En algunas formas de realización, el cáncer, tumor o neoplasia o cáncer, tumor o neoplasia sospechosos son de pulmón. En algunas formas de realización, el cáncer, tumor o neoplasia o el cáncer, tumor o neoplasia sospechosos son del colon o del recto. En algunas formas de realización, el cáncer, tumor o neoplasia o el cáncer, tumor o neoplasia sospechosos son de la mama. En algunas formas de realización, el cáncer, tumor o neoplasia o el cáncer, tumor o neoplasia sospechosos son de la próstata. En cualquiera de las formas de realización anteriores, el sujeto puede ser un sujeto humano.

[0136] En algunas formas de realización, al sujeto se le diagnosticó previamente un cáncer, por ejemplo, cualquiera de los cánceres indicados anteriormente o en otra parte del presente documento. Dicho sujeto puede haber recibido previamente uno o más tratamientos contra el cáncer, por ejemplo, cirugía, quimioterapia, radiación y/o inmunoterapia. En algunas formas de realización, se obtiene una muestra (por ejemplo, ADNlc) de un sujeto diagnosticado y tratado previamente en uno o más puntos temporales preseleccionados después de uno o más tratamientos contra el cáncer previos.

[0137] La muestra (por ejemplo, ADNlc) obtenida del sujeto puede secuenciarse para proporcionar un conjunto de información de secuencia, que puede incluir la secuenciación de moléculas de ADN capturadas del conjunto de regiones objetivo variables de secuencia a una mayor profundidad de secuenciación que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas, como se describe en detalle en otra parte del presente documento.

[0138] En cualquiera de las formas de realización anteriores, el ADN puede obtenerse de una muestra de tejido (por ejemplo, una biopsia, un aspirado con aguja fina o una muestra fijada con formalina e incluida en parafina).

5. Secuenciación

[0139] Los ácidos nucleicos de muestra, después de los pasos de aislamiento o partición y otros pasos como se describe en este documento, con o sin amplificación previa, pueden someterse a secuenciación. Los métodos de secuenciación o formatos disponibles comercialmente que se utilizan opcionalmente incluyen, por ejemplo, secuenciación de Sanger, secuenciación de alto rendimiento, pirosecuenciación, secuenciación por síntesis, secuenciación de una sola molécula, secuenciación basada en nanoporos, secuenciación de semiconductores, secuenciación por ligadura, secuenciación por hibridación, RNA-Seq (Illumina), expresión génica digital (Helicos), secuenciación de próxima generación (NGS), secuenciación de una sola molécula por síntesis (SMSS) (Helicos), secuenciación masiva paralela, matriz de una sola molécula clonal (Solexa), secuenciación shotgun, Ion Torrent, Oxford Nanopore, Roche Genia, secuenciación Maxam-Gilbert, primer walking, secuenciación utilizando plataformas PacBio, SOLiD, Ion Torrent o Nanopore. Las reacciones de secuenciación se pueden realizar en una variedad de unidades de procesamiento de muestras, que pueden incluir múltiples carriles, múltiples canales, múltiples pocillos u otros medios para procesar múltiples conjuntos de muestras de manera prácticamente simultánea. Las unidades de procesamiento de muestras también pueden incluir múltiples cámaras de muestras para permitir el procesamiento de múltiples ejecuciones simultáneamente.

[0140] Las reacciones de secuenciación se pueden realizar en uno o más tipos de fragmentos de ácido nucleico o regiones que contienen marcadores de cáncer o de otras enfermedades. Las reacciones de secuenciación también se pueden realizar en cualquier fragmento de ácido nucleico presente en la muestra. Las reacciones de secuenciación se pueden realizar en al menos aproximadamente el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 % del genoma. En otros casos, las reacciones de secuencia se pueden realizar en menos de aproximadamente el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 %, 99,9 % o 100 % del genoma.

[0141] Se pueden realizar reacciones de secuenciación simultáneas utilizando técnicas de secuenciación multiplex. En algunas formas de realización, los polinucleótidos (por ejemplo, polinucleótidos libres de células) se secuencian con al menos aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 5.0000 o 100.000 reacciones de secuenciación. En otras formas de realización, los polinucleótidos se secuencian con menos de aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 5.0000 o 100.000 reacciones de secuenciación. Las reacciones de secuenciación se realizan normalmente de forma secuencial o simultánea. El análisis de datos posterior se realiza generalmente en todas o en parte de las reacciones de secuenciación. En algunas formas de realización, el análisis de datos se realiza en al menos aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 5.0000 o 100.000 reacciones de secuenciación. En otras formas de realización, el análisis

de datos se puede realizar en menos de aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 5.0000 o 100.000 reacciones de secuenciación. Un ejemplo de una profundidad de lectura es de aproximadamente 1.000 a aproximadamente 5.0000 lecturas por locus (por ejemplo, posición de base).

5 a. Profundidad diferencial de secuenciación

[0142] En algunas formas de realización, los ácidos nucleicos correspondientes al conjunto de regiones objetivo de secuencia variable se secuencian a una mayor profundidad de secuenciación que los ácidos nucleicos correspondientes al conjunto de regiones objetivo epigenéticas. Por ejemplo, la profundidad de secuenciación para ácidos nucleicos correspondientes al conjunto de regiones objetivo de variantes de secuencia puede ser al menos 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 veces mayor, o de 1,25 a 1,5, de 1,5 a 1,75, de 1,75 a 2, de 2 a 2,25, de 2,25 a 2,5, de 2,5 a 2,75, de 2,75 a 3, de 3 a 3,5, de 3,5 a 4, 4 a 4,5, 4,5 a 5, 5 a 5,5, 5,5 a 6, 6 a 7, 7 a 8, 8 a 9, 9 a 10, 10 a 11, 11 a 12, 13 a 14, 14 a 15 veces o 15 a 100 veces mayor que la profundidad de secuenciación de los ácidos nucleicos correspondientes al conjunto de regiones objetivo epigenéticas. En algunas formas de realización, dicha profundidad de secuenciación es al menos 2 veces mayor. En algunas formas de realización, dicha profundidad de secuenciación es al menos 5 veces mayor. En algunas formas de realización, dicha profundidad de secuenciación es al menos 10 veces mayor. En algunas formas de realización, dicha profundidad de secuenciación es de 4 a 10 veces mayor. En algunas formas de realización, dicha profundidad de secuenciación es de 4 a 100 veces mayor. Cada una de estas formas de realización se refiere al grado en el que los ácidos nucleicos correspondientes al conjunto de regiones objetivo de secuencia variable se secuencian a una mayor profundidad de secuenciación que los ácidos nucleicos correspondientes al conjunto de regiones objetivo epigenéticas.

[0143] En algunas formas de realización, las regiones objetivo de secuencia variable se secuencian con una cobertura de al menos 1.000X, tal como una cobertura de 1.000X-2.0000X, o una cobertura de 2.000X-2.0000X. En algunas formas de realización, la cobertura se refiere al número medio o mediano de moléculas únicas en las que se secuencia una posición genómica en una región objetivo de secuencia variable. En algunas formas de realización, las regiones objetivo epigenéticas se secuencian con una cobertura de al menos 1.000X, tal como una cantidad de cobertura en el rango de 1.000X-10.000X. En algunas formas de realización, las regiones objetivo de secuencia variable se secuencian con una mayor cantidad de cobertura que las regiones objetivo epigenéticas. Por ejemplo, la cobertura para los ácidos nucleicos correspondientes al conjunto de regiones objetivo de la variante de secuencia puede ser al menos 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 veces mayor, o de 1,25 a 1,5, de 1,5 a 1,75, de 1,75 a 2, de 2 a 2,25, de 2,25 a 2,5, de 2,5 a 2,75, de 2,75 a 3, de 3 a 3,5, de 3,5 a 4, 4 a 4,5, 4,5 a 5, 5 a 5,5, 5,5 a 6, 6 a 7, 7 a 8, 8 a 9, 9 a 10, 10 a 11, 11 a 12, 13 a 14, 14 a 15 veces o 15 a 100 veces mayor que la cobertura de los ácidos nucleicos correspondientes al conjunto de regiones objetivo epigenéticas.

[0144] En algunas formas de realización, la primera pluralidad de regiones objetivo capturadas se agrupan con la segunda pluralidad de regiones objetivo capturadas antes de la secuenciación. Como tal, la primera y segunda pluralidad de regiones objetivo capturadas pueden secuenciarse simultáneamente, por ejemplo, en la misma célula de secuenciación (tal como la célula de flujo de un Illumina u otro secuenciador) y/o en la misma composición.

6. Análisis

[0145] La secuenciación puede generar una pluralidad de lecturas de secuencia o lecturas. Las lecturas de secuencia o lecturas pueden incluir secuencias de datos de nucleótidos de menos de aproximadamente 150 bases de longitud, o menos de aproximadamente 90 bases de longitud. En algunas formas de realización, las lecturas tienen entre aproximadamente 80 bases y aproximadamente 90 bases, por ejemplo, aproximadamente 85 bases de longitud. En algunas formas de realización, los métodos de la presente divulgación se aplican a lecturas muy cortas, por ejemplo, menos de aproximadamente 50 bases o aproximadamente 30 bases de longitud. Los datos de lectura de secuencia pueden incluir los datos de secuencia, así como metainformación.

[0146] En algunas formas de realización, los recuentos de moléculas se determinan a partir de los datos de secuenciación para una pluralidad de particiones, por ejemplo, particiones hipermetiladas e hipometiladas. Un recuento de moléculas indica el número de moléculas únicas correspondientes a una región objetivo dada que están representadas en los datos de secuencia. Las lecturas se pueden agrupar en familias derivadas de moléculas únicas, por ejemplo, utilizando secuencias de etiqueta, solas o en combinación con otros aspectos de la secuencia, como la longitud o uno o ambos puntos finales de la secuencia de la molécula alineada con el genoma de referencia. En algunas formas de realización, los recuentos de moléculas se pueden utilizar para determinar el grado de metilación (u otras características, dependiendo del tipo de partición que se realizó) en regiones objetivo de interés, como regiones objetivo variables de hipermetilación. Cuando se secuencia todo o sustancialmente todo el ADN en una partición, el recuento de moléculas para una región objetivo dada se puede determinar como el número de moléculas únicas correspondientes a esa región objetivo de la que se obtienen los datos de secuencia. En algunas formas de realización, cuando la segunda pluralidad de regiones objetivo capturadas comprende regiones objetivo capturadas de una parte (menos del 100 %) de la partición hipometilada, el recuento de moléculas se puede estimar mediante una simple multiplicación con un factor de escala que compensa la secuenciación de solo una parte del ADN (por ejemplo, cuando se secuencia un tercio del ADN, el factor de escala sería 3). Por lo tanto, el factor de escala puede ser el recíproco de la fracción de la partición hipometilada que se incluyó en el segundo grupo. Sin embargo, existe la posibilidad de una representación sesgada debido a uno o más de los

procedimientos de enriquecimiento, amplificación (cuando se utilicen) y secuenciación, de modo que pueden ser deseables enfoques alternativos en algunas situaciones.

[0147] En algunas formas de realización, se estima un recuento de moléculas (por ejemplo, para regiones objetivo epigenéticas en la partición hipometilada) utilizando una relación de anclaje. Por ejemplo, se pueden capturar regiones de control, cuya frecuencia relativa en las particiones hipermetiladas e hipometiladas no se espera que varíe sustancialmente de una muestra a otra, y se puede determinar la relación de anclaje a partir de las regiones de control. Como alternativa, se puede determinar la relación de anclaje en función de los niveles de diversidad, por ejemplo, el número de moléculas únicas detectadas en las regiones objetivo epigenéticas en comparación con las regiones objetivo de secuencia variable para las particiones hipermetiladas e hipometiladas.

[0148] Alternativamente, cuando un lote de muestras se encuentra en proceso de procesamiento, se puede calibrar un factor de escala mediante la secuenciación de toda la partición hipometilada de algunas muestras en el lote. El factor de escala se determina, por ejemplo, mediante la determinación de la diferencia de frecuencia media o mediana de las regiones objetivo epigenéticas en los datos de secuencia de la partición hipometilada de muestras en las que se secuenció toda la partición hipometilada frente a muestras en las que solo se secuenció una parte de la partición hipometilada.

[0149] Alternativamente, cuando un lote de muestras se encuentra en proceso de procesamiento, se puede determinar un factor de escala para un lote secuenciando una muestra del lote dos veces o una pluralidad de muestras del lote dos veces, una vez utilizando una fracción de la partición hipometilada y una vez utilizando la partición hipometilada completa. La relación de los recuentos de moléculas entre los dos conjuntos de datos se puede utilizar como el factor de escala. Por lo tanto, se puede determinar un factor de escala a partir de una diferencia de pliegue media o mediana en la frecuencia de regiones objetivo epigenéticas en datos de secuencia de partición hipometilada de una pluralidad de conjuntos de datos de secuencia de una o una pluralidad de muestras, comprendiendo los conjuntos de datos de secuencia datos de secuencia en los que se secuenció una fracción de la partición hipometilada y datos de secuencia en los que se secuenció la partición hipometilada completa.

[0150] Como alternativa, se puede capturar un subconjunto de regiones objetivo epigenéticas junto con las regiones objetivo variables de secuencia, por ejemplo, incluyendo sus sondas junto con las regiones objetivo variables de secuencia, para proporcionar un punto de referencia para determinar un factor de escala.

[0151] Como alternativa, se puede determinar un factor de escala utilizando la relación entre lecturas y moléculas únicas (por ejemplo, número de lecturas por molécula única) para inferir qué recuento de moléculas habría resultado de la captura de regiones objetivo epigenéticas de toda la partición hipometilada. Por ejemplo, el factor de escala se puede inferir mediante el muestreo descendente digital de las lecturas en cada muestra, calculando luego el número de moléculas para cada condición de muestreo descendente. Este cálculo se puede utilizar para determinar la variación en nuevas moléculas únicas como una función de lecturas adicionales. Mientras que, en una biblioteca teórica de complejidad infinita, cada nueva lectura sería una nueva molécula, en la práctica, el factor de escala para una muestra donde, por ejemplo, se incluye el 20 % de una partición hipometilada generalmente será menor de 5 veces porque no cada lectura producirá una molécula única. Ajustar la relación entre lecturas y moléculas a diferentes tasas de muestreo descendente puede permitir una inferencia de cuántas moléculas estarían representadas si hubiera algún número de lecturas adicionales. De este modo, la relación entre las lecturas de secuenciación y las moléculas observadas puede hacerse universal mediante esta manipulación matemática. Una vez establecida esta relación, se puede inferir la cantidad de moléculas únicas a partir de la cantidad de lecturas de secuenciación observadas.

[0152] Los datos de lectura de secuencias se pueden almacenar en cualquier formato de archivo adecuado, incluidos, por ejemplo, archivos VCF, archivos FASTA o archivos FASTQ. FASTA puede hacer referencia a un programa informático para buscar bases de datos de secuencias, y el nombre FASTA también puede hacer referencia a un formato de archivo estándar. FASTA se describe, por ejemplo, en Pearson & Lipman, 1988, Improved tools for biological sequence comparison, PNAS 85:2444-2448. Una secuencia en formato FASTA comienza con una descripción de una sola línea, seguida de líneas de datos de secuencia. La línea de descripción se distingue de los datos de secuencia por un símbolo mayor que (">") en la primera columna. La palabra que sigue al símbolo ">" es el identificador de la secuencia, y el resto de la línea es la descripción (ambos son opcionales). No puede haber espacio entre el ">" y la primera letra del identificador. Se recomienda que todas las líneas de texto sean más cortas que 80 caracteres. La secuencia termina si aparece otra línea que comience con un ">"; Esto indica el inicio de otra secuencia.

[0153] El formato FASTQ es un formato basado en texto para almacenar tanto una secuencia biológica (normalmente una secuencia de nucleótidos) como sus puntuaciones de calidad correspondientes. Es similar al formato FASTA, pero con puntuaciones de calidad a continuación de los datos de la secuencia. Tanto la letra de la secuencia como la puntuación de calidad se codifican con un único carácter ASCII para abreviar. El formato FASTQ es un estándar de facto para almacenar el resultado de instrumentos de secuenciación de alto rendimiento como el analizador de genoma Illumina, como describen, por ejemplo, Cock et al. ("The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants", Nucleic Acids Res 38(6):1767-1771, 2009).

[0154] Para los archivos FASTA y FASTQ, la metainformación incluye la línea de descripción y no las líneas de datos de secuencia. En algunas formas de realización, para los archivos FASTQ, la metainformación incluye las puntuaciones de

calidad. Para los archivos FASTA y FASTQ, los datos de secuencia comienzan después de la línea de descripción y están presentes típicamente utilizando algún subconjunto de códigos de ambigüedad IUPAC opcionalmente con "-". En una forma de realización, los datos de secuencia pueden utilizar los caracteres A, T, C, G y N, incluyendo opcionalmente "-" o U según sea necesario (por ejemplo, para representar espacios o uracilo).

[0155] En algunas formas de realización, el al menos un archivo de lectura de secuencia maestra y el archivo de salida se almacenan como archivos de texto simple (por ejemplo, utilizando una codificación como ASCII; ISO/IEC 646; EBCDIC; UTF-8; o UTF-16). Un sistema informático proporcionado por la presente divulgación puede incluir un programa de edición de texto capaz de abrir los archivos de texto simple. Un programa de edición de texto puede referirse a un programa informático capaz de presentar el contenido de un archivo de texto (como un archivo de texto simple) en una pantalla de ordenador, lo que permite que un ser humano edite el texto (por ejemplo, utilizando un monitor, un teclado y un ratón). Los ejemplos de editores de texto incluyen, sin limitación, Microsoft Word, emacs, pico, vi, BBEdit y TextWrangler. El programa de edición de texto puede ser capaz de mostrar los archivos de texto simple en una pantalla de ordenador, mostrando la metainformación y las lecturas de secuencia en un formato legible para humanos (por ejemplo, no codificado en binario sino utilizando caracteres alfanuméricos como pueden utilizarse en la impresión o en la escritura humana).

[0156] Aunque se han analizado métodos con referencia a archivos FASTA o FASTQ, los métodos y sistemas de la presente divulgación se pueden utilizar para comprimir cualquier formato de archivo de secuencia adecuado, incluidos, por ejemplo, archivos en el formato de formato de llamada de variante (VCF). Un archivo VCF típico puede incluir una sección de encabezado y una sección de datos. El encabezado contiene un número arbitrario de líneas de metainformación, cada una comenzando con caracteres '##', y una línea de definición de campo delimitada por tabuladores que comienza con un solo carácter '#'. La línea de definición de campo nombra ocho columnas obligatorias y la sección de cuerpo contiene líneas de datos que llenan las columnas definidas por la línea de definición de campo. El formato VCF se describe, por ejemplo, por Danecek et al. ("The variant call format and VCF tools", Bioinformatics 27(15):2156-2158, 2011). La sección de encabezado se puede tratar como la metainformación para escribir en los archivos comprimidos y la sección de datos se puede tratar como las líneas, cada una de las cuales se puede almacenar en un archivo maestro solo si es única.

[0157] Algunas formas de realización prevén el ensamblaje de lecturas de secuencias. En el ensamblaje por alineamiento, por ejemplo, las lecturas de secuencias se alinean entre sí o se alinean con una secuencia de referencia. Al alinear cada lectura, a su vez con un genoma de referencia, todas las lecturas se posicionan en relación entre sí para crear el ensamblaje. Además, la alineación o mapeo de la lectura de secuencia con una secuencia de referencia también se puede utilizar para identificar secuencias variantes dentro de la lectura de secuencia. La identificación de secuencias variantes se puede utilizar en combinación con los métodos y sistemas descritos en el presente documento para ayudar aún más en el diagnóstico o pronóstico de una enfermedad o afección, o para guiar las decisiones de tratamiento.

[0158] En algunas formas de realización, cualquiera o todos los pasos están automatizados. Alternativamente, los métodos de la presente divulgación pueden incorporarse total o parcialmente en uno o más programas dedicados, por ejemplo, cada uno opcionalmente escrito en un lenguaje compilado tal como C++, luego compilado y distribuido como un binario. Los métodos de la presente divulgación pueden implementarse total o parcialmente como módulos dentro de, o invocando funcionalidad dentro de, plataformas de análisis de secuencias existentes. En algunas formas de realización, los métodos de la presente divulgación incluyen una serie de pasos que se invocan todos automáticamente en respuesta a una única cola de inicio (por ejemplo, uno o una combinación de eventos desencadenantes provenientes de la actividad humana, otro programa informático o una máquina). Por lo tanto, la presente divulgación proporciona métodos en los que cualquiera o los pasos o cualquier combinación de los pasos pueden ocurrir automáticamente en respuesta a una cola. "Automáticamente" generalmente significa sin intervención de entrada, influencia o interacción humana (por ejemplo, en respuesta solo a la actividad humana original o previa a la cola).

[0159] Los métodos de la presente divulgación también pueden abarcar varias formas de salida, que incluyen una interpretación precisa y sensible de la muestra de ácido nucleico de un sujeto. La salida de la recuperación se puede proporcionar en el formato de un archivo de computadora. En algunas formas de realización, la salida es un archivo FASTA, un archivo FASTQ o un archivo VCF. La salida se puede procesar para producir un archivo de texto o un archivo XML que contiene datos de secuencia, como una secuencia del ácido nucleico alineada con una secuencia del genoma de referencia. En otras formas de realización, el procesamiento produce una salida que contiene coordenadas o una cadena que describe una o más mutaciones en el ácido nucleico del sujeto en relación con el genoma de referencia. Las cadenas de alineación pueden incluir el Informe de Alineación Simple sin Espacios (SUGAR, por sus siglas en inglés), el Informe de Alineación con Espacios Etiquetados, Verbosos y Útiles (VULGAR, por sus siglas en inglés) y el Informe de Alineación con Espacios Idiosincrásicos Compactos (CIGAR, por sus siglas en inglés) (como se describe, por ejemplo, en Ning et al., Genome Research 11(10):1725-9, 2001). Estas cadenas pueden implementarse, por ejemplo, en el software de alineamiento de secuencias Exonerate del Instituto Europeo de Bioinformática (Hinxton, Reino Unido).

[0160] En algunas formas de realización, se produce una alineación de secuencias, como, por ejemplo, un archivo de mapa de alineación de secuencias (SAM) o mapa de alineación binaria (BAM), que comprende una cadena CIGAR (el formato SAM se describe, por ejemplo, por Li et al., "The Sequence Alignment/Map format and SAMtools", Bioinformatics, 25(16):2078-9, 2009). En algunas formas de realización, CIGAR muestra o incluye alineaciones con espacios en blanco, una por línea. CIGAR es un formato de alineación por pares comprimido que se informa como una cadena CIGAR. Una

cadena CIGAR puede ser útil para representar alineaciones por pares largas (por ejemplo, genómicas). Una cadena CIGAR se puede utilizar en formato SAM para representar alineaciones de lecturas con una secuencia de genoma de referencia.

[0161] Una cadena CIGAR puede seguir un motivo establecido. Cada carácter está precedido por un número, que proporciona los recuentos de base del evento. Los caracteres utilizados pueden incluir M, I, D, N y S (M = coincidencia; I = inserción; D = eliminación; N = espacio; S = sustitución). La cadena CIGAR define la secuencia de coincidencias y/o desajustes y eliminaciones (o espacios). Por ejemplo, la cadena CIGAR 2MD3M2D2M puede indicar que la alineación contiene 2 coincidencias, 1 eliminación (se omite el número 1 para ahorrar algo de espacio), 3 coincidencias, 2 eliminaciones y 2 coincidencias.

[0162] En algunas formas de realización, se prepara una población de ácidos nucleicos para secuenciación mediante la formación enzimática de extremos romos en ácidos nucleicos bicatenarios con salientes monocatenarios en uno o ambos extremos. En estas formas de realización, la población se trata típicamente con una enzima que tiene una actividad de ADN polimerasa 5'-3' y una actividad de exonucleasa 3'-5' en presencia de los nucleótidos (por ejemplo, A, C, G y T o U). Los ejemplos de enzimas o fragmentos catalíticos de las mismas que se pueden utilizar opcionalmente incluyen el fragmento grande de Klenow y la polimerasa T4. En los salientes 5', la enzima típicamente extiende el extremo 3' rebajado en la hebra opuesta hasta que está al ras del extremo 5' para producir un extremo romo. En los salientes 3', la enzima generalmente digiere desde el extremo 3' hasta y, a veces, más allá del extremo 5' de la hebra opuesta. Si esta digestión se produce más allá del extremo 5' de la cadena opuesta, el hueco puede ser rellenado por una enzima que tenga la misma actividad polimerasa que se utiliza para los salientes 5'. La formación de extremos romos en los ácidos nucleicos de doble cadena facilita, por ejemplo, la unión de adaptadores y la posterior amplificación.

[0163] En algunas formas de realización, las poblaciones de ácidos nucleicos se someten a un procesamiento adicional, como la conversión de ácidos nucleicos monocatenarios en ácidos nucleicos bicatenarios y/o la conversión de ARN en ADN (por ejemplo, ADN complementario o ADNc). Estas formas de ácido nucleico también se unen opcionalmente a adaptadores y se amplifican.

[0164] Con o sin amplificación previa, los ácidos nucleicos sujetos al proceso de formación de extremos romos descrito anteriormente, y opcionalmente otros ácidos nucleicos en una muestra, pueden secuenciarse para producir ácidos nucleicos secuenciados. Un ácido nucleico secuenciado puede referirse a la secuencia de un ácido nucleico (por ejemplo, información de secuencia) o a un ácido nucleico cuya secuencia ha sido determinada. La secuenciación puede realizarse de manera que proporcione datos de secuencia de moléculas de ácido nucleico individuales en una muestra, ya sea directa o indirectamente a partir de una secuencia de consenso de productos de amplificación de una molécula de ácido nucleico individual en la muestra.

[0165] En algunas formas de realización, los ácidos nucleicos bicatenarios con salientes monocatenarios en una muestra después de la formación de extremos romos se unen en ambos extremos a adaptadores que incluyen códigos de barras, y la secuenciación determina las secuencias de ácidos nucleicos, así como los códigos de barras en línea introducidos por los adaptadores. Las moléculas de ADN de extremos romos se ligan opcionalmente a un extremo romo de un adaptador al menos parcialmente bicatenario (por ejemplo, un adaptador en forma de Y o en forma de campana). Alternativamente, los extremos romos de los ácidos nucleicos de muestra y los adaptadores se pueden unir con nucleótidos complementarios para facilitar la ligadura (por ejemplo, ligadura de extremos pegajosos).

[0166] La muestra de ácido nucleico se pone en contacto típicamente con un número suficiente de adaptadores de modo que exista una baja probabilidad (por ejemplo, menos de aproximadamente 1 o 0,1 %) de que dos copias cualesquiera del mismo ácido nucleico reciban la misma combinación de códigos de barras de adaptadores de los adaptadores unidos en ambos extremos. El uso de adaptadores de esta manera puede permitir la identificación de familias de secuencias de ácidos nucleicos con los mismos puntos de inicio y detención en un ácido nucleico de referencia y unidos a la misma combinación de códigos de barras. Dicha familia puede representar secuencias de productos de amplificación de un ácido nucleico en la muestra antes de la amplificación. Las secuencias de los miembros de la familia se pueden compilar para derivar uno o más nucleótidos de consenso o una secuencia de consenso completa para una molécula de ácido nucleico en la muestra original, modificada por la formación de extremos romos y la unión de adaptadores. En otras palabras, el nucleótido que ocupa una posición específica de un ácido nucleico en la muestra se puede determinar como el consenso de nucleótidos que ocupan esa posición correspondiente en las secuencias de miembros de la familia. Las familias pueden incluir secuencias de una o ambas cadenas de un ácido nucleico bicatenario. Si los miembros de una familia incluyen secuencias de ambas cadenas de un ácido nucleico bicatenario, las secuencias de una cadena pueden convertirse en sus complementos con el fin de compilar secuencias para derivar nucleótidos o secuencias de consenso. Algunas familias incluyen solo una secuencia de un solo miembro. En este caso, esta secuencia puede tomarse como la secuencia de un ácido nucleico en la muestra antes de la amplificación. Alternativamente, las familias con solo una secuencia de un solo miembro pueden eliminarse del análisis posterior.

[0167] Las variaciones de nucleótidos (por ejemplo, SNV o indeles) en ácidos nucleicos secuenciados se pueden determinar comparando ácidos nucleicos secuenciados con una secuencia de referencia. La secuencia de referencia es a menudo una secuencia conocida, por ejemplo, una secuencia de genoma completo o parcial conocida de un sujeto (por ejemplo, una secuencia de genoma completo de un sujeto humano). La secuencia de referencia puede ser, por ejemplo,

hG19 o hG38. Los ácidos nucleicos secuenciados pueden representar secuencias determinadas directamente para un ácido nucleico en una muestra, o un consenso de secuencias de productos de amplificación de dicho ácido nucleico, como se describió anteriormente. Se puede realizar una comparación en una o más posiciones designadas en una secuencia de referencia. Se puede identificar un subconjunto de ácidos nucleicos secuenciados que incluya una posición que corresponda con una posición designada de la secuencia de referencia cuando las secuencias respectivas estén alineadas al máximo. Dentro de dicho subconjunto se puede determinar cuáles, si los hay, ácidos nucleicos secuenciados incluyen una variación de nucleótido en la posición designada, y opcionalmente cuáles, si los hay, incluyen un nucleótido de referencia (por ejemplo, el mismo que en la secuencia de referencia). Si el número de ácidos nucleicos secuenciados en el subconjunto que incluye una variante de nucleótido excede un umbral seleccionado, entonces se puede llamar a un nucleótido variante en la posición designada. El umbral puede ser un número simple, tal como al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 ácidos nucleicos secuenciados dentro del subconjunto que incluye la variante de nucleótido o puede ser una proporción, tal como al menos 0,5, 1, 2, 3, 4, 5, 10, 15 o 20, de ácidos nucleicos secuenciados dentro del subconjunto que incluye la variante de nucleótido, entre otras posibilidades. La comparación se puede repetir para cualquier posición designada de interés en la secuencia de referencia. A veces se puede realizar una comparación para posiciones designadas que ocupan al menos aproximadamente 20, 100, 200 o 300 posiciones contiguas en una secuencia de referencia, por ejemplo, aproximadamente 20-500, o aproximadamente 50-300 posiciones contiguas.

[0168] Se pueden encontrar métodos y sistemas ejemplares útiles para analizar las variaciones epigenéticas tales como el patrón de metilación y fragmentación en WO2021/202752 y WO2016/015058.

[0169] También se proporcionan detalles adicionales sobre la secuenciación de ácidos nucleicos, incluidos los formatos y aplicaciones descritos en este documento, en, por ejemplo, Levy et al., Annual Review of Genomics and Human Genetics, 17: 95-115 (2016), Liu et al., J. of Biomedicine and Biotechnology, Volumen 2012, ID de artículo 251364:1-11 (2012), Voelkerding et al., Clinical Chem., 55: 641-658 (2009), MacLean et al., Nature Rev. Microbiol., 7: 287-296 (2009), Astier et al., J Am Chem Soc., 128(5):1705-10 (2006), Patente de EE. UU. N.º 6.210.891, Patente de EE. UU. N.º 6.258.568, Patente de EE. UU. N.º 6.258.568 No. 6.833,246, Patente de EE. UU. N.º 7,115,400, Patente de EE. UU. N.º 6,969,488, Patente de EE. UU. N.º 5,912,148, Patente de EE. UU. N.º 6,130,073, Patente de EE. UU. N.º 7,169,560, Patente de EE. UU. N.º 7,282,337, Patente de EE. UU. N.º 7,482,120, Patente de EE. UU. N.º 7,501,245, Patente de EE. UU. N.º 6,818,395, Patente de EE. UU. N.º 6,911,345, Patente de EE. UU. N.º 7,501,245, Patente de EE. UU. N.º 7,329,492, Patente de EE. UU. N.º 7,170,050, Patente de EE. UU. N.º 7,302,146, Patente de EE. UU. N.º 7,313,308, y Patente de EE. UU. N.º 7,476,503.

7. Método ejemplar para la identificación de etiquetas moleculares de bibliotecas particionadas con perlas de MBD

[0170] Un método ejemplar para la identificación de etiquetas moleculares de bibliotecas particionadas con perlas de MBD a través de NGS es el siguiente:

i) Partición física de una muestra de ADN extraída (por ejemplo, ADN de plasma sanguíneo extraído de una muestra humana, que opcionalmente se ha sometido a captura de objetivo como se describe en este documento) utilizando un kit de purificación de perlas de proteína de dominio de unión a metilo, guardando todas las eluciones del proceso para el procesamiento posterior.

ii) Aplicación paralela de códigos de barras moleculares diferenciales y secuencias adaptadoras que permiten NGS a cada partición. Por ejemplo, la metilación residual hipermetilada (por ejemplo, recogida de un paso de "lavado" entre la recogida del flujo continuo y la elución durante la partición utilizando una columna) y las particiones hipometiladas se ligan con adaptadores NGS con códigos de barras moleculares (por ejemplo, códigos de barras moleculares específicos de la partición).

iii) Preparar un primer pool que comprende una porción (por ejemplo, 70-90 %, tal como aproximadamente 80 %) de la partición hipometilada, y un segundo pool que comprende la partición hipermetilada y una porción (por ejemplo, 10-30 %, tal como aproximadamente 20 %) de la partición hipometilada y opcionalmente cualquier partición intermedia, y amplificación posterior del primer y segundo pool utilizando secuencias de cebadores de ADN específicos del adaptador.

iv) Captura/hibridación de los pools amplificados, dirigiéndose a regiones genómicas de interés (por ejemplo, variantes genéticas específicas del cáncer para el primer pool, y variantes genéticas específicas del cáncer y regiones metiladas diferencialmente para el segundo pool).

v) Re-amplificación del ADN capturado del primer y segundo pool (que se puede combinar para dar una biblioteca antes de la reamplificación), añadiendo una etiqueta de muestra. Se agrupan diferentes muestras y se analizan en multiplex en un instrumento NGS.

vi) Análisis bioinformático de datos de NGS, en el que se utilizan etiquetas moleculares para identificar moléculas únicas, así como la deconvolución de la muestra en moléculas que se dividieron en MED de forma diferencial. Este

análisis puede proporcionar información sobre la 5-metilcitosina relativa para las regiones genómicas, de manera simultánea con la secuenciación genética estándar y la detección de variantes.

[0171] El método ejemplar expuesto anteriormente puede comprender además cualquier característica compatible de los métodos de acuerdo con esta divulgación expuestos en otra parte del presente documento.

8. Flujos de trabajo ejemplares

[0172] En el presente documento se proporcionan flujos de trabajo ejemplares para la partición y preparación de bibliotecas. En algunas formas de realización, algunas o todas las características de los flujos de trabajo de partición y preparación de bibliotecas se pueden utilizar en combinación. Los flujos de trabajo ejemplares expuestos anteriormente pueden comprender además cualquier característica compatible de los métodos de acuerdo con esta divulgación expuestos en otra parte del presente documento.

15 a. Partición

[0173] En algunas formas de realización, el ADN de muestra (por ejemplo, entre 1 y 300 ng) se mezcla con una cantidad apropiada de tampón de dominio de unión a metilo (MBD) (la cantidad de tampón MBD depende de la cantidad de ADN utilizado) y perlas magnéticas conjugadas con proteínas MBD e incubadas durante la noche. El ADN metilado (ADN hipermetilado) se une a la proteína MBD en las perlas magnéticas durante esta incubación. El ADN no metilado (ADN hipometilado) o menos metilado (ADN intermedicamente metilado) se elimina de las perlas con tampones que contienen concentraciones crecientes de sal. Por ejemplo, se pueden obtener una, dos o más fracciones que contienen ADN no metilado, hipometilado y/o intermedicamente metilado a partir de dichos lavados. Finalmente, se utiliza un tampón con alto contenido de sal para eluir el ADN altamente metilado (ADN hipermetilado) de la proteína MBD. En algunas formas de realización, estos lavados dan como resultado tres particiones (partición hipometilada, fracción/partición intermedicamente metilada y partición hipermetilada) de ADN que tienen niveles crecientes de metilación.

[0174] En algunas formas de realización, las tres particiones de ADN se desalinizan y se concentran en preparación para los pasos enzimáticos de preparación de la biblioteca.

30 b. Preparación de la biblioteca

[0175] En algunas formas de realización (por ejemplo, después de concentrar el ADN en las particiones), el ADN particionado se hace ligable, por ejemplo, extendiendo los salientes de los extremos de las moléculas de ADN y añadiendo residuos de adenosina a los extremos 3' de los fragmentos y fosforilando el extremo 5' de cada fragmento de ADN. Se añaden ligasa de ADN y adaptadores para ligar cada molécula de ADN particionada con un adaptador en cada extremo. Estos adaptadores contienen etiquetas de partición (por ejemplo, códigos de barras no aleatorios, no únicos) que se pueden distinguir de las etiquetas de partición en los adaptadores utilizados en las otras particiones. Después de la ligación, las tres particiones se agrupan y se amplifican (por ejemplo, mediante PCR, como con cebadores específicos para los adaptadores).

[0176] Después de la PCR, el ADN amplificado puede limpiarse y concentrarse antes de la captura. El ADN amplificado se pone en contacto con una colección de sondas descritas en este documento (que pueden ser, por ejemplo, sondas de ARN biotiniladas) que se dirigen a regiones específicas de interés. La mezcla se incuba, por ejemplo, durante la noche, por ejemplo, en un tampón de sal. Las sondas se capturan (por ejemplo, utilizando perlas magnéticas de estreptavidina) y se separan del ADN amplificado que no se capturó, como por una serie de lavados con sal, proporcionando así un conjunto capturado de ADN. Después de la captura, el ADN del conjunto capturado se amplifica por PCR. En algunas formas de realización, los cebadores de PCR contienen una etiqueta de muestra, incorporando así la etiqueta de muestra en las moléculas de ADN. En algunas formas de realización, el ADN de diferentes muestras se agrupa y luego se secuencia de forma multiplexada, por ejemplo, utilizando un secuenciador Illumina NovaSeq.

III. CARACTERÍSTICAS GENERALES DE LOS MÉTODOS

1. Muestras

[0177] Una muestra puede ser cualquier muestra biológica aislada de un sujeto. Una muestra puede ser una muestra corporal. Las muestras pueden incluir tejidos corporales, como tumores sólidos conocidos o sospechosos, sangre completa, plaquetas, suero, plasma, heces, glóbulos rojos, glóbulos blancos o leucocitos, células endoteliales, biopsias de tejido, líquido cefalorraquídeo, líquido sinovial, líquido linfático, líquido ascítico, líquido intersticial o extracelular, el líquido en espacios entre células, incluido el líquido crevicular gingival, médula ósea, derrames pleurales, líquido cefalorraquídeo, saliva, mucosidad, esputo, semen, sudor, uretra, etc. Las muestras son preferiblemente fluidos corporales, particularmente sangre y fracciones de la misma, y orina. Una muestra puede estar en la forma originalmente aislada de un sujeto o puede haber sido sometida a un procesamiento adicional para eliminar o agregar componentes, como células, o enriquecer un componente en relación con otro. Por lo tanto, un fluido corporal preferido para el análisis es plasma o suero que contiene ácidos nucleicos libres de células. Una muestra puede aislarse u obtenerse de un sujeto y transportarse a un sitio de análisis de muestra. La muestra puede conservarse y enviarse a una temperatura deseada,

por ejemplo, temperatura ambiente, 4 °C, -20 °C y/o -80 °C. Una muestra puede aislarse u obtenerse de un sujeto en el sitio del análisis de la muestra. El sujeto puede ser un ser humano, un mamífero, un animal, un animal de compañía, un animal de servicio o una mascota. El sujeto puede tener un cáncer. El sujeto puede no tener cáncer o un síntoma de cáncer detectable. El sujeto puede haber sido tratado con una o más terapias contra el cáncer, por ejemplo, una o más de las siguientes quimioterapias, anticuerpos, vacunas o productos biológicos. El sujeto puede estar en remisión. El sujeto puede o no ser diagnosticado de ser susceptible al cáncer o cualquier mutación/trastorno genético asociado al cáncer.

[0178] El volumen de plasma puede depender de la profundidad de lectura deseada para las regiones secuenciadas. Los volúmenes ejemplares son 0,4-40 ml, 5-20 ml, 10-20 ml. Por ejemplo, el volumen puede ser 0,5 ml, 1 ml, 5 ml, 10 ml, 20 ml, 30 mL o 40 ml. Un volumen de plasma muestreado puede ser de 5 a 20 ml.

[0179] Una muestra puede comprender diversas cantidades de ácido nucleico que contienen equivalentes de genoma. Por ejemplo, una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10.000 (10⁴) equivalentes de genoma humano haploide y, en el caso de ADNlc, aproximadamente 200 mil millones (2x10¹¹) de moléculas de polinucleótidos individuales. De manera similar, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30.000 equivalentes de genoma humano haploide y, en el caso de ADNlc, aproximadamente 600 mil millones de moléculas individuales.

[0180] Una muestra puede comprender ácidos nucleicos de diferentes fuentes, por ejemplo, de células y células libres del mismo sujeto, de células y células libres de diferentes sujetos. Una muestra puede comprender ácidos nucleicos que portan mutaciones. Por ejemplo, una muestra puede comprender ADN que porta mutaciones de la línea germinal y/o mutaciones somáticas. Las mutaciones de la línea germinal se refieren a mutaciones existentes en el ADN de la línea germinal de un sujeto. Las mutaciones somáticas se refieren a mutaciones que se originan en células somáticas de un sujeto, por ejemplo, células cancerosas. Una muestra puede comprender ADN que porta mutaciones asociadas al cáncer (por ejemplo, mutaciones somáticas asociadas al cáncer). Una muestra puede comprender una variante epigenética (es decir, una modificación química o proteica), en donde la variante epigenética está asociada con la presencia de una variante genética tal como una mutación asociada al cáncer. En algunas formas de realización, la muestra comprende una variante epigenética asociada con la presencia de una variante genética, en donde la muestra no comprende la variante genética.

[0181] Las cantidades ejemplares de ácidos nucleicos libres de células en una muestra antes de la amplificación varían de aproximadamente 1 fg a aproximadamente 1 µg, por ejemplo, de 1 pg a 200 ng, de 1 ng a 100 ng, de 10 ng a 1.000 ng. Por ejemplo, la cantidad puede ser de hasta aproximadamente 600 ng, hasta aproximadamente 500 ng, hasta aproximadamente 400 ng, hasta aproximadamente 300 ng, hasta aproximadamente 200 ng, hasta aproximadamente 100 ng, hasta aproximadamente 50 ng o hasta aproximadamente 20 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de al menos 1 fg, al menos 10 fg, al menos 100 fg, al menos 1 pg, al menos 10 pg, al menos 100 pg, al menos 1 ng, al menos 10 ng, al menos 100 ng, al menos 150 ng o al menos 200 ng de moléculas de ácido nucleico libres de células. La cantidad puede ser de hasta 1 femtogramo (fg), 10 fg, 100 fg, 1 picogramo (pg), 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 150 ng, 200 ng, 250 ng o 300 ng de moléculas de ácido nucleico libres de células. El método puede comprender la obtención de 1 femtogramo (fg) a 200 ng.

[0182] Los ácidos nucleicos libres de células son ácidos nucleicos que no están contenidos dentro de una célula o unidos de otro modo a ella o, en otras palabras, ácidos nucleicos que permanecen en una muestra después de retirar células intactas. Los ácidos nucleicos libres de células incluyen ADN, ARN e híbridos de los mismos, incluyendo ADN genómico, ADN mitocondrial, ARNi, miARN, ARN circulante (ARNc), ARNt, ARNr, ARN nucleolar pequeño (ARNsno), ARN que interactúa con Piwi (ARNpi), ARN largo no codificante (ARNnc largo) o fragmentos de cualquiera de estos. Los ácidos nucleicos libres de células pueden ser bicatenarios, monocatenarios o un híbrido de los mismos. Un ácido nucleico libre de células puede liberarse en un fluido corporal a través de procesos de secreción o muerte celular, por ejemplo, necrosis celular y apoptosis. Algunos ácidos nucleicos libres de células se liberan en un fluido corporal a partir de células cancerosas, por ejemplo, ADN tumoral circulante (ADNct). Otros se liberan a partir de células sanas. En algunas formas de realización, el ADNlc es ADN fetal libre de células (ADNflic). En algunas formas de realización, los ácidos nucleicos libres de células son producidos por células tumorales. En algunas formas de realización, los ácidos nucleicos libres de células son producidos por una mezcla de células tumorales y células no tumorales.

[0183] Ácidos nucleicos libres de células tienen una distribución de tamaño ejemplar de aproximadamente 100-500 nucleótidos, con moléculas de 110 a aproximadamente 230 nucleótidos que representan aproximadamente el 90 % de las moléculas, con una moda de aproximadamente 168 nucleótidos y un segundo pico menor en un rango entre 240 a 440 nucleótidos.

[0184] Los ácidos nucleicos libres de células se pueden aislar de los fluidos corporales mediante un paso de fraccionamiento o partición en el que los ácidos nucleicos libres de células, tal como se encuentran en solución, se separan de las células intactas y otros componentes no solubles del fluido corporal. La partición puede incluir técnicas tales como centrifugación o filtración. Alternativamente, las células en fluidos corporales se pueden lisar y los ácidos nucleicos libres de células y celulares se pueden procesar juntos. Generalmente, después de la adición de tampones y pasos de lavado, los ácidos nucleicos se pueden precipitar con un alcohol. Se pueden utilizar pasos de limpieza adicionales tales como columnas a base de sílice para eliminar contaminantes o sales. Se pueden agregar ácidos nucleicos portadores a granel

no específicos, tales como ADN C1, ADN o proteína para secuenciación con bisulfito, hibridación y/o ligación, a lo largo de la reacción para optimizar ciertos aspectos del procedimiento tales como el rendimiento.

[0185] Después de dicho procesamiento, las muestras pueden incluir varias formas de ácido nucleico, incluyendo ADN bicatenario, ADN monocatenario y ARN monocatenario. En algunas formas de realización, el ADN y el ARN monocatenarios se pueden convertir en formas bicatenarias para que se incluyan en los pasos de procesamiento y análisis posteriores.

[0186] Las moléculas de ADN de doble cadena en una muestra y las moléculas de ácido nucleico de cadena sencilla convertidas en moléculas de ADN de doble cadena se pueden unir a adaptadores en un extremo o en ambos extremos. Típicamente, las moléculas de doble cadena se vuelven romas en sus extremos mediante el tratamiento con una polimerasa con una polimerasa 5'-3' y una exonucleasa 3'-5' (o función de corrección de pruebas), en presencia de los cuatro nucleótidos estándar. El fragmento grande de Klenow y la polimerasa T4 son ejemplos de polimerasas adecuadas. Las moléculas de ADN de extremos romos se pueden ligar con un adaptador al menos parcialmente bicatenario (por ejemplo, un adaptador en forma de Y o de campana). Alternativamente, se pueden agregar nucleótidos complementarios a los extremos romos de los ácidos nucleicos de muestra y adaptadores para facilitar la ligadura. En el presente documento se contemplan tanto la ligadura de extremos romos como la ligadura de extremos pegajosos. En la ligadura de extremos romos, tanto las moléculas de ácido nucleico como las etiquetas adaptadoras tienen extremos romos. En la ligadura de extremos pegajosos, típicamente, las moléculas de ácido nucleico tienen un saliente "A" y los adaptadores tienen un saliente "T".

2. Etiquetas

[0187] Las etiquetas que comprenden códigos de barras se pueden incorporar o unir de otra manera a los adaptadores. Las etiquetas se pueden incorporar mediante ligadura, PCR de extensión por superposición, entre otros métodos.

a. Estrategias de etiquetado molecular

[0188] El etiquetado molecular se refiere a una práctica de etiquetado que permite diferenciar moléculas de las que se originaron lecturas de secuencia. Las estrategias de etiquetado se pueden dividir en estrategias de etiquetado único y estrategias de etiquetado no único. En el etiquetado único, todas o sustancialmente todas las moléculas en una muestra tienen una etiqueta diferente, de modo que las lecturas se pueden asignar a moléculas originales basándose únicamente en la información de la etiqueta. Las etiquetas utilizadas en dichos métodos a veces se denominan "etiquetas únicas". En el etiquetado no único, diferentes moléculas en la misma muestra pueden tener la misma etiqueta, de modo que se utiliza otra información además de la información de la etiqueta para asignar una lectura de secuencia a una molécula original. Dicha información puede incluir las coordenadas de inicio y fin, las coordenadas a las que se asigna la molécula, las coordenadas de inicio o fin únicamente, etc. Las etiquetas utilizadas en dichos métodos a veces se denominan "etiquetas no únicas". Por consiguiente, no es necesario etiquetar de forma única cada molécula de una muestra. Es suficiente etiquetar de forma única las moléculas que pertenecen a una clase identificable dentro de una muestra. Por lo tanto, las moléculas de diferentes familias identificables pueden llevar la misma etiqueta sin pérdida de información sobre la identidad de la molécula etiquetada.

[0189] En ciertas formas de realización de etiquetado no único, el número de etiquetas diferentes utilizadas puede ser suficiente para que exista una probabilidad muy alta (por ejemplo, al menos 99 %, al menos 99,9 %, al menos 99,99 % o al menos 99,999 %) de que todas las moléculas de un grupo particular lleven una etiqueta diferente. Cabe señalar que cuando se utilizan códigos de barras como etiquetas, y cuando los códigos de barras se unen, por ejemplo, aleatoriamente, a ambos extremos de una molécula, la combinación de códigos de barras, en conjunto, puede constituir una etiqueta. Este número, en términos, es una función del número de moléculas que caen en las llamadas. Por ejemplo, la clase puede ser todas las moléculas que se asignan a la misma posición de inicio-parada en un genoma de referencia. La clase puede ser todas las moléculas que se asignan a través de un locus genético particular, por ejemplo, una base particular o una región particular (por ejemplo, hasta 100 bases o un gen o un exón de un gen). En ciertas formas de realización, el número de etiquetas diferentes utilizadas para identificar de forma única un número de moléculas, z , en una clase puede estar entre cualquiera de 2^z , 3^z , 4^z , 5^z , 6^z , 7^z , 8^z , 9^z , 10^z , 11^z , 12^z , 13^z , 14^z , 15^z , 16^z , 17^z , 18^z , 19^z , 20^z o 100^z (por ejemplo, límite inferior) y cualquiera de 100.000^z , 10.000^z , 1.000^z o 100^z (por ejemplo, límite superior).

[0190] Por ejemplo, en una muestra de aproximadamente 3 ng a 30 ng de ADN libre de células humanas, se espera que alrededor de 10^3 - 10^4 moléculas se asignen a una coordenada de nucleótido particular, y entre aproximadamente 3 y 10 moléculas que tengan cualquier coordenada de inicio compartan la misma coordenada de parada. En consecuencia, pueden ser suficientes aproximadamente de 50 a aproximadamente 50.000 etiquetas diferentes (por ejemplo, entre aproximadamente 6 y 220 combinaciones de códigos de barras) para etiquetar de forma única todas esas moléculas. Para etiquetar de forma única todas las 10^3 - 10^4 moléculas que se asignan a través de una coordenada de nucleótido, se requerirían aproximadamente de 1 millón a aproximadamente 20 millones de etiquetas diferentes.

[0191] En general, la asignación de códigos de barras de etiquetas únicas o no únicas en reacciones sigue los métodos y sistemas descritos en las solicitudes de patente de EE. UU. 20010053519, 20030152490, 20110160078 y las patentes

de EE. UU. n.º 6.582.908, n.º 7.537.898 y n.º 9.598.731. Las etiquetas se pueden vincular a ácidos nucleicos de muestra de forma aleatoria o no aleatoria.

[0192] En algunas formas de realización, los ácidos nucleicos marcados se secuencian después de cargarlos en una placa de micropocillos. La placa de micropocillos puede tener 96, 384 o 1536 micropocillos. En algunos casos, se introducen en una proporción esperada de etiquetas únicas a micropocillos. Por ejemplo, las etiquetas únicas se pueden cargar de modo que se carguen más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1.000, 5.000, 10.000, 50.000, 100.000, 500.000, 1.000.000, 10.000.000, 50.000.000 o 1.000.000.000 etiquetas únicas por muestra de genoma. En algunos casos, las etiquetas únicas se pueden cargar de modo que se carguen menos de aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1.000, 5.000, 10.000, 50.000, 100.000, 500.000, 1.000.000, 10.000.000, 50.000.000 o 1.000.000.000 de etiquetas únicas por muestra de genoma. En algunos casos, el número promedio de etiquetas únicas cargadas por muestra de genoma es menor o mayor que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1.000, 5.000, 10.000, 50.000, 100.000, 500.000, 1.000.000, 10.000.000, 50.000.000 o 1.000.000.000 de etiquetas únicas por muestra de genoma.

[0193] Un formato preferido utiliza de 20 a 50 etiquetas diferentes (por ejemplo, códigos de barras) ligadas a ambos extremos de los ácidos nucleicos objetivo. Por ejemplo, 35 etiquetas diferentes (por ejemplo, códigos de barras) ligadas a ambos extremos de las moléculas objetivo crean 35 x 35 permutaciones, lo que equivale a 1225 combinaciones de etiquetas para 35 etiquetas. Dichas cantidades de etiquetas son suficientes para que diferentes moléculas que tienen los mismos puntos de inicio y detención tengan una alta probabilidad (por ejemplo, al menos 94 %, 99,5 %, 99,99 %, 99,999 %) de recibir diferentes combinaciones de etiquetas. Otras combinaciones de códigos de barras incluyen cualquier número entre 10 y 500, por ejemplo, aproximadamente 15 x 15, aproximadamente 35 x 35, aproximadamente 75 x 75, aproximadamente 100 x 100, aproximadamente 250 x 250, aproximadamente 500 x 500.

[0194] En algunos casos, las etiquetas únicas pueden ser oligonucleótidos de secuencia predeterminados o aleatorios o semialeatorios. En otros casos, se puede utilizar una pluralidad de códigos de barras de modo que los códigos de barras no sean necesariamente únicos entre sí en la pluralidad. En este ejemplo, los códigos de barras se pueden ligar a moléculas individuales de modo que la combinación del código de barras y la secuencia a la que se puede ligar crea una secuencia única que se puede rastrear individualmente. Como se describe en el presente documento, la detección de códigos de barras no únicos en combinación con datos de secuencia de partes de inicio (inicio) y final (fin) de lecturas de secuencia puede permitir la asignación de una identidad única a una molécula particular. La longitud o el número de pares de bases de una lectura de secuencia individual también se puede utilizar para asignar una identidad única a dicha molécula. Como se describe en el presente documento, los fragmentos de una sola cadena de ácido nucleico a los que se les ha asignado una identidad única pueden permitir de ese modo la identificación posterior de fragmentos de la cadena parental.

3. Amplificación

[0195] Los ácidos nucleicos de muestra flanqueados por adaptadores pueden amplificarse mediante PCR y otros métodos de amplificación. La amplificación puede iniciarse mediante cebadores que se unen a los sitios de unión de cebadores en los adaptadores que flanquean una molécula de ADN que se va a amplificar. Los métodos de amplificación pueden implicar ciclos de desnaturalización, hibridación y extensión, resultantes del termociclado, o pueden ser isotérmicos como en la amplificación mediada por transcripción. Otros métodos de amplificación incluyen la reacción en cadena de la ligasa, la amplificación por desplazamiento de cadena, la amplificación basada en secuencias de ácidos nucleicos y la replicación basada en secuencias autosostenidas.

[0196] En algunas formas de realización, los métodos presentes comprenden realizar ligaduras de ADNbc "TV A" con adaptadores de cola T y de cola C, que dan como resultado la amplificación de al menos el 50, 60, 70 u 80 % de los ácidos nucleicos bicatenarios antes de la unión a los adaptadores. En algunas formas de realización, los métodos presentes aumentan la cantidad o el número de moléculas amplificadas en relación con los métodos de control realizados con adaptadores de cola T solos en al menos el 10, 15 o 20 %.

4. Conjuntos de cebos; fracciones de captura; enriquecimiento

[0197] Como se ha comentado anteriormente, los ácidos nucleicos (por ejemplo, en grupos preparados a partir de porciones de ADN de muestra particionada) pueden someterse a una etapa de captura, en la que se capturan moléculas que tienen secuencias objetivo para su posterior análisis. La captura objetivo puede implicar el uso de un conjunto de cebos que comprende cebos de oligonucleótidos marcados con una fracción de captura, como biotina o los otros ejemplos que se indican a continuación. Las sondas pueden tener secuencias seleccionadas para colocarlas en mosaico en un panel de regiones, como genes. En algunas formas de realización, un conjunto de cebos puede tener rendimientos de captura más altos y más bajos para conjuntos de regiones objetivo, como los del conjunto de regiones objetivo de secuencia variable y el conjunto de regiones objetivo epigenéticas, respectivamente, como se analiza en otra parte del presente documento. Dichos conjuntos de cebos se combinan con una muestra en condiciones que permiten la hibridación de las moléculas objetivo con los cebos. A continuación, las moléculas capturadas se aíslan utilizando la fracción de captura. Por ejemplo, una fracción de captura de biotina mediante estreptavidina basada en perlas. Dichos métodos se describen con más detalle, por ejemplo, en la patente estadounidense 9.850.523, emitida el 26 de diciembre de 2017.

[0198] Las fracciones de captura incluyen, sin limitación, biotina, avidina, estreptavidina, un ácido nucleico que comprende una secuencia de nucleótidos particular, un hapteno reconocido por un anticuerpo y partículas atraíbles magnéticamente. La fracción de extracción puede ser un miembro de un par de unión, tal como biotina/estreptavidina o hapteno/anticuerpo.

5 En algunas formas de realización, una fracción de captura que está unida a un analito es capturada por su par de unión que está unido a una fracción aislable, tal como una partícula atraíble magnéticamente o una partícula grande que puede sedimentarse mediante centrifugación. La fracción de captura puede ser cualquier tipo de molécula que permita la separación por afinidad para los ácidos nucleicos que portan la fracción de captura de los ácidos nucleicos que carecen de la fracción de captura. Las fracciones de captura ejemplares son biotina que permite la separación por afinidad mediante la unión a estreptavidina unida o enlazable a una fase sólida o un oligonucleótido, que permite la separación por afinidad mediante la unión a un oligonucleótido complementario unido o enlazable a una fase sólida.

[0199] En algunas formas de realización, se utiliza una colección de sondas específicas al objetivo en una etapa de captura. La colección utilizada para la captura del primer grupo (primera colección) puede comprender sondas de unión a la objetivo específicas para un conjunto de regiones objetivo de secuencia variable. La colección utilizada para la captura del segundo grupo (segunda colección) puede comprender sondas de unión a la objetivo específicas para un conjunto de regiones objetivo de secuencia variable y sondas de unión a la objetivo específicas para un conjunto de regiones objetivo epigenéticas.

[0200] Con respecto a cualquier colección que comprende sondas de unión a dianas específicas para un conjunto de regiones objetivo variables de secuencia y sondas de unión a dianas específicas para un conjunto de regiones objetivo epigenéticas (por ejemplo, una segunda colección como se describió anteriormente), en algunas formas de realización, el rendimiento de captura de las sondas de unión a dianas específicas para el conjunto de regiones objetivo variables de secuencia es mayor (por ejemplo, al menos 2 veces mayor) que el rendimiento de captura de las sondas de unión a dianas específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la colección de sondas específicas de dianas está configurada para tener un rendimiento de captura específico para el conjunto de regiones objetivo variables de secuencia mayor (por ejemplo, al menos 2 veces mayor) que su rendimiento de captura específico para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es al menos 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 veces mayor que el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es de 1,25 a 1,5, de 1,5 a 1,75, de 1,75 a 2, de 2 a 2,25, de 2,25 a 2,5, de 2,5 a 2,75, de 2,75 a 3, de 3 a 3,5, de 3,5 a 4, de 4 a 4,5, de 4,5 a 5, de 5 a 5,5, de 5,5 a 6, de 6 a 7, de 7 a 8, de 8 a 9, de 9 a 10, de 10 a 11, de 11 a 12, de 13 a 14 o de 14 a 15 veces mayor que el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es al menos 10 veces mayor que el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas, por ejemplo, de 10 a 20 veces mayor que el rendimiento de captura de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la colección de sondas específicas del objetivo está configurada para tener un rendimiento de captura específico para el conjunto de regiones objetivo de secuencia variable al menos 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 veces mayor que su rendimiento de captura para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la colección de sondas específicas del objetivo está configurada para tener un rendimiento de captura específico para el conjunto de regiones objetivo de secuencia variable de 1,25 a 1,5, de 1,5 a 1,75, de 1,75 a 2, de 2 a 2,25, de 2,25 a 2,5, de 2,5 a 2,75, de 2,75 a 3, de 3 a 3,5, de 3,5 a 4, de 4 a 4,5, de 4,5 a 5, de 5 a 5,5, de 5,5 a 6, de 6 a 7, de 7 a 8, de 8 a 9, de 9 a 10, de 10 a 11, de 11 a 12, de 13 a 14 o 14 a 15 veces mayor que su rendimiento de captura específico para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la colección de sondas específicas al objetivo está configurada para tener un rendimiento de captura específico para el conjunto de regiones objetivo de secuencia variable al menos 10 veces mayor que su rendimiento de captura para el conjunto de regiones objetivo epigenéticas, por ejemplo, 10 a 20 veces mayor que su rendimiento de captura para el conjunto de regiones objetivo epigenéticas.

[0201] La colección de sondas se puede configurar para proporcionar mayores rendimientos de captura para el conjunto de regiones objetivo de secuencia variable de diversas maneras, incluyendo concentración, diferentes longitudes y/o químicas (por ejemplo, que afectan la afinidad), y combinaciones de las mismas. La afinidad se puede modular ajustando la longitud de la sonda y/o incluyendo modificaciones de nucleótidos como se analiza a continuación.

[0202] En algunas formas de realización, las sondas específicas al objetivo específicas para el conjunto de regiones objetivo de secuencia variable están presentes en una concentración más alta que las sondas específicas al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la concentración de las sondas de unión a la objetivo específicas para el conjunto de regiones objetivo de secuencia variable es al menos 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 veces mayor que la concentración de las sondas de unión a la objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es de 1,25 a 1,5, de 1,5 a 1,75, de 1,75 a 2, de 2 a 2,25, de 2,25 a 2,5, de 2,5 a 2,75, de 2,75 a 3, de

3 a 3,5, de 3,5 a 4, de 4 a 4,5, de 4,5 a 5, de 5 a 5,5, de 5,5 a 6, de 6 a 7, de 7 a 8, de 8 a 9, de 9 a 10, de 10 a 11, de 11 a 12, de 13 a 14 o de 14 a 15 veces mayor que la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es al menos 2 veces mayor que la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En algunas formas de realización, la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable es al menos 10 veces mayor que la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas, por ejemplo, de 10 a 20 veces mayor que la concentración de las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas. En dichas formas de realización, la concentración puede referirse a la concentración promedio de masa por volumen de sondas individuales en cada conjunto.

[0203] En algunas formas de realización, las sondas específicas de objetivo específicas para el conjunto de regiones objetivo de secuencia variable tienen una afinidad más alta por sus dianas que las sondas específicas de objetivo específicas para el conjunto de regiones objetivo epigenéticas. La afinidad se puede modular de cualquier manera conocida por los expertos en la materia, incluyendo mediante el uso de diferentes químicas de sonda. Por ejemplo, ciertas modificaciones de nucleótidos, como la 5-metilación de citosina (en ciertos contextos de secuencia), modificaciones que proporcionan un heteroátomo en la posición de azúcar 2' y nucleótidos LNA, pueden aumentar la estabilidad de los ácidos nucleicos bicatenarios, lo que indica que los oligonucleótidos con dichas modificaciones tienen una afinidad relativamente más alta por sus secuencias complementarias. Véase, por ejemplo, Severin et al., *Nucleic Acids Res.* 39: 8740-8751 (2011); Freier et al., *Nucleic Acids Res.* 25: 4429-4443 (1997); Patente de EE. UU. N.º 9.738.894. Además, las longitudes de secuencia más largas generalmente proporcionarán una mayor afinidad. Otras modificaciones de nucleótidos, como la sustitución de la nucleobase hipoxantina por guanina, reducen la afinidad al reducir la cantidad de enlaces de hidrógeno entre el oligonucleótido y su secuencia complementaria. En algunas formas de realización, las sondas específicas del objetivo específicas para el conjunto de regiones objetivo de secuencia variable tienen modificaciones que aumentan su afinidad por sus objetivos. En algunas formas de realización, de manera alternativa o adicional, las sondas específicas del objetivo específicas para el conjunto de regiones objetivo epigenéticas tienen modificaciones que disminuyen su afinidad por sus objetivos. En algunas formas de realización, las sondas específicas del objetivo específicas para el conjunto de regiones objetivo de secuencia variable tienen longitudes promedio más largas y/o temperaturas de fusión promedio más altas que las sondas específicas del objetivo específicas para el conjunto de regiones objetivo epigenéticas. Estas formas de realización se pueden combinar entre sí y/o con diferencias en la concentración como se discutió anteriormente para lograr una diferencia de pliegue deseada en el rendimiento de captura, como cualquier diferencia de pliegue o rango de la misma descrito anteriormente.

[0204] En algunas formas de realización, las sondas específicas al objetivo comprenden una fracción de captura. La fracción de captura puede ser cualquiera de las fracciones de captura descritas en el presente documento, por ejemplo, biotina. En algunas formas de realización, las sondas específicas al objetivo están unidas a un soporte sólido, por ejemplo, de forma covalente o no covalente, como, por ejemplo, a través de la interacción de un par de unión de fracciones de captura. En algunas formas de realización, el soporte sólido es una perla, como una perla magnética.

[0205] En algunas formas de realización, las sondas específicas del objetivo específicas para el conjunto de regiones objetivo de secuencia variable y/o las sondas específicas del objetivo específicas para el conjunto de regiones objetivo epigenéticas son un conjunto cebo como se discutió anteriormente, por ejemplo, sondas que comprenden fracciones de captura y secuencias seleccionadas para distribuirse en mosaico a través de un panel de regiones, tales como genes.

[0206] En algunas formas de realización, las sondas específicas al objetivo se proporcionan en una única composición. La única composición puede ser una solución (líquida o congelada). Alternativamente, puede ser un liofilizado.

[0207] Alternativamente, las sondas específicas al objetivo pueden proporcionarse como una pluralidad de composiciones, por ejemplo, que comprenden una primera composición que comprende sondas específicas para el conjunto de regiones objetivo epigenéticas y una segunda composición que comprende sondas específicas para el conjunto de regiones objetivo de secuencia variable. Estas sondas pueden mezclarse en proporciones apropiadas para proporcionar una composición de sonda combinada con cualquiera de las diferencias de pliegue anteriores en la concentración y/o el rendimiento de captura. Alternativamente, pueden usarse en procedimientos de captura separados (por ejemplo, con alícuotas de una muestra o secuencialmente con la misma muestra) para proporcionar una primera y una segunda composición que comprenden regiones objetivo epigenéticas capturadas y regiones objetivo de secuencia variable, respectivamente.

a. Sondas específicas para regiones objetivo epigenéticas

[0208] Las sondas para el conjunto de regiones objetivo epigenéticas pueden comprender sondas específicas para uno o más tipos de regiones objetivo que probablemente diferencien el ADN de células neoplásicas (por ejemplo, tumorales o cancerosas) de células sanas, por ejemplo, células circulantes no neoplásicas. En este documento se analizan en detalle tipos ejemplares de dichas regiones, por ejemplo, en las secciones anteriores relativas a los conjuntos capturados. Las sondas para el conjunto de regiones objetivo epigenéticas también pueden comprender sondas para una o más regiones de control, por ejemplo, como se describe en este documento.

[0209] En algunas formas de realización, las sondas para el conjunto de sondas de la región objetivo epigenética tienen una huella de al menos 100 kb, por ejemplo, al menos 200 kb, al menos 300 kb o al menos 400 kb. En algunas formas de realización, las sondas para el conjunto de la región objetivo epigenética tienen una huella en el intervalo de 100-1.000 kb, por ejemplo, 100-200 kb, 200-300 kb, 300-400 kb, 400-500 kb, 500-600 kb, 600-700 kb, 700-800 kb, 800-900 kb y 900-1.000 kb.

i. Regiones objetivo variables de hipermetilación

[0210] En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas comprenden sondas específicas para una o más regiones objetivo variables de hipermetilación. Las regiones objetivo variables de hipermetilación pueden ser cualquiera de las establecidas anteriormente. Por ejemplo, en algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para una pluralidad de loci enumerados en la Tabla 1, p. ej., al menos el 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los loci enumerados en la Tabla 1. En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para una pluralidad de loci enumerados en la Tabla 2, p. ej., al menos el 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los loci enumerados en la Tabla 2. En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para una pluralidad de loci enumerados en la Tabla 1 o la Tabla 2, p. ej., al menos 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los loci enumerados en la Tabla 1 o la Tabla 2. En algunas formas de realización, para cada locus incluido como región objetivo, puede haber una o más sondas con un sitio de hibridación que se une entre el sitio de inicio de la transcripción y el codón de parada (el último codón de parada para genes que se empalman alternativamente) del gen. En algunas formas de realización, las una o más sondas se unen dentro de los 300 pb de la posición enumerada, por ejemplo, dentro de los 200 o 100 pb. En algunas formas de realización, una sonda tiene un sitio de hibridación que se superpone a la posición enumerada anteriormente. En algunas formas de realización, las sondas específicas para las regiones objetivo de hipermetilación incluyen sondas específicas para uno, dos, tres, cuatro o cinco subconjuntos de regiones objetivo de hipermetilación que colectivamente muestran hipermetilación en uno, dos, tres, cuatro o cinco cánceres de mama, colon, riñón, hígado y pulmón.

[0211] En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para al menos uno o más genes enumerados en la Tabla 6, por ejemplo, al menos 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1.000, 1100, 1150 o 1200 genes enumerados en la Tabla 6. La hipermetilación de estos genes puede ser útil para detectar contribuciones de células inmunes a una muestra de ADN. En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para una pluralidad de genes enumerados en la Tabla 6, p. ej., al menos el 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los genes enumerados en la Tabla 6. En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipermetilación comprenden sondas específicas para todos los genes enumerados en la Tabla 6.

Tabla 6. Regiones objetivo variables de hipermetilación ejemplares (genes o partes de las mismas)

GENES						
CCL5	RSAD2	DBNDD1	CD80	CD2	ITPK1	PRR5
CD19	ISG15	SNRPB	CEACAM4	G0S2	JAG1	PTGS2
CD37	MMP3	MNAT1	SH2D1A	MT1G	MMRN2	PTPN11
CD3D	MARCKSL1	NUP35	SAA2-SAA4	CARD16	NFIL3	PVRIG
CD3E	IGF2R	TCEB1	TMEM150B	MT1H	PDGFB	RELA
CD3G	LAMB1	HSPB11	PYHIN1	SERPINA1	PGPEP1	RICTOR
CD3Z	SPARC	C19orf48	CCL7	ACP5	RPL7A	RIPK2
CD79A	FN1	ID3	PDCD1	CXCL5	SERPINB5	SPP1
CD79B	ITGA4	IPO4	BNC1	CD74	SERPINH1	STAT4
CD8A	SMO	FARSB	ICOS	C1QB	STC1	TNFRSF14
CD8B1	MMP19	EIF4G1	ITK	HLA-DRB5	TNFAIP6	TNFRSF25
IGHG3	ITGB8	SKA1	CLEC4E	FADS3	VCAN	TNFSF9
IGJ	ITGA5	MFSD11	GZMK	IL15RA	VEGFA	TRAT1
IGLC1	NID1	PLAUR	UBASH3A	MFI2	VEGFB	TSLP
CD14	TIMP1	MARVELD2	CRTAM	IL4I1	VEGFC	VTCN1

(Continuación)

	GENES						
5	LCK	SEMA3F	MCM3	EOMES	HLA-A	ATF3	CCL13
	LTB	RHOQ	DHFR	P2RY10	TM4SF19	BATF3	CCL14
	MS4A1	CTNNB1	RNF41	NCF1	MMP9	CD1C	CCL18
10	CORO1A	MMP2	ID2	CARD17	ACY3	CD36	CCL19
	MNDA	SERPINE1	H2AFZ	ZBP1	KRT14	CD4	CCL20
	CCRL2	EPHB2	CDK2	TNFRSF9	MT1X	CDC20	CCL21
15	SLC7A7	COL16A1	NCLN	PDCD1LG2	MT1M	CTSW	CCL22
	HLA-DMA	EPHA2	ZWILCH	IL12RB1	IFITM1	CXCL1	CCL3/L1
	FYB	TNC	DYNLT1	LILRA6	IFI35	DTX3L	CSF1
20	RNASE6	JUP	C16orf61	CXCR3	CD2	ITPK1	CSF2
	TLR2	ITGA3	SLC25A40	SLAMF1	G0S2	JAG1	CSF2RB
	CTSC	TCF7L2	RHOC	SLAMF6	MT1G	MMRN2	CSF3R
25	LILRB4	COL3A1	CCT5	RHOH	CARD16	NFIL3	CX3CL1
	PSCDBP	CDH6	PDIA4	GPR114	MT1H	PDGFB	CX3CR1
	CTSS	WNT2B	SNRPA	SLA2	SERPINA1	PGPEP1	CXCL12
30	RASSF4	ADAM9	RBM14	IL18RAP	ACP5	RPL7A	CXCL14
	MSN	DSP	PDLIM7	GPR171	CXCL5	SERPINB5	CXCL16
	CYBB	HSPG2	PITPNC1	SCIMP	CD74	SERPINH1	CXCL2
35	LAPTM5	ARHGAP1	TPM3	P2RY13	C1QB	STC1	CXCL3
	DOCK2	ITGB5	CORO1C	LAX1	HLA-DRB5	TNFAIP6	CXCL6
	FCGR1A	IGFBP5	ERLIN1	AOAH	FADS3	VCAN	CXCL8
40	EVI2B	ARHGDIA	PAICS	CXCL11	GAL	KLRD1	CXCR2
	ADCY7	LRP1	TPRKB	CD38	CXCL13	MRC1	GNG4
	CD48	IGFBP2	SKA2	GRAP2	RAC2	PSMB10	IL10
45	ARHGAP1 5	CTNNA1	MYBL1	WDFY4	IL1R2	PSMB5	IL11
	ARRB2	LRRC17	SH3BP5L	LTA	LRFN4	PSMB8	IL11RA
	SYK	MMP14	BRCA2	GZMH	CLEC4A	PSMC4	IL12RB2
50	BTK	NEO1	SAR1A	PTPN22	SLAMF8	THBD	IL16
	TNFAIP3	EFNA5	POLR3K	IRF4	SH2D2A	TNF	IL1A
	FCGR2A	ITGB3	MRPS28	GNLY	CDC25B	TRIM21	IL1B
55	VSIG4	EPHB3	NUP107	CCR5	PSME2	UBA7	IL1RN
	FPRL2	CD44	TUBG1	SNX20	GNGT2	UBB	IL22RA1
	IL10RA	IGFBP4	PNN	LILRB5	TYROBP	UBE2C	IL24
60	IFI16	TNFRSF1A	FAM167A	RGS18	AIF1	ULBP2	IL33
	ITGB2	RAC1	RFC3	EMR1	WARS	VHL	IL34
	IL7R	PXN	MYL6	TIGIT	CD300A	AKT1	IL6
65	TBXAS1	PLAT	MCM7	SIGLEC7	C1QC	APC	IL6R
	FMNL1	COL8A1	MAGOHB	GPR65	ARPC1B	BAD	JAK2
	FLI1	WNT8B	FAM89B	TNFRSF17	CIITA	BAX	JAK3
	RASSF2	IGFBP3	TOMM40	ADAMDEC1	HAVCR2	BBC3	PF4
	LYZ	RHOA	CDCA4	HK3	SPI1	BCL2	PIK3R5
	CD163	EPHB4	MT3	LILRB2	MIR3123	BCL2L1	SHC2

(Continuación)

GENES						
CD97	MMP1	MTHFD1	CTLA4	ACAA2	BCL6B	STAT2
CCL2	PAK1	PSMD12	SIGLEC12	KRTAP19-8	BID	STAT3
FCGR2B	MTA1	MYBL2	CXorf21	LPCAT1	BIRC3	CBLC
MERTK	THBS2	CKLF	SPIB	CARTPT	BIRC5	CD47
CD84	CSPG2	NRIP3	CLEC4G	ARHGAP26	BLM	CNTFR
CD53	MMP17	EZR	CD300C	RPA3	CASP3	GHR
CD86	CD59	C12orf24	CD300E	MIR129-1	CASP8	GZMM
HMHA1	DVL3	GPLD1	LILRB1	FAM150A	CASP9	IFITM2
CTSL	RHOB	SRM	LAG3	KCNB2	CDH1	IRF9
EVI2A	COL6A3	RAB3B	HMGA2	TERF1	FADD	JAK1
TNFRSF1B	NOTCH2	NLN	VNN1	KCNK9	HMGB1	KLRB1
CXCR4	BSG	MT1F	IL2RA	PTK2	LY96	KLRK1
LCP1	MMP11	TNFRSF12A	NCR3	GPR20	RIPK1	LIF
SAMHD1	COL1A2	TPI1	CCR8	TSNARE1	RIPK3	PRLR
CPVL	ZYX	HAS2	TFEC	MIR4475	ROCK1	SIRPA
HLA-DRB1	RND3	APOO	CCR2	IGFBPL1	TICAM1	SPRY4
C13orf18	THBS1	FBXO41	STAP1	LOC100129316	TLR3	DDB2
GIMAP4	RHOG	MRPL37	IDO1	QRFP	TLR4	ERCC3
SAMSN1	ICAM1	GSTCD	MYO1G	CAMSAP1	TNFRSF10B	FANCA
PLCG2	LAMA4	SDC1	SIT1	GPR37L1	TNFSF10	MGMT
OSBPL3	DVL1	WDR54	CD209	CCDC85C	TP53	NEIL1
RUNX3	PAK2	RNF138	LILRA5	GACAT2	BNIP3	PARP4
FCGR3A	COL6A1	APITD1	SIGLEC1	LINC01443	DEPTOR	POLR2A
AMPD3	FGD1	RMND5B	UBD	CDH2	HIF1A	TNKS
MYO1F	CEP78	ENO1	SPN	MIR4436A	HRAS	UBE2T
CECR1	LSM3	MAP3K8	CXCL9	INSM1	KRAS	XCL1
LYN	LRRC40	TMEM130	CD70	LOC101929625	MAP3K7	XCL2
MPP1	STK17A	SNX17	GZMB	COL18A1	MAPK10	ARID1A
LRMP	RPN1	KRR1	SIGLEC10	OPA1-AS1	MTOR	BRD3
FGL2	JUNB	TAGLN	CD300LF	C9	NRAS	BRD4
NCKAP1L	NUP85	PA2G4	CCL8	MIR585	PIK3CA	DNMT1
HCLS1	FLNC	RUVBL1	PTPN7	MAD1L1	PIK3CD	HDAC11
SELL	HMG2	SNRPD1	CCL4	CPA2	PIK3R1	HDAC3
CASP1	RPP40	LOXL2	CMKLR1	LOC100130298	PIK3R2	HDAC4
SELPLG	UQCRL10	POLE2	ICAM3	C8orf34	PRKAA2	HDAC5
CD33	AIMP2	MAPRE1	SIRPG	OC90	PRKACB	HELLS
GPNUMB	CHEK1	IMP4	NLRP3	TRAPPC9	PTEN	HMGA1
NCF2	VTA1	EMP2	MARCO	MED27	RPS6KB1	KAT2B
FNBP1	EXOSC8	PSMD2	CCL4L1	MLH1	RPTOR	MAP3K12
IL18	CENPO	MET	CD180	MSH2	ANLN	BMP2
B2M	PNO1	IFRD2	SIGLEC9	MSH6	ATM	GAS1

(Continuación)

GENES						
SP140	SLC16A1	LMNB2	KRT6B	PMS2	CCNA1	GLI1
FCER1G	WDR77	PLOD2	NFAM1	BRCA1	CCNB1	IHH
LCP2	UBE2J1	NCEH1	GPR84	TERT	CCND1	PRKX
LY86	NOP16	NME1	KLK5	EphA2	CCND3	WNT10A
LAIR1	NUDT1	STRA13	CCL23	cGAS	CCNO	WNT11
IFI30	SMC2	ACTL6A	ZBED2	STING	CDC25C	WNT2
TNFSF13B	SLC25A5	DLEU1	CD28	CD8	CDK6	WNT3A
LST1	NUPL1	SNRPA1	CEACAM21	PDL2	CDKN1A	WNT4
FGR	DLEU2	CBX1	PRF1	ASXL3	CDKN1C	WNT5A
NPL	PDAP1	LYAR	CSF3	SPEDF	CDKN2A	WNT5B
PLEK	CCBL2	PTPLB	KIAA0226L	PSMA	CDKN2B	WNT7B
PTPRC	COX17	PFN1	CXCR6	GPC3	CENPF	ALDOA
GNPTAB	BCCIP	CENPJ	CCL24	RAD51	CEP55	EGF
SLC1A3	PLG	COTL1	ARHGAP9	ACVR1	EXO1	EGFR
HCK	RGS8	SPRYD7	LILRB3	APLNR	H2AFX	ERBB2
NPC2	SNRPC	USPL1	VNN2	CDC27	MELK	HK1
C3AR1	PLK4	MRPL12	ZAP70	CDK4	MKI67	HK2
PIK3CG	NUTF2	ADAMTS1	TNFSF8	HLA-C	MYC	IFNGR1
DAPK1	LSM4	GLRX3	HEPHL1	HLA-E	NBN	IFNGR2
ALOX5AP	SMS	WSB2	CCL3	HLA-G	POLD1	LDHA
CSF1R	EBNA1BP2	MRPS16	CLEC10A	IL1R	PRKCA	NOS2
CUGBP2	C13orf27	DCLRE1B	CCL3L3	IL2RB	RAD50	PDK1
APOE	VDAC1	MKKS	CCR4	INFR1	RAD51C	PFKFB3
APOC1	PSMD14	C3orf26	ITGAL	INFR2	RB1	SLC2A1
CD52	MYCBP	CPEB4	SAA2	IRF1	RBL2	TFRC
LHFPL2	SMURF2	SPAG17	GZMA	IRF2	RRM2	CD276
C1orf54	GNG11	MLF1IP	BCL2A1	KIR2DS2	TYMS	CD6
IKZF1	F3	UAP1	RNASE2	KIR2DS3	ADORA2A	CDH5
SH2B3	BRIP1	COQ2	SAA1	PIAS4	BTLA	CLEC14A
WIPF1	HNRNPA2B1	WDHD1	NKG7	PTPN2	CD247	CLEC5A
IFI44	DCK	DCBLD2	MT2A	SERPINB3	CD27	CLEC7A
IFI44L	ALKBH7	KIAA0090	FAM26F	SERPINB4	CD274	CLECL1
DDX58	HN1L	SAR1B	PSMB9	SPOP	CD40	ICAM2
IFI6	TPM1	PSMA7	TYMP	TAPBP	CD40LG	ICAM5
IFI27	HYLS1	PSMC3	CD7	TGFBR1	CD69	ITGA1
IFIT2	HAUS1	COPS6	SYT12	CD58	CHUK	ITGA2
IFIT1	NUP93	DUT	S100A7	ANGPT1	DPP4	ITGAE
IFIT3	SNRPE	PPIH	HLA-F	ANGPT2	EGR1	ITGAM
CXCL10	ITGA6	PHF19	MT1E	ANGPTL4	FYN	ITGAX
MX1	CENPN	TPM2	ISG20	CCND2	ICOSLG	NCAM1
OAS1	C11orf24	MCTS1	C1QA	CCNE1	IKBKB	NECTIN1
OAS2	GGH	EIF4EBP1	TNFRSF18	CES3	IKBKG	PECAM1

(Continuación)

GENES						
OAS3	PFKP	HNRNPR	IL32	DLL4	IL15	PVR
HERC5	FARSA	IFNG	SOCS1	E2F3	IL17A	SELE
SAMD9	EIF2AK1	TLR8	HLA-DQB1	EDN1	IL18R1	SELP
HERC6	CENPW	FASLG	IFITM3	EZH2	IL2	THY1
DDX60	TUBA4A	KIR2DL4	PI3	FGF18	IL4	VCAM1
RTP4	TRA2B	IL21R	HCST	FGFR1	LY9	EIF2AK2
IFIH1	UMPS	GBP5	SUSD3	FLT1	NECTIN2	FLNB
STAT1	MRT04	ZNF683	HLA-B	FSTL3	NFATC2	GBP1
TAP1	NUDT15	PLA2G2D	IL2RG	HEY1	NFKB1	GBP2
OASL	PGM2	TBX21	HLA-DRA	ITGAV	NFKBIA	GBP4
TPSD1	GATA2	SERPINF1	FCER1A	DAPP1	CCR7	EREG
SLC18A2	HPGDS	PLAC8	CD1A	C10ORF54	XCR1	CD81
IL1RL1	TCF4	JCHAIN	AXL	ANPEP	MTSS1	PRDM1
VWA5A	MZB1	SLC15A4	CD1E	CST7	MSA4A7	ABL2
TPSAB2	IL3RA	ITM2C	PKIB	FLT3	LILRA1	SDC4
IFNA1	PDGFA	MMP7	CRABP2	TNFRSF4	FZD9	PNOC
IFNAR1	PDGFRB	NID2	DAB2	TNFSF12	GPC4	HLA-DR
IRF3	PLA1A	RELN	FCAR	TNFSF13	SFRP1	LIN
IRF5	PLA2G2A	AQP9	FCN1	TNFSF18	SFRP4	CD11b
IRF7	RASAL1	CEBPB	FOSL1	TNFSF4	SOX11	CD15
BLK	RASGRF1	ERO1A	FPR1	APH1B	SOX2	CD11c
CD5	RELB	FBP1	FPR3	DLL1	FOXP3	CD34
CD96	TGFB1	G6PD	IER3	DTX4	NCR1	CD83
F2RL1	TGFB2	GLS	OLR1	HES1	CEACAM3	PD-L1
SLAMF7	TGFB3	GOT1	PDZK1IP1	JAG2	FCGR3A/B	CD45
DUSP1	TGFB2	GOT2	S100A12	MAML2	SIGLEC5	CD124
DUSP2	A2M	LDHB	S100A8	MFNG	CPA3	CD66
DUSP5	COL11A1	PC	S100A9	NOTCH1	HDC	CD3
FAS	COL11A2	PCK2	SIGLEC8	TCL1A	MS4A2	CD13
FGF13	COL17A1	PFKM	SIRPB2	ACVR1C	TPSAB1/B2	TPSAB1
FGF9	COL4A5	PKM	SLC11A1	BAMBI	CD68	LILRA4
KDR	COL5A1	SGK1	TLR1	ID4	MS4A4A	CD16
KIT	COMP	SLC1A5	TREM1	INHBA	CD244	PLTP
MAP3K5	LAMA1	SLC7A5	TREM2	SMAD5	PTGER4	ADM
NF1	LAMB3	AREG	NFKBIE	AXIN1	HSD11B1	CSTB
NFKB2	LAMC2	ARG1	TNFRSF11A	DKK1	FAM30A	PPARG
NGFR	LTBP1	C5AR1	TNFRSF11B	FZD8	FCRL2	RGCC

ii. Regiones objetivo variables de hipometilación

[0212] En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas comprenden sondas específicas para una o más regiones objetivo variables de hipometilación. Las regiones objetivo variables de hipometilación pueden ser cualquiera de las establecidas anteriormente. Por ejemplo, las sondas específicas para una o más regiones objetivo variables de hipometilación pueden incluir sondas para regiones tales como elementos repetidos, por ejemplo, elementos LINE1, elementos Alu, repeticiones en tándem centroméricas, repeticiones en tándem

pericentroméricas y ADN satélite, y las regiones intergénicas que normalmente están metiladas en células sanas pueden mostrar una metilación reducida en células tumorales.

[0213] En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipometilación incluyen sondas específicas para elementos repetidos y/o regiones intergénicas. En algunas formas de realización, las sondas específicas para elementos repetidos incluyen sondas específicas para uno, dos, tres, cuatro o cinco elementos de LINE I, elementos Alu, repeticiones en tándem centroméricas, repeticiones en tándem pericentroméricas y/o ADN satélite.

[0214] Las sondas ejemplares específicas para regiones genómicas que muestran hipometilación asociada al cáncer incluyen sondas específicas para los nucleótidos 8403565-8953708 y/o 151104701-151106035 del cromosoma humano 1. En algunas formas de realización, las sondas específicas para regiones objetivo variables de hipometilación incluyen sondas específicas para regiones que se superponen o comprenden los nucleótidos 8403565-8953708 y/o 151104701-151106035 del cromosoma humano 1.

iii. Regiones de unión de CTCF

[0215] En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas incluyen sondas específicas para regiones de unión de CTCF. En algunas formas de realización, las sondas específicas para regiones de unión de CTCF comprenden sondas específicas para al menos 10, 20, 50, 100, 200 o 500 regiones de unión de CTCF, o 10-20, 20-50, 50-100, 100-200, 200-500 o 500-1.000 regiones de unión de CTCF, por ejemplo, como las regiones de unión de CTCF descritas anteriormente o en uno o más de los artículos CTCFBSDB o Cuddapah et al., Martin et al. o Rhee et al. citados anteriormente. En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas comprenden al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 750 pb o al menos 1.000 pb de regiones corriente arriba y corriente abajo de los sitios de unión de CTCF.

iv. Sitios de inicio de la transcripción

[0216] En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas incluyen sondas específicas para sitios de inicio de la transcripción. En algunas formas de realización, las sondas específicas para los sitios de inicio de la transcripción comprenden sondas específicas para al menos 10, 20, 50, 100, 200 o 500 sitios de inicio de la transcripción, o 10-20, 20-50, 50-100, 100-200, 200-500 o 500-1.000 sitios de inicio de la transcripción, por ejemplo, como los sitios de inicio de la transcripción enumerados en DBTSS. En algunas formas de realización, las sondas para el conjunto de regiones objetivo epigenéticas comprenden sondas para secuencias de al menos 100 pb, al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 750 pb o al menos 1.000 pb corriente arriba y corriente abajo de los sitios de inicio de la transcripción.

v. Amplificaciones focales

[0217] Como se señaló anteriormente, aunque las amplificaciones focales son mutaciones somáticas, se pueden detectar mediante secuenciación basada en la frecuencia de lectura de una manera análoga a los enfoques para detectar ciertos cambios epigenéticos tales como cambios en la metilación. Como tal, las regiones que pueden mostrar amplificaciones focales en el cáncer se pueden incluir en el conjunto de regiones objetivo epigenéticas, como se discutió anteriormente. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo epigenéticas incluyen sondas específicas para amplificaciones focales. En algunas formas de realización, las sondas específicas para amplificaciones focales incluyen sondas específicas para uno o más de AR, BRAF, CCND1, CCND2, CCNE1, CDK4, CDK6, EGFR, ERBB2, FGFR1, FGFR2, KIT, KRAS, MET, MYC, PDGFRA, PIK3CA y RAF1. Por ejemplo, en algunas formas de realización, las sondas específicas para amplificaciones focales incluyen sondas específicas para uno o más de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 o 18 de los objetivos anteriores.

vi. Regiones de control

[0218] Puede ser útil incluir regiones de control para facilitar la validación de datos. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo epigenéticas incluyen sondas específicas para regiones metiladas de control que se espera que estén metiladas en esencialmente todas las muestras. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo epigenéticas incluyen sondas específicas para regiones hipometiladas de control que se espera que estén hipometiladas en esencialmente todas las muestras.

b. Sondas específicas para regiones objetivo de secuencia variable

[0219] Las sondas para el conjunto de regiones objetivo de secuencia variable pueden comprender sondas específicas para una pluralidad de regiones que se sabe que sufren mutaciones somáticas en el cáncer. Las sondas pueden ser específicas para cualquier conjunto de regiones objetivo de secuencia variable descrito en el presente documento. En el presente documento se analizan en detalle conjuntos de regiones objetivo de secuencia variable a modo de ejemplo, por ejemplo, en las secciones anteriores relativas a los conjuntos capturados.

[0220] En algunas formas de realización, el conjunto de sondas de región objetivo de secuencia variable tiene una huella de al menos 10 kb, por ejemplo, al menos 20 kb, al menos 30 kb o al menos 40 kb. En algunas formas de realización, el conjunto de sondas de región objetivo epigenética tiene una huella en el intervalo de 10-100 kb, por ejemplo, 10-20 kb, 20-30 kb, 30-40 kb, 40-50 kb, 50-60 kb, 60-70 kb, 70-80 kb, 80-90 kb y 90-100 kb.

[0221] En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos una porción de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65 o 70 de los genes de la Tabla 3. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65 o 70 de los SNV de la Tabla 3. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos 1, al menos 2, al menos 3, al menos 4, al menos 5 o 6 de las fusiones de la Tabla 3. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos una parte de al menos 1, al menos 2 o 3 de los indeles de la Tabla 3. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos una parte de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70 o 73 de los genes de la Tabla 4. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70 o 73 de los SNV de la Tabla 4. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo variables de secuencia comprenden sondas específicas para al menos 1, al menos 2, al menos 3, al menos 4, al menos 5 o 6 de las fusiones de la Tabla 4. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo variables de secuencia comprenden sondas específicas para al menos una parte de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29, al menos 30, al menos 31, al menos 32, al menos 33, al menos 34, al menos 35, al menos 36, al menos 37, al menos 38, al menos 39, al menos 40, al menos 41, al menos 42, al menos 43, al menos 44, al menos 45, al menos 46, al menos 47, al menos 48, al menos 49, al menos 50, al menos 51, al menos 52, al menos 53, al menos 54, al menos 55, al menos 56, al menos 57, al menos 58, al menos 59, al menos 60, al menos 61, al menos 62, al menos 63, al menos 64, al menos 65, al menos 66, al menos 67, al menos 68, al menos 69, al menos 70, al menos 71, al menos 72, al menos 73, al menos 74, al menos 75, al menos 76, al menos 77, al menos 78, al menos 79, al menos 80, al menos 81, al menos 82, al menos 83, al menos 84, al menos 85, al menos 86, al menos 87, al menos 88, al menos 89, al menos 90, al menos 91, al menos 92, al menos 93, al menos 94, al menos 95, al menos 96, al menos 97, al menos 98, al menos 99 o al menos 100 de los indeles de la Tabla 4. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para al menos una porción de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19 o al menos 20 de los genes de la Tabla 5.

[0222] En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para regiones objetivo de al menos 10, 20, 30 o 35 genes relacionados con el cáncer, tales como AKT1, ALK, BRAF, CCND1, CDK2A, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FOXL2, GATA3, GNA11, GNAQ, GNAS, HRAS, IDH1, IDH2, KIT, KRAS, MED12, MET, MYC, NFE2L2, NRAS, PDGFRA, PIK3CA, PPP2RIA, PTEN, RET, STK11, TP53 y U2AF1.

[0223] En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para regiones objetivo de al menos un gen relacionado con el cáncer, tal como uno o más de HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, DQA1, DQB1, DPA1, DPB1, DRB1, DRB3, DRB4, DRB5, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5A, KIR2DL5B, KIR2DP1, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, KIR3DS1, IFNA1, IFNA10, IFNA13, IFNA14, IFNA16, IFNA17, IFNA2, IFNA21, IFNA22P, IFNA4, IFNA5, IFNA6, IFNA7, IFNA8, IFNB1, IFNE, IFNW1, MIR31 y PTPLAD2. En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo de secuencia variable comprenden sondas específicas para regiones objetivo para uno o más genes que comprenden mutaciones de puntos críticos de la vacuna contra el cáncer. En algunas formas de realización, una mutación de punto crítico de la vacuna contra el cáncer es una mutación que se ha producido en un cáncer en una pluralidad de sujetos diferentes y que puede actuar como un neoantígeno del cáncer. En algunas formas de realización, genes como BRAF y EGFR pueden comprender mutaciones de puntos críticos de la vacuna contra el cáncer. En algunas formas de realización, las mutaciones de puntos críticos de la vacuna contra el cáncer se pueden obtener de bases de datos sobre el cáncer como, entre otras, el Programa Atlas del Genoma del Cáncer (TCGA) y el Catálogo de Mutaciones Somáticas en el Cáncer (COSMIC). En algunas formas de realización, las sondas específicas para el conjunto de regiones objetivo variables de secuencia comprenden sondas específicas para regiones objetivo para uno o más genes virales o bacterianos. En algunas formas de realización, el virus puede ser un virus del papiloma humano (HPV) o un virus de Epstein Barr (EBV).

c. Composiciones de sondas

[0224] En algunas formas de realización, se proporciona una única composición que comprende sondas para el conjunto de regiones objetivo variables de secuencia y sondas para el conjunto de regiones objetivo epigenéticas. Las sondas

pueden proporcionarse en dicha composición en cualquier proporción de concentración descrita en el presente documento.

[0225] En algunas formas de realización, se proporciona una primera composición que comprende sondas para el conjunto de regiones objetivo epigenéticas y una segunda composición que comprende sondas para el conjunto de regiones objetivo de secuencia variable. La relación de la concentración de las sondas en la primera composición con la concentración de las sondas en la segunda composición puede ser cualquiera de las relaciones descritas en el presente documento.

IV. SISTEMAS INFORMÁTICOS

[0226] Los métodos de la presente divulgación se pueden implementar utilizando, o con la ayuda de, sistemas informáticos. Por ejemplo, dichos métodos, que pueden comprender: dividir el ADN de la muestra en una pluralidad de particiones, comprendiendo la pluralidad al menos una partición hipermetilada y una partición hipometilada; etiquetar diferencialmente el ADN de la partición hipermetilada y el ADN de la partición hipometilada; preparar un primer conjunto que comprende al menos una parte del ADN de la partición hipometilada; preparar un segundo conjunto que comprende al menos una parte del ADN de la partición hipermetilada; capturar al menos un primer conjunto de regiones objetivo del primer grupo, en donde el primer conjunto comprende regiones objetivo de secuencia variable; capturar una segunda pluralidad de conjuntos de regiones objetivo del segundo grupo, en donde la segunda pluralidad comprende regiones objetivo de secuencia variable y regiones objetivo epigenéticas; secuenciar las moléculas de ADN capturadas; obtener una pluralidad de lecturas de secuencia generadas por un secuenciador de ácidos nucleicos a partir de la secuenciación de las moléculas de ADN capturadas; mapear la pluralidad de lecturas de secuencia a una o más secuencias de referencia para generar lecturas de secuencia mapeadas; y procesar las lecturas de secuencia mapeadas correspondientes al conjunto de regiones objetivo de secuencia variable y al conjunto de regiones objetivo epigenéticas para determinar la probabilidad de que el sujeto tenga cáncer.

[0227] La FIG. 2 muestra un sistema informático 201 que está programado o configurado de otro modo para implementar los métodos de la presente divulgación. El sistema informático 201 puede regular diversos aspectos de la preparación, secuenciación y/o análisis de muestras. En algunos ejemplos, el sistema informático 201 está configurado para realizar la preparación y el análisis de muestras, incluida la secuenciación de ácidos nucleicos.

[0228] El sistema informático 201 incluye una unidad central de procesamiento (CPU, también "procesador" y "procesador informático" en el presente documento) 205, que puede ser un procesador de un solo núcleo o de múltiples núcleos, o una pluralidad de procesadores para procesamiento en paralelo. El sistema informático 201 también incluye una memoria o ubicación de memoria 210 (por ejemplo, memoria de acceso aleatorio, memoria de solo lectura, memoria flash), una unidad de almacenamiento electrónico 215 (por ejemplo, disco duro), una interfaz de comunicación 220 (por ejemplo, adaptador de red) para comunicarse con uno o más sistemas diferentes, y dispositivos periféricos 225, como caché, otra memoria, almacenamiento de datos y/o adaptadores de pantalla electrónica. La memoria 210, la unidad de almacenamiento 215, la interfaz 220 y los dispositivos periféricos 225 están en comunicación con la CPU 205 a través de una red de comunicación o bus (líneas continuas), como una placa base. La unidad de almacenamiento 215 puede ser una unidad de almacenamiento de datos (o repositorio de datos) para almacenar datos. El sistema informático 201 puede estar acoplado operativamente a una red informática 230 con la ayuda de la interfaz de comunicación 220. La red informática 230 puede ser Internet, una red de Internet y/o extranet, o una intranet y/o extranet que esté en comunicación con Internet. La red informática 230 en algunos casos es una red de telecomunicaciones y/o de datos. La red informática 230 puede incluir uno o más servidores informáticos, que pueden permitir la computación distribuida, como la computación en la nube. La red informática 230, en algunos casos con la ayuda del sistema informático 201, puede implementar una red de igual a igual, que puede permitir que los dispositivos acoplados al sistema informático 201 se comporten como un cliente o un servidor.

[0229] La CPU 205 puede ejecutar una secuencia de instrucciones legibles por máquina, que pueden estar incorporadas en un programa o software. Las instrucciones pueden estar almacenadas en una ubicación de memoria, como la memoria 210. Ejemplos de operaciones realizadas por la CPU 205 pueden incluir búsqueda, decodificación, ejecución y reescritura.

[0230] La unidad de almacenamiento 215 puede almacenar archivos, tales como controladores, bibliotecas y programas guardados. La unidad de almacenamiento 215 puede almacenar programas generados por usuarios y sesiones grabadas, así como salidas asociadas con los programas. La unidad de almacenamiento 215 puede almacenar datos de usuario, por ejemplo, preferencias de usuario y programas de usuario. El sistema informático 201 en algunos casos puede incluir una o más unidades de almacenamiento de datos adicionales que son externas al sistema informático 201, tales como ubicadas en un servidor remoto que está en comunicación con el sistema informático 201 a través de una intranet o Internet. Los datos pueden transferirse de una ubicación a otra utilizando, por ejemplo, una red de comunicación o transferencia física de datos (por ejemplo, utilizando un disco duro, una memoria USB u otro mecanismo de almacenamiento de datos).

[0231] El sistema informático 201 puede comunicarse con uno o más sistemas informáticos remotos a través de la red 230. Para la forma de realización, el sistema informático 201 puede comunicarse con un sistema informático remoto de un usuario (por ejemplo, un operador). Los ejemplos de sistemas informáticos remotos incluyen ordenadores personales

(por ejemplo, PC portátiles), tabletas o tabletas (por ejemplo, iPad de Apple®, Galaxy Tab de Samsung®), teléfonos, teléfonos inteligentes (por ejemplo, iPhone de Apple®, dispositivo habilitado con Android, Blackberry®) o asistentes digitales personales. El usuario puede acceder al sistema informático 201 a través de la red 230.

5 **[0232]** Los métodos como los descritos en este documento pueden implementarse por medio de un código ejecutable por máquina (por ejemplo, un procesador de ordenador) almacenado en una ubicación de almacenamiento electrónico del sistema informático 201, como, por ejemplo, en la memoria 210 o la unidad de almacenamiento electrónico 215. El código ejecutable por máquina o legible por máquina puede proporcionarse en forma de software. Durante el uso, el código puede ser ejecutado por el procesador 205. En algunos casos, el código puede recuperarse de la unidad de almacenamiento 215 y almacenarse en la memoria 210 para un acceso rápido por parte del procesador 205. En algunas situaciones, la unidad de almacenamiento electrónico 215 puede excluirse, y las instrucciones ejecutables por máquina se almacenan en la memoria 210.

15 **[0233]** En un aspecto, la presente divulgación proporciona un medio legible por computadora no transitorio que comprende instrucciones ejecutables por computadora que, cuando son ejecutadas por al menos un procesador electrónico, realizan al menos una parte de un método que comprende: dividir el ADN de la muestra en una pluralidad de particiones, comprendiendo la pluralidad al menos una partición hipermetilada y una partición hipometilada; etiquetar diferencialmente el ADN de la partición hipermetilada y el ADN de la partición hipometilada; preparar un primer grupo que comprende al menos una parte del ADN de la partición hipometilada; preparar un segundo grupo que comprende al menos una parte del ADN de la partición hipermetilada; capturar al menos un primer conjunto de regiones objetivo del primer grupo, en donde el primer conjunto comprende regiones objetivo de secuencia variable; capturar una segunda pluralidad de conjuntos de regiones objetivo del segundo grupo, en donde la segunda pluralidad comprende regiones objetivo de secuencia variable y regiones objetivo epigenéticas; secuenciar las moléculas de ADN capturadas, en donde las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable se secuencian a una mayor profundidad de secuenciación que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas; obtener una pluralidad de lecturas de secuencia generadas por un secuenciador de ácidos nucleicos a partir de la secuenciación de las moléculas de ADN capturadas; mapear la pluralidad de lecturas de secuencia a una o más secuencias de referencia para generar lecturas de secuencia mapeadas; y procesar las lecturas de secuencia mapeadas correspondientes al conjunto de regiones objetivo de secuencia variable y al conjunto de regiones objetivo epigenéticas para determinar la probabilidad de que el sujeto tenga cáncer.

25 **[0234]** El código puede ser precompilado y configurado para su uso con una máquina que tenga un procesador adaptado para ejecutar el código o puede ser compilado durante el tiempo de ejecución. El código puede ser suministrado en un lenguaje de programación que puede seleccionarse para permitir que el código se ejecute de manera precompilada o compilada.

35 **[0235]** Los aspectos de los sistemas y métodos proporcionados en este documento, como el sistema informático 201, pueden incorporarse en la programación. Varios aspectos de la tecnología pueden considerarse como "productos" o "artículos de fabricación" típicamente en forma de código ejecutable por máquina (o procesador) y/o datos asociados que se transportan o incorporan en un tipo de medio legible por máquina. El código ejecutable por máquina puede almacenarse en una unidad de almacenamiento electrónico, tal memoria (por ejemplo, memoria de solo lectura, memoria de acceso aleatorio, memoria flash) o un disco duro. Los medios de tipo "almacenamiento" pueden incluir cualquiera o todas las memorias tangibles de los ordenadores, procesadores o similares, o módulos asociados a los mismos, como varias memorias de semiconductores, unidades de cinta, unidades de disco y similares, que pueden proporcionar almacenamiento no transitorio en cualquier momento para la programación del software.

45 **[0236]** Todo o parte del software puede en ocasiones comunicarse a través de Internet o de otras redes de telecomunicaciones. Dichas comunicaciones, por ejemplo, pueden permitir la carga del software desde una computadora o procesador a otro, por ejemplo, desde un servidor de administración o computadora host a la plataforma de computadora de un servidor de aplicaciones. Por lo tanto, otro tipo de medio que puede soportar los elementos del software incluye ondas ópticas, eléctricas y electromagnéticas, como las que se utilizan en interfaces físicas entre dispositivos locales, a través de redes terrestres ópticas y cableadas, y en varios enlaces aéreos. Los elementos físicos que transportan dichas ondas, como enlaces cableados o inalámbricos, enlaces ópticos o similares, también pueden considerarse como medios que soportan el software. Tal como se utilizan en el presente documento, a menos que se restrinjan a medios de "almacenamiento" tangibles no transitorios, los términos como "medio legible por computadora" o "medio legible por máquina" se refieren a cualquier medio que participe en el suministro de instrucciones a un procesador para su ejecución.

50 **[0237]** Por lo tanto, un medio legible por máquina, tal como un código ejecutable por computadora puede tomar muchas formas, incluyendo, pero sin limitarse a, un medio de almacenamiento tangible, un medio de onda portadora o un medio de transmisión física. Los medios de almacenamiento no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, tales como cualquiera de los dispositivos de almacenamiento en cualquier computadora o similar, tales como los que se pueden usar para implementar las bases de datos, etc., que se muestran en los dibujos. Los medios de almacenamiento volátiles incluyen memoria dinámica, tal como la memoria principal de dicha plataforma informática. Los medios de transmisión tangibles incluyen cables coaxiales, alambre de cobre y fibra óptica, incluyendo los cables que comprenden un bus dentro de un sistema informático. Los medios de transmisión de onda portadora pueden tomar la forma de señales eléctricas o electromagnéticas, u ondas acústicas o de luz tales como las generadas durante las comunicaciones de datos

por radiofrecuencia (RF) e infrarrojos (IR). Por lo tanto, las formas comunes de medios legibles por computadora incluyen, por ejemplo: un disquete, un disco flexible, un disco duro, una cinta magnética, cualquier otro medio magnético, un CD-ROM, un DVD o un DVD-ROM, cualquier otro medio óptico, tarjetas perforadas, cinta de papel, cualquier otro medio de almacenamiento físico con patrones de agujeros, una RAM, una ROM, una PROM y una EPROM, una FLASH EPROM, cualquier otro chip o cartucho de memoria, una onda portadora que transporte datos o instrucciones, cables o enlaces que transporten dicha onda portadora, o cualquier otro medio desde el cual una computadora pueda leer código de programación y/o datos. Muchas de estas formas de medios legibles por computadora pueden estar involucradas en llevar una o más secuencias de una o más instrucciones a un procesador para su ejecución.

[0238] El sistema informático 201 puede incluir o estar en comunicación con una pantalla electrónica que comprende una interfaz de usuario (UI) para proporcionar, por ejemplo, uno o más resultados de análisis de muestras. Los ejemplos de UI incluyen, sin limitación, una interfaz gráfica de usuario (GUI) y una interfaz de usuario basada en web.

[0239] También se proporcionan detalles adicionales relacionados con sistemas y redes informáticas, bases de datos y productos de programas informáticos en, por ejemplo, Peterson, Computer Networks: A Systems Approach, Morgan Kaufmann, 5.^a edición (2011), Kurose, Computer Networking: A Top-Down Approach, Pearson, 7.^a edición (2016), Elmasri, Fundamentals of Database Systems, Addison Wesley, 6.^a edición (2010), Coronel, Database Systems: Design, Implementation, & Management, Cengage Learning, 11.^a edición (2014), Tucker, Programming Languages, McGraw-Hill Science/Engineering/Math, 2.^a edición (2006), y Rhoton, Cloud Computing Architected: Solution Design Handbook, Recursive Press (2011).

V. APLICACIONES

1. Cáncer y otras enfermedades

[0240] Los métodos presentes se pueden utilizar para diagnosticar la presencia o ausencia de afecciones, particularmente cáncer, en un sujeto, para caracterizar afecciones (por ejemplo, estadificación del cáncer o determinación de la heterogeneidad de un cáncer), monitorear la respuesta al tratamiento de una afección, determinar el riesgo de pronóstico de desarrollar una afección o el curso posterior de una afección. En algunas formas de realización, la presencia o ausencia de un cáncer se determina en un sujeto basándose al menos en parte en datos generados mediante la secuenciación de la primera y segunda pluralidades de conjuntos de regiones objetivo. La presente divulgación también puede ser útil para determinar la eficacia de una opción de tratamiento particular. Las opciones de tratamiento exitosas pueden aumentar la cantidad de variación del número de copias o mutaciones raras detectadas en la sangre del sujeto si el tratamiento es exitoso ya que más cánceres pueden morir y desprender ADN. En otros ejemplos, esto puede no ocurrir. En otro ejemplo, tal vez ciertas opciones de tratamiento se pueden correlacionar con perfiles genéticos de cánceres a lo largo del tiempo. Esta correlación puede ser útil para seleccionar una terapia.

[0241] Además, si se observa que un cáncer está en remisión después del tratamiento, los métodos actuales se pueden utilizar para monitorear la enfermedad residual o la recurrencia de la enfermedad.

[0242] En algunas formas de realización, los métodos y sistemas divulgados en el presente documento pueden utilizarse para identificar terapias personalizadas o dirigidas para tratar una enfermedad o afección determinada en pacientes basándose en la clasificación de una variante de ácido nucleico como de origen somático o de línea germinal. Normalmente, la enfermedad en consideración es un tipo de cáncer. Los ejemplos no limitantes de dichos cánceres incluyen cáncer de las vías biliares, cáncer de vejiga, carcinoma de células transicionales, carcinoma urotelial, cáncer cerebral, gliomas, astrocitomas, carcinoma de mama, carcinoma metaplásico, cáncer de cuello uterino, carcinoma de células escamosas cervical, cáncer rectal, carcinoma colorrectal, cáncer de colon, cáncer colorrectal hereditario sin poliposis, adenocarcinomas colorrectales, tumores del estroma gastrointestinal (GIST), carcinoma endometrial, sarcomas del estroma endometrial, cáncer de esófago, carcinoma de células escamosas de esófago, adenocarcinoma de esófago, melanoma ocular, melanoma uveal, carcinomas de vesícula biliar, adenocarcinoma de vesícula biliar, carcinoma de células renales, carcinoma de células renales de células claras, carcinoma de células transicionales, carcinomas uroteliales, tumor de Wilms, leucemia, leucemia linfocítica aguda (LLA), leucemia mieloide aguda (LMA), leucemia linfocítica crónica (LLC), leucemia leucemia mieloide crónica (LMC), leucemia mielomonocítica crónica (LMMC), cáncer de hígado, carcinoma de hígado, hepatoma, carcinoma hepatocelular, colangiocarcinoma, hepatoblastoma, cáncer de pulmón, cáncer de pulmón de células no pequeñas (CPCNP), mesotelioma, linfomas de células B, linfoma no Hodgkin, linfoma difuso de células B grandes, linfoma de células del manto, linfomas de células T, linfoma no Hodgkin, linfoma/leucemia linfoblástica de precursores T, linfomas de células T periféricas, mieloma múltiple, carcinoma nasofaríngeo (CNP), neuroblastoma, cáncer orofaríngeo, carcinomas de células escamosas de la cavidad oral, osteosarcoma, carcinoma de ovario, cáncer de páncreas, adenocarcinoma ductal pancreático, neoplasias pseudopapilares, carcinomas de células acinares. Cáncer de próstata, adenocarcinoma de próstata, cáncer de piel, melanoma, melanoma maligno, melanoma cutáneo, carcinomas de intestino delgado, cáncer de estómago, carcinoma gástrico, tumor del estroma gastrointestinal (GIST), cáncer de útero o sarcoma uterino. El tipo y/o estadio del cáncer se pueden detectar a partir de variaciones genéticas, incluidas mutaciones, mutaciones raras, indeles, variaciones en el número de copias, transversiones, translocaciones, inversión, delecciones, aneuploidía, aneuploidía parcial, poliploidía, inestabilidad cromosómica, alteraciones de la estructura cromosómica, fusiones genéticas, fusiones cromosómicas, truncamientos genéticos, amplificación genética, duplicaciones genéticas, lesiones cromosómicas, lesiones del ADN,

cambios anormales en las modificaciones químicas de los ácidos nucleicos, cambios anormales en los patrones epigenéticos y cambios anormales en la 5-metilcitosina del ácido nucleico.

[0243] Los datos genéticos también se pueden utilizar para caracterizar una forma específica de cáncer. Los cánceres suelen ser heterogéneos tanto en composición como en estadificación. Los datos del perfil genético pueden permitir la caracterización de subtipos específicos de cáncer que pueden ser importantes en el diagnóstico o tratamiento de ese subtipo específico. Esta información también puede proporcionar a un sujeto o médico pistas sobre el pronóstico de un tipo específico de cáncer y permitir que un sujeto o médico adapte las opciones de tratamiento de acuerdo con el progreso de la enfermedad. Algunos cánceres pueden progresar hasta volverse más agresivos y genéticamente inestables. Otros cánceres pueden permanecer benignos, inactivos o latentes. El sistema y los métodos de esta divulgación pueden ser útiles para determinar la progresión de la enfermedad.

[0244] Además, los métodos de la divulgación se pueden utilizar para caracterizar la heterogeneidad de una condición anormal en un sujeto. Dichos métodos pueden incluir, por ejemplo, generar un perfil genético de polinucleótidos extracelulares derivados del sujeto, en donde el perfil genético comprende una pluralidad de datos resultantes de la variación del número de copias y análisis de mutaciones raras. En algunas formas de realización, una condición anormal es cáncer. En algunas formas de realización, la condición anormal puede ser una que resulte en una población genómica heterogénea. En el ejemplo del cáncer, se sabe que algunos tumores comprenden células tumorales en diferentes etapas del cáncer. En otros ejemplos, la heterogeneidad puede comprender múltiples focos de enfermedad. Nuevamente, en el ejemplo del cáncer, puede haber múltiples focos tumorales, quizás donde uno o más focos son el resultado de metástasis que se han propagado desde un sitio primario.

[0245] Los presentes métodos pueden utilizarse para generar un perfil, huella o un conjunto de datos que sea una suma de información genética derivada de diferentes células en una enfermedad heterogénea. Este conjunto de datos puede comprender variación del número de copias, variación epigenética y análisis de mutaciones solos o en combinación.

[0246] Los métodos presentes se pueden utilizar para diagnosticar, pronosticar, controlar u observar cánceres u otras enfermedades. En algunas formas de realización, los métodos descritos en el presente documento no implican el diagnóstico, pronóstico o control de un feto y, como tales, no están dirigidos a pruebas prenatales no invasivas. En otras formas de realización, estas metodologías se pueden emplear en una embarazada para diagnosticar, pronosticar, controlar u observar cánceres u otras enfermedades en un feto cuyo ADN y otros polinucleótidos pueden co-circular con moléculas maternas.

[0247] Ejemplos no limitantes de otras enfermedades, trastornos o afecciones de base genética que se evalúan opcionalmente utilizando los métodos y sistemas divulgados en el presente documento incluyen acondroplasia, deficiencia de alfa-1 antitripsina, síndrome antifosfolípido, autismo, enfermedad renal poliquística autosómica dominante, Charcot-Marie-Tooth (CMT), cri du chat, enfermedad de Crohn, fibrosis quística, enfermedad de Dercum, síndrome de Down, síndrome de Duane, distrofia muscular de Duchenne, trombofilia del factor V Leiden, hipercolesterolemia familiar, fiebre mediterránea familiar, síndrome del cromosoma X frágil, enfermedad de Gaucher, hemocromatosis, hemofilia, holoprosencefalia, enfermedad de Huntington, síndrome de Klinefelter, síndrome de Marfan, distrofia miotónica, neurofibromatosis, síndrome de Noonan, osteogénesis imperfecta, enfermedad de Parkinson, fenilcetonuria, anomalía de Poland, porfiria, progeria, retinitis pigmentosa, inmunodeficiencia combinada grave (SCID), anemia falciforme, enfermedad de las células madre, atrofia muscular espinal, enfermedad de Tay-Sachs, talasemia, trimetilaminuria, síndrome de Turner, síndrome velocardiofacial, síndrome de WAGR, enfermedad de Wilson o similares.

[0248] En algunas formas de realización, un método descrito en el presente documento comprende detectar la presencia o ausencia de ADN originado o derivado de una célula tumoral en un punto temporal preseleccionado después de un tratamiento previo contra el cáncer de un sujeto previamente diagnosticado con cáncer utilizando un conjunto de información de secuencia obtenida como se describe en el presente documento. El método puede comprender además determinar una puntuación de recurrencia del cáncer que sea indicativa de la presencia o ausencia del ADN originado o derivado de la célula tumoral para el sujeto de prueba.

[0249] Cuando se determina una puntuación de recurrencia del cáncer, puede utilizarse además para determinar un estado de recurrencia del cáncer. El estado de recurrencia del cáncer puede ser de riesgo de recurrencia del cáncer, por ejemplo, cuando la puntuación de recurrencia del cáncer está por encima o por debajo de un umbral predeterminado. El estado de recurrencia del cáncer puede ser de riesgo bajo o menor de recurrencia del cáncer, por ejemplo, cuando la puntuación de recurrencia del cáncer está por encima o por debajo de un umbral predeterminado. En formas de realización particulares, una puntuación de recurrencia del cáncer igual al umbral predeterminado puede dar como resultado un estado de recurrencia del cáncer de riesgo de recurrencia del cáncer o de riesgo bajo o menor de recurrencia del cáncer.

[0250] En algunas formas de realización, se compara una puntuación de recurrencia del cáncer con un umbral de recurrencia del cáncer predeterminado, y el sujeto de prueba se clasifica como candidato para un tratamiento posterior del cáncer cuando la puntuación de recurrencia del cáncer está por encima del umbral de recurrencia del cáncer o no es candidato para terapia cuando la puntuación de recurrencia del cáncer está por debajo del umbral de recurrencia del cáncer. En algunas formas de realización, se compara una puntuación de recurrencia del cáncer con un umbral de recurrencia del cáncer predeterminado, y el sujeto de prueba se clasifica como candidato para un tratamiento posterior

del cáncer cuando la puntuación de recurrencia del cáncer cumple con el umbral de recurrencia del cáncer o no es candidato para terapia cuando la puntuación de recurrencia del cáncer no cumple con el umbral de recurrencia del cáncer. En formas de realización particulares, una puntuación de recurrencia del cáncer igual al umbral de recurrencia del cáncer puede dar como resultado la clasificación como candidato para un tratamiento posterior del cáncer o no candidato para terapia.

[0251] Los métodos analizados anteriormente pueden comprender además cualquier característica o características compatibles establecidas en otra parte del presente documento, incluida la sección relativa a los métodos para determinar un riesgo de recurrencia del cáncer en un sujeto de prueba y/o clasificar a un sujeto de prueba como candidato para un tratamiento posterior contra el cáncer.

2. Métodos para determinar un riesgo de recurrencia del cáncer en un sujeto de prueba y/o clasificar a un sujeto de prueba como candidato para un tratamiento posterior contra el cáncer

[0252] En algunas formas de realización, un método proporcionado en el presente documento es un método para determinar un riesgo de recurrencia del cáncer en un sujeto de prueba. En algunas formas de realización, un método proporcionado en el presente documento es un método para clasificar a un sujeto de prueba como candidato para un tratamiento posterior contra el cáncer.

[0253] Cualquiera de dichos métodos puede comprender la recolección de ADN (por ejemplo, originado o derivado de una célula tumoral) del sujeto de prueba diagnosticado con el cáncer en uno o más puntos temporales preseleccionados después de uno o más tratamientos previos contra el cáncer al sujeto de prueba. El sujeto puede ser cualquiera de los sujetos descritos en este documento. El ADN puede ser ADNlc. El ADN puede obtenerse de una muestra de tejido.

[0254] Cualquiera de dichos métodos puede comprender la captura de una pluralidad de conjuntos de regiones objetivo a partir del ADN del sujeto, en donde la pluralidad de conjuntos de regiones objetivo comprende un conjunto de regiones objetivo de secuencia variable y un conjunto de regiones objetivo epigenéticas, con lo que se produce un conjunto capturado de moléculas de ADN. La etapa de captura puede realizarse de acuerdo con cualquiera de las formas de realización descritas en otra parte del presente documento.

[0255] En cualquiera de dichos métodos, el tratamiento previo del cáncer puede comprender cirugía, administración de una composición terapéutica y/o quimioterapia.

[0256] Cualquiera de dichos métodos puede comprender la secuenciación de las moléculas de ADN capturadas, con lo que se produce un conjunto de información de secuencia. Las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable pueden secuenciarse a una mayor profundidad de secuenciación que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas.

[0257] Cualquiera de dichos métodos puede comprender la detección de la presencia o ausencia de ADN originado o derivado de una célula tumoral en un punto temporal preseleccionado utilizando el conjunto de información de secuencia. La detección de la presencia o ausencia de ADN originado o derivado de una célula tumoral puede realizarse de acuerdo con cualquiera de las formas de realización descritas en otra parte del presente documento.

[0258] Los métodos para determinar un riesgo de recurrencia de cáncer en un sujeto de prueba pueden comprender la determinación de una puntuación de recurrencia de cáncer que sea indicativa de la presencia o ausencia, o cantidad, del ADN originado o derivado de la célula tumoral para el sujeto de prueba. La puntuación de recurrencia de cáncer puede utilizarse además para determinar un estado de recurrencia de cáncer. El estado de recurrencia de cáncer puede ser de riesgo de recurrencia de cáncer, por ejemplo, cuando la puntuación de recurrencia de cáncer está por encima o por debajo de un umbral predeterminado. El estado de recurrencia de cáncer puede ser de riesgo bajo o menor de recurrencia de cáncer, por ejemplo, cuando la puntuación de recurrencia de cáncer está por debajo o por encima de un umbral predeterminado. En formas de realización particulares, una puntuación de recurrencia de cáncer igual al umbral predeterminado puede dar como resultado un estado de recurrencia de cáncer de riesgo de recurrencia de cáncer o de riesgo bajo o menor de recurrencia de cáncer.

[0259] Los métodos para clasificar a un sujeto de prueba como candidato para un tratamiento posterior contra el cáncer pueden comprender comparar la puntuación de recurrencia del cáncer del sujeto de prueba con un umbral de recurrencia del cáncer predeterminado, clasificando de ese modo al sujeto de prueba como candidato para el tratamiento posterior contra el cáncer cuando la puntuación de recurrencia del cáncer está por encima del umbral de recurrencia del cáncer o como no candidato para la terapia cuando la puntuación de recurrencia del cáncer está por debajo del umbral de recurrencia del cáncer. En formas de realización particulares, una puntuación de recurrencia del cáncer igual al umbral de recurrencia del cáncer puede dar como resultado la clasificación como candidato para un tratamiento posterior contra el cáncer o como no candidato para la terapia. En algunas formas de realización, el tratamiento posterior contra el cáncer comprende quimioterapia o la administración de una composición terapéutica.

[0260] Cualquiera de dichos métodos puede comprender la determinación de un período de supervivencia libre de enfermedad (SLE) para el sujeto de prueba basándose en la puntuación de recurrencia del cáncer; por ejemplo, el período de SLE puede ser de 1 año, 2 años, 3 años, 4 años, 5 años o 10 años.

- 5 **[0261]** En algunas formas de realización, el conjunto de información de secuencia comprende secuencias de región objetivo de secuencia variable, y la determinación del puntaje de recurrencia del cáncer puede comprender la determinación de al menos un primer subpuntaje indicativo de la cantidad de SNV, inserciones/deleciones, CNV y/o fusiones presentes en secuencias de región objetivo de secuencia variable.
- 10 **[0262]** En algunas formas de realización, una cantidad de mutaciones en las regiones objetivo de secuencia variable elegidas entre 1, 2, 3, 4 o 5 es suficiente para que la primera subpuntuación dé como resultado una puntuación de recurrencia del cáncer clasificada como positiva para la recurrencia del cáncer. En algunas formas de realización, la cantidad de mutaciones se elige entre 1, 2 o 3.
- 15 **[0263]** En algunas formas de realización, el conjunto de información de secuencia comprende secuencias de la región objetivo epigenética, y la determinación de la puntuación de recurrencia del cáncer comprende la determinación de una segunda puntuación parcial indicativa de la cantidad de lecturas de secuencia anormales en las secuencias de la región objetivo epigenética. Las lecturas de secuencia anormales pueden ser lecturas indicativas de un estado epigenético diferente del ADN encontrado en una muestra correspondiente de un sujeto sano (por ejemplo, ADNcf encontrado en una muestra de sangre de un sujeto sano, o ADN encontrado en una muestra de tejido de un sujeto sano donde la muestra de tejido es del mismo tipo de tejido que el obtenido del sujeto de prueba). Las lecturas anormales pueden ser consistentes con cambios epigenéticos asociados con el cáncer, por ejemplo, metilación de regiones objetivo variables de hipermetilación y/o fragmentación perturbada de regiones objetivo variables de fragmentación, donde "perturbado" significa diferente del ADN encontrado en una muestra correspondiente de un sujeto sano.
- 20 **[0264]** En algunas formas de realización, una proporción de lecturas correspondientes al conjunto de regiones objetivo variables de hipermetilación y/o al conjunto de regiones objetivo variables de fragmentación que indican hipermetilación en el conjunto de regiones objetivo variables de hipermetilación y/o fragmentación anormal en el conjunto de regiones objetivo variables de fragmentación mayor o igual a un valor en el intervalo de 0,001 %-100 % es suficiente para que la segunda subpuntuación se clasifique como positiva para la recurrencia del cáncer. El intervalo puede ser de 0,001 %-10 %, 0,001 %-1 %, 0,005 %-1 %, 0,01 %-5 %, 0,01 %-2 % o 0,01 %-1 %. Alternativamente, el intervalo puede ser de 10 % a 100 %.
- 25 **[0265]** En algunas formas de realización, cualquiera de dichos métodos puede comprender la determinación de una fracción de ADN tumoral a partir de la fracción de lecturas en el conjunto de información de secuencia que indican una o más características indicativas de origen a partir de una célula tumoral. Esto se puede hacer para lecturas correspondientes a algunas o todas las regiones objetivo epigenéticas, por ejemplo, incluyendo una o ambas de las regiones objetivo variables de hipermetilación y regiones objetivo variables de fragmentación (la hipermetilación de una región objetivo variable de hipermetilación y/o la fragmentación anormal de una región objetivo variable de fragmentación se pueden considerar indicativas de origen a partir de una célula tumoral). Esto se puede hacer para lecturas correspondientes a regiones objetivo variables de secuencia, por ejemplo, lecturas que comprenden alteraciones consistentes con el cáncer, tales como SNV, indeles, CNV y/o fusiones. La fracción de ADN tumoral se puede determinar basándose en una combinación de lecturas correspondientes a regiones objetivo epigenéticas y lecturas correspondientes a regiones objetivo variables de secuencia.
- 30 **[0266]** La determinación de una puntuación de recurrencia del cáncer puede basarse al menos en parte en la fracción de ADN tumoral, en donde una fracción de ADN tumoral mayor que un umbral en el rango de 10^{-11} a 1 o 10^{-10} a 1 es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer. En algunas formas de realización, una fracción de ADN tumoral mayor o igual que un umbral en el rango de 10^{-10} a 10^{-9} , 10^{-9} a 10^{-8} , 10^{-8} a 10^{-7} , 10^{-7} a 10^{-6} , 10^{-6} a 10^{-5} , 10^{-5} a 10^{-4} , 10^{-4} a 10^{-3} , 10^{-3} a 10^{-2} , o 10^{-2} a 10^{-1} es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer. En algunas formas de realización, la fracción de ADN tumoral mayor que un umbral de al menos 10^{-7} es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer. Se puede realizar una determinación de que una fracción de ADN tumoral es mayor que un umbral, como un umbral correspondiente a cualquiera de las formas de realización anteriores, en función de una probabilidad acumulativa. Por ejemplo, la muestra se consideró positiva si la probabilidad acumulativa de que la fracción tumoral fuera mayor que un umbral en cualquiera de los rangos anteriores supera un umbral de probabilidad de al menos 0,5, 0,75, 0,9, 0,95, 0,98, 0,99, 0,995 o 0,999. En algunas formas de realización, el umbral de probabilidad es al menos 0,95, como 0,99.
- 35 **[0267]** En algunas formas de realización, el conjunto de información de secuencia comprende secuencias de regiones objetivo variables de secuencia y secuencias de regiones objetivo epigenéticas, y determinar la puntuación de recurrencia del cáncer comprende determinar una primera subpuntuación indicativa de la cantidad de SNV, inserciones/deleciones, CNV y/o fusiones presentes en secuencias de regiones objetivo variables de secuencia y una segunda subpuntuación indicativa de la cantidad de lecturas de secuencia anormales en secuencias de regiones objetivo epigenéticas, y combinar la primera y la segunda subpuntuaciones para proporcionar la puntuación de recurrencia del cáncer. Cuando se combinan la primera y la segunda subpuntuaciones, se pueden combinar aplicando un umbral a cada subpuntuación de forma
- 40
- 45
- 50
- 55
- 60
- 65

independiente (por ejemplo, mayor que un número predeterminado de mutaciones (por ejemplo, > 1) en regiones objetivo variables de secuencia, y mayor que una fracción predeterminada de lecturas anormales (por ejemplo, tumor) en regiones objetivo epigenéticas), o entrenar un clasificador de aprendizaje automático para determinar el estado en función de una pluralidad de muestras de entrenamiento positivas y negativas.

[0268] En algunas formas de realización, un valor para la puntuación combinada en el rango de -4 a 2 o de -3 a 1 es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer.

[0269] En cualquier realización donde una puntuación de recurrencia de cáncer se clasifica como positiva para la recurrencia de cáncer, el estado de recurrencia de cáncer del sujeto puede estar en riesgo de recurrencia de cáncer y/o el sujeto puede clasificarse como candidato para un tratamiento de cáncer posterior.

[0270] En algunas formas de realización, el cáncer es cualquiera de los tipos de cáncer descritos en otras partes de este documento, por ejemplo, cáncer colorrectal.

3. Terapias y administración relacionada

[0271] En ciertas formas de realización, los métodos divulgados en este documento se relacionan con la identificación y administración de terapias personalizadas a pacientes dado el estado de una variante de ácido nucleico como de origen somático o de línea germinal. En algunas formas de realización, esencialmente cualquier terapia contra el cáncer (por ejemplo, terapia quirúrgica, radioterapia, quimioterapia y/o similares) puede incluirse como parte de estos métodos. En algunas formas de realización, las terapias personalizadas incluyen al menos una inmunoterapia (o un agente inmunoterapéutico). La inmunoterapia se refiere en general a métodos para mejorar una respuesta inmunitaria contra un tipo de cáncer determinado. En ciertas formas de realización, la inmunoterapia se refiere a métodos para mejorar una respuesta de células T contra un tumor o cáncer.

[0272] En ciertas formas de realización, el estado de una variante de ácido nucleico de una muestra de un sujeto como de origen somático o de línea germinal se puede comparar con una base de datos de resultados comparativos de una población de referencia para identificar terapias personalizadas o dirigidas para ese sujeto. En algunas formas de realización, la población de referencia incluye pacientes con el mismo tipo de cáncer o enfermedad que el sujeto de prueba y/o pacientes que están recibiendo, o que han recibido, la misma terapia que el sujeto de prueba. Una terapia (o terapias) personalizada o dirigida se puede identificar cuando la variante de ácido nucleico y los resultados comparativos satisfacen ciertos criterios de clasificación (por ejemplo, son una coincidencia sustancial o aproximada).

[0273] En ciertas formas de realización, las terapias personalizadas descritas en el presente documento se administran típicamente por vía parenteral (por ejemplo, por vía intravenosa o subcutánea). Las composiciones farmacéuticas que contienen un agente inmunoterapéutico se administran típicamente por vía intravenosa. Ciertos agentes terapéuticos se administran por vía oral. Sin embargo, las terapias personalizadas (por ejemplo, agentes inmunoterapéuticos, etc.) también se pueden administrar por métodos tales como, por ejemplo, bucal, sublingual, rectal, vaginal, intrauretral, tópico, intraocular, intranasal y/o intraauricular, cuya administración puede incluir comprimidos, cápsulas, gránulos, suspensiones acuosas, geles, aerosoles, supositorios, ungüentos, pomadas o similares.

VI. EJEMPLOS

1. Detección de cáncer mediante la detección de reordenamientos genómicos y uno o ambos conjuntos de regiones objetivo de secuencia variable y un conjunto de regiones objetivo epigenéticas

[0274] El ADN (por ejemplo, ADNlc) se procesa particionándolo en al menos particiones hipermetiladas e hipometiladas a través de la preparación de la biblioteca estándar y PCR, y el ADN marcado resultante se divide en una primera y una segunda alícuotas. La preparación de la biblioteca estándar (por ejemplo, reparación de extremos, ligadura de adaptadores en ambos extremos) y la amplificación a través de PCR se realizan en cada partición. Se prepara un primer grupo que comprende aproximadamente el 80 % de la partición hipometilada y se prepara un segundo grupo que comprende la partición hipermetilada y aproximadamente el 20 % de la partición hipometilada. Las regiones objetivo de secuencia variable se capturan del primer grupo. Las regiones objetivo de secuencia variable y las regiones objetivo epigenéticas se capturan del segundo grupo. La captura se realiza poniendo en contacto el ADN del grupo con sondas de región objetivo para uno o ambos conjuntos de regiones objetivo de secuencia variable y el conjunto de regiones objetivo epigenéticas.

[0275] Las sondas para el conjunto de regiones objetivo variables de secuencia pueden tener una huella de aproximadamente 50 kb, mientras que las sondas para el conjunto de regiones objetivo epigenéticas pueden tener una huella de región objetivo de aproximadamente 500 kb. Las sondas para el conjunto de regiones objetivo variables de secuencia pueden comprender oligonucleótidos que se dirigen a una selección de regiones identificadas en las Tablas 3-5 y las sondas para el conjunto de regiones objetivo epigenéticas pueden comprender oligonucleótidos que se dirigen a una selección de regiones objetivo variables de hipermetilación, regiones objetivo variables de hipometilación, regiones objetivo de unión a CTCF, regiones objetivo del sitio de inicio de la transcripción, regiones objetivo de amplificación focal y regiones de control de metilación.

[0276] El ADNlc capturado aislado de esta manera puede luego secuenciarse utilizando un secuenciador Illumina HiSeq o NovaSeq, y se analizarán los resultados. Las secuencias de la región objetivo de secuencia variable se analizan detectando alteraciones genómicas tales como SNV, inserciones, deleciones y fusiones que se pueden invocar con suficiente respaldo para discriminar variantes tumorales reales de errores técnicos. Las secuencias de la región objetivo epigenética se analizan de forma independiente para detectar fragmentos metilados en regiones que han demostrado estar metiladas de forma diferencial en el cáncer en comparación con las células sanguíneas.

[0277] Los resultados de la secuenciación de las regiones objetivo de secuencia variable y las regiones objetivo epigenéticas se analizan y combinan para producir una determinación final de tumor presente/ausente para determinar si los resultados son consistentes con el cáncer, por ejemplo, con una especificidad del 95 %.

2. Caracterización de los datos de secuenciación obtenidos a través de un método divulgado a modo de ejemplo

[0278] Para preparar el ADNlc para la secuenciación, el ADNlc extraído de plasma de donante sano o de CRC se dividió. Una partición hipometilada se eluyó por separado. Una partición intermedia (también denominada residual) se eluyó, y ese eluato se utilizó posteriormente para eluir la partición hipermetilada. Las particiones hipometiladas y residuales/hipermetiladas se ligaron con etiquetas moleculares específicas de partición y se limpiaron. La partición hipometilada se dividió de manera que el 80 % de la misma se mantuvo separada como un primer grupo. El 20 % restante de la partición hipometilada se combinó con las particiones residuales e hipermetiladas, dando un segundo grupo. Paralelamente, las mismas muestras de ADNlc se procesaron en tres lotes separados sin dividirlos y se combinaron en un solo grupo después de la partición y el etiquetado.

[0279] Los grupos primero y segundo de las muestras divididas y las muestras de control no divididas se sometieron a PCR de preparación de biblioteca, limpieza e hibridación para capturar el primer y segundo conjunto de regiones objetivo del primer y segundo grupo, respectivamente. Las sondas para regiones objetivo de secuencia variable se hibridaron con el primer grupo. Por separado, las sondas para regiones objetivo de secuencia variable y regiones objetivo epigenéticas se hibridaron con el segundo grupo en un pocillo de ensayo diferente del primer grupo. Las sondas para regiones objetivo de secuencia variable y regiones objetivo epigenéticas se hibridaron con las muestras de control no divididas. Después de la hibridación, los grupos se combinaron en el mismo pocillo de ensayo por muestra y se añadieron perlas revestidas con estreptavidina. Se realizaron lavados de enriquecimiento, PCR y limpieza. Las muestras se agruparon en proporciones equimolares para la secuenciación en una célula de flujo NovaSeq S4.

[0280] Para cada muestra, después de la secuenciación, las lecturas de secuenciación se alinearon con el genoma humano. Antes de la determinación de la presencia o ausencia de ADN derivado de células tumorales en función de la metilación ("llamado epigenómico"), las muestras con una partición hipo del 20 % se ampliaron para reflejar la diversidad esperada a partir del 100 %. El aumento de escala se realizó mediante dos métodos independientes: 1) multiplicando los recuentos de moléculas de la región observados en las muestras hipo del 20 % por 4,5x ("escalado plano de 4,5x"); 2) determinando un factor de escala distinto de 4,5 en función de la relación de regiones de control metiladas y no metiladas. A continuación, se realizó el llamado epigenómico. El llamado de mutaciones somáticas se realizó de acuerdo con técnicas estándar.

[0281] Como se muestra en la Fig. 3A ("Múltiple") las muestras terminaron siendo secuenciadas a una profundidad similar a los Controles 1-3. Como se muestra en la Fig. 3B, los datos de la muestra Múltiple tuvieron una diversidad menor en el panel de la región objetivo epigenómica, como se esperaba. Como se muestra en la Fig. 3C, hubo una fuerte correlación en la fracción estimada de ADN originado a partir de células tumorales ("theta exp") entre los datos estándar ("100 % hipo sin escalar") de los controles y los datos de la muestra dividida ("20 % escalado").

[0282] El uso del flujo de trabajo para las muestras divididas consumió menos ancho de banda de secuenciación al menos porque se capturaron y secuenciaron menos moléculas de la partición hipometilada correspondiente a las regiones objetivo epigenéticas. No obstante, los resultados se correlacionaron bien con las muestras de control (Fig. 3C) a pesar de la menor diversidad. En consecuencia, los métodos divulgados pueden reducir el consumo de recursos sin comprometer sustancialmente la detección de ADNct o la estimación de la fracción tumoral.

3. Determinación de la fracción de alelo mutante

[0283] Las muestras del Ejemplo 2 de sujetos con cáncer colorrectal (CCR) se mezclaron con ADN sano para dar muestras con valores conocidos de fracción de alelo mutante (MAF) de 0,05 %, 0,3 %, 8,99 % y 19,74 %, como se indica en las Figs. 4A-B. A continuación, las muestras se dividieron, se procesaron como se describió anteriormente para el Ejemplo 2 para proporcionar muestras divididas y de control, y se secuenciaron. Los datos de secuencia se analizaron para identificar mutaciones somáticas y medir MAF.

Como se muestra en las Figs. 4A-B, la determinación de MAF y la determinación de MAF se realizaron de manera similar para las muestras divididas y de control. Las réplicas de secuenciación individuales se representan individualmente a lo largo del eje horizontal. Debe notarse que la determinación de MAF negativa de las muestras en el 0.3 que tienen el MAF esperado pero que se denominan negativas es un comportamiento esperado. Esto se consideró un artefacto de la dilución de plasma que interfiere con los MAF de las variantes de línea germinal de las muestras normales durante el filtrado de

variantes tumorales. Estos resultados confirman que los métodos descritos en este documento pueden reducir el consumo de recursos sin comprometer sustancialmente la calidad del resultado, como la detección de mutaciones y la medición de MAF.

5

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Un método para aislar ADN de una muestra, comprendiendo el método:

- 5 dividir el ADN de la muestra en una pluralidad de particiones, comprendiendo la pluralidad al menos una partición hipermetilada y una partición hipometilada;
preparar un primer grupo que comprende al menos una primera porción del ADN de la partición hipometilada;
preparar un segundo grupo que comprende al menos una primera porción del ADN de la partición hipermetilada;
capturar al menos un primer conjunto de regiones objetivo del primer grupo; y
10 capturar al menos un segundo conjunto de regiones objetivo del segundo grupo, en donde el primer conjunto de regiones objetivo y el segundo conjunto de regiones objetivo no son idénticos.

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

- 15 (i) el primer conjunto de regiones objetivo comprende un conjunto de regiones objetivo de secuencia variable;
(ii) el primer conjunto de regiones objetivo comprende un conjunto de regiones objetivo de fragmentación variable;
(iii) el primer conjunto de regiones objetivo comprende un conjunto de regiones objetivo de hipometilación variable;
(iv) el segundo conjunto de regiones objetivo comprende un conjunto de regiones objetivo epigenéticas,
opcionalmente en donde el conjunto de regiones objetivo epigenéticas comprende un conjunto de regiones objetivo
20 de hipermetilación variable;
(v) el segundo conjunto de regiones objetivo comprende un conjunto de regiones objetivo de fragmentación variable;
(vi) el segundo conjunto de regiones objetivo comprende un conjunto de regiones objetivo de secuencia variable;
(vii) se añade una etiqueta de muestra al ADN de la muestra;
25 (viii) el primer conjunto de regiones objetivo o el ADN amplificado a partir de las mismas se secuencian por separado del segundo conjunto de regiones objetivo o el ADN amplificado a partir de las mismas, opcionalmente en donde una pluralidad de primeros conjuntos de regiones objetivo de diferentes muestras se agrupan y secuencian juntos, y/o una pluralidad de segundos conjuntos de regiones objetivo de diferentes muestras se agrupan y secuencian juntos; y/o
30 (ix) el método comprende además etiquetar diferencialmente el ADN de la partición hipermetilada y el ADN de la partición hipometilada.

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

- 35 (i) el primer conjunto de regiones objetivo comprende un conjunto de regiones objetivo de secuencia variable;
(ii) el sujeto de prueba fue diagnosticado previamente con un cáncer y recibió uno o más tratamientos previos contra el cáncer; y
(iii) las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable se secuencian, generando de ese modo un conjunto de información de secuencia.

4. Un método para aislar ADN de una muestra, comprendiendo el método:

- particionar el ADN de la muestra en una pluralidad de particiones, comprendiendo la pluralidad al menos una
partición hipermetilada y una partición hipometilada;
45 preparar un primer pool que comprenda al menos una primera porción del ADN de la partición
partición hipometilada;
preparar un segundo grupo que comprenda al menos una primera porción del ADN de la partición
partición hipermetilada;
capturar al menos un primer conjunto de regiones objetivo del primer grupo, en el que el primer conjunto
50 comprende regiones objetivo de secuencia variable; y
capturar una segunda pluralidad de conjuntos de regiones objetivo del segundo conjunto, en el que la
segunda pluralidad comprende regiones objetivo de secuencia variable y regiones objetivo epigenéticas.

5. Procedimiento según cualquiera de las reivindicaciones 1 a 4, en el que:

- 55 (i) capturar el primer conjunto de regiones objetivo del primer conjunto comprende poner en contacto el ADN del primer conjunto con un primer conjunto de sondas específicas al objetivo, opcionalmente en el que el primer conjunto de sondas específicas al objetivo comprende sondas de unión a la objetivo específicas para regiones objetivo variables de secuencia;
60 (ii) capturar la segunda pluralidad de conjuntos de regiones objetivo o el segundo conjunto de regiones objetivo del segundo grupo comprende poner en contacto el ADN del segundo grupo con un segundo conjunto de sondas específicas al objetivo, opcionalmente en donde el segundo conjunto de sondas específicas al objetivo comprende sondas de unión a la objetivo específicas para regiones objetivo de secuencia variable y/o sondas de unión a la objetivo específicas para regiones objetivo epigenéticas;
65 (iii) el ADN comprende ADN libre de células (ADNlc);

- (iv) la primera porción del ADN de la partición hipometilada comprende al menos aproximadamente el 50 % del ADN de la partición hipometilada;
- (v) la primera porción del ADN de la partición hipometilada comprende aproximadamente el 50-95 % del ADN de la partición hipometilada; y/o
- (vi) la primera porción del ADN de la partición hipometilada comprende al menos aproximadamente el 80 % del ADN de la partición hipometilada.

6. El método de cualquiera de las reivindicaciones 1-5, en donde:

- (i) el segundo grupo comprende una segunda porción del ADN de la partición hipometilada;
- (ii) la primera porción del ADN de la partición hipometilada comprende una mayor cantidad de ADN de la partición hipometilada que la segunda porción del ADN de la partición hipometilada;
- (iii) la segunda porción del ADN de la partición hipometilada comprende menos o igual a aproximadamente el 50 % del ADN de la partición hipometilada; y/o
- (iv) la segunda porción del ADN de la partición hipometilada comprende menos o igual a aproximadamente el 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o 5 % del ADN de la partición hipometilada.

7. El método de cualquiera de las reivindicaciones 1 a 5, en el que el primer grupo comprende sustancialmente todo el ADN de la partición hipometilada.

8. El método de cualquiera de las reivindicaciones 1-7, en el que:

- (i) la segunda porción comprende al menos aproximadamente el 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % del ADN de la partición hipermetilada;
- (ii) el segundo grupo comprende sustancialmente todo el ADN de la partición hipermetilada;
- (iii) la pluralidad de particiones comprende además una partición intermedia, opcionalmente en el que el segundo grupo comprende al menos una porción de la partición intermedia, tal como en el que el segundo grupo comprende sustancialmente toda la partición intermedia;
- (iv) el segundo conjunto de regiones objetivo o la segunda pluralidad de conjuntos de regiones objetivo comprende un mayor número de regiones objetivo epigenéticas que el primer conjunto de regiones objetivo;
- (v) el primer conjunto de regiones objetivo comprende una mayor cantidad de regiones objetivo de secuencia variable que el segundo conjunto de regiones objetivo o la segunda pluralidad de conjuntos de regiones objetivo;
- (vi) el primer conjunto de regiones objetivo no comprende regiones objetivo epigenéticas;
- (vii) el conjunto de regiones objetivo epigenéticas comprende un conjunto de regiones objetivo variables de hipermetilación; y/o
- (viii) el conjunto de regiones objetivo epigenéticas comprende un conjunto de regiones objetivo variables de fragmentación o el primer conjunto de regiones objetivo comprende regiones objetivo variables de fragmentación, opcionalmente en donde el conjunto de regiones objetivo variables de fragmentación comprende:
 - (a) regiones del sitio de inicio de la transcripción; y/o
 - (b) regiones de unión a CTCF.

9. El método de cualquiera de las reivindicaciones 1-8, en donde:

- (I) la segunda pluralidad de conjuntos de regiones objetivo comprende regiones objetivo variables de fragmentación;
- (II) al menos una región objetivo variable de hipermetilación se captura del segundo grupo, pero no del primer grupo, opcionalmente en donde una pluralidad de regiones objetivo variables de hipermetilación se capturan del segundo grupo pero no del primer grupo; y/o
- (III) el método comprende además secuenciar la primera y segunda pluralidades de conjuntos de regiones objetivo, opcionalmente en donde:
 - (A) las moléculas de ADN correspondientes al conjunto de regiones objetivo de secuencia variable se secuencian a una mayor profundidad de secuenciación que las moléculas de ADNlc correspondientes al conjunto de regiones objetivo epigenéticas;
 - (B) la secuenciación genera una pluralidad de lecturas de secuencia y el método comprende además mapear la pluralidad de lecturas de secuencia a una o más secuencias de referencia para generar lecturas de secuencia mapeadas;
 - (C) la muestra es de un sujeto y el método comprende además determinar la presencia o ausencia de un cáncer en el sujeto basándose al menos en parte en datos generados por la secuenciación de la primera y segunda pluralidades de conjuntos de regiones objetivo;
 - (D) la muestra es de un sujeto y el método comprende además determinar una probabilidad de que el sujeto tenga cáncer basándose al menos en parte en datos generados por la secuenciación de la primera y segunda pluralidades de conjuntos de regiones objetivo;

(E) el método comprende además mapear la pluralidad de lecturas de secuencia a una o más secuencias de referencia para generar lecturas de secuencia mapeadas y procesar las lecturas de secuencia mapeadas correspondientes al conjunto de regiones objetivo de secuencia variable y al conjunto de regiones objetivo epigenéticas para determinar la probabilidad de que el sujeto tenga cáncer; y/o

(F) los recuentos de moléculas se determinan a partir de los resultados de secuenciación de las particiones hipometiladas e hipometiladas, como en donde una fracción del ADN de la partición hipometilada se incluyó en el segundo grupo, por ejemplo, en donde los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman mediante la multiplicación de los recuentos de moléculas observados con un factor de escala, opcionalmente en donde:

(a) el factor de escala es el recíproco de la fracción de la partición hipometilada que se incluyó en el segundo grupo;

(b) los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman utilizando una relación de anclaje determinada en base a frecuencias de la región de control;

(c) los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman utilizando una relación de anclaje determinada en base a los niveles de diversidad;

(d) los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman mediante la multiplicación con un factor de escala determinado (i) a partir de una diferencia media o mediana en la frecuencia de las regiones objetivo epigenéticas en los datos de secuencia de la partición hipometilada de muestras en las que se secuenció una partición hipometilada completa frente a muestras en las que solo se secuenció una parte de una partición hipometilada, o (ii) a partir de una diferencia media o mediana en la frecuencia de las regiones objetivo epigenéticas en los datos de secuencia de la partición hipometilada de una pluralidad de conjuntos de datos de secuencia de una o una pluralidad de muestras, comprendiendo los conjuntos de datos de secuencia datos de secuencia en los que se secuenció una fracción de la partición hipometilada y datos de secuencia en los que se secuenció la partición hipometilada completa;

(e) los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman mediante la multiplicación con un factor de escala determinado utilizando frecuencias de regiones objetivo epigenéticas para las que se incluyeron sondas en la captura de regiones objetivo de secuencia variable del primer grupo; o

(f) los recuentos de moléculas para las regiones objetivo epigenéticas en la partición hipometilada se estiman utilizando una relación entre lecturas y moléculas únicas para inferir un recuento de moléculas que habría resultado de la captura de regiones objetivo epigenéticas de toda la partición hipometilada.

10. El método de cualquiera de las reivindicaciones 1-9, en el que el sujeto de prueba fue diagnosticado previamente con un cáncer y recibió uno o más tratamientos previos contra el cáncer, opcionalmente en el que el ADN se obtiene en uno o más puntos temporales preseleccionados después de uno o más tratamientos previos contra el cáncer.

11. El método de la reivindicación 10, en el que las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable se secuencian, generando de este modo un conjunto de información de secuencia.

12. El método de la reivindicación 11, en el que:

(I) las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable se secuencian a una mayor profundidad de secuenciación que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas; y/o

(II) el método comprende además detectar una presencia o ausencia de ADN originado o derivado de una célula tumoral en un punto temporal preseleccionado utilizando el conjunto de información de secuencia, opcionalmente en donde:

(a) el método comprende además determinar una puntuación de recurrencia de cáncer que es indicativa de la presencia o ausencia del ADN originado o derivado de la célula tumoral para el sujeto de prueba, opcionalmente en donde:

(i) el método comprende además determinar un estado de recurrencia de cáncer basado en la puntuación de recurrencia de cáncer, en donde el estado de recurrencia de cáncer del sujeto de prueba se determina como en riesgo de recurrencia de cáncer cuando se determina que una puntuación de recurrencia de cáncer está en o por encima de un umbral predeterminado o el estado de recurrencia de cáncer del sujeto de prueba se determina como en menor riesgo de recurrencia de cáncer cuando la puntuación de recurrencia de cáncer está por debajo del umbral predeterminado; y/o

(ii) el método comprende además comparar la puntuación de recurrencia del cáncer del sujeto de prueba con un umbral de recurrencia del cáncer predeterminado, y el sujeto de prueba se clasifica como candidato para un tratamiento posterior del cáncer cuando la puntuación de recurrencia del cáncer está por encima del umbral de recurrencia del cáncer o no es candidato para un tratamiento posterior del cáncer cuando la puntuación de recurrencia del cáncer está por debajo del umbral de recurrencia del cáncer; o el sujeto de prueba está en riesgo de recurrencia del cáncer y se clasifica como candidato para un tratamiento posterior

del cáncer, opcionalmente en donde el tratamiento posterior del cáncer comprende quimioterapia o administración de una composición terapéutica;

- (b) el ADN originado o derivado de una célula tumoral es ADN libre de células; y/o
- (c) el ADN originado o derivado de una célula tumoral se obtiene de una muestra de tejido.

13. El método de cualquiera de las reivindicaciones 10 a 12, que comprende además determinar un período de supervivencia libre de enfermedad (SLE) para el sujeto de prueba basándose en la puntuación de recurrencia del cáncer, en donde opcionalmente el período de SLE es 1 año, 2 años, 3 años, 4 años, 5 años o 10 años.

14. El método de cualquiera de las reivindicaciones 11-13, en el que:

(I) el conjunto de información de secuencia comprende secuencias de regiones objetivo variables de secuencia, y la determinación de la puntuación de recurrencia del cáncer comprende determinar al menos una primera subpuntuación indicativa de la cantidad de SNV, inserciones/deleciones, CNV y/o fusiones presentes en secuencias de regiones objetivo variables de secuencia, opcionalmente en el que un número de mutaciones en las regiones objetivo variables de secuencia elegidas entre 1, 2, 3, 4 o 5 es suficiente para que la primera subpuntuación dé como resultado una puntuación de recurrencia del cáncer clasificada como positiva para la recurrencia del cáncer, opcionalmente en el que el número de mutaciones se elige entre 1, 2 o 3;

(II) el conjunto de información de secuencia comprende secuencias de la región objetivo epigenética, y la determinación de la puntuación de recurrencia del cáncer comprende la determinación de una segunda subpuntuación indicativa de la cantidad de lecturas de secuencia anormales en las secuencias de la región objetivo epigenética, opcionalmente en donde las lecturas de secuencia anormales comprenden lecturas indicativas de metilación de secuencias objetivo variables de hipermetilación y/o lecturas indicativas de fragmentación anormal en regiones objetivo variables de fragmentación, tal como en donde una proporción de lecturas correspondientes al conjunto de la región objetivo variable de hipermetilación y/o al conjunto de la región objetivo variable de fragmentación que indican hipermetilación en el conjunto de la región objetivo variable de hipermetilación y/o fragmentación anormal en el conjunto de la región objetivo variable de fragmentación mayor o igual a un valor en el rango de 0,001 %-10 % es suficiente para que la segunda subpuntuación se clasifique como positiva para la recurrencia del cáncer, por ejemplo, en donde el rango es:

- (a) 0,001 %-1 % o 0,005 %-1 %;
- (b) 0,01 %-5 % o 0,01 %-2 %; o
- (c) 0,01 %-1 %;

(III) el método comprende además determinar una fracción de ADN tumoral a partir de la fracción de lecturas en la pluralidad de lecturas de secuencia que indican una o más características indicativas de origen a partir de una célula tumoral, opcionalmente en donde:

(a) las una o más características indicativas de origen a partir de una célula tumoral comprenden una o más alteraciones en una región objetivo variable de secuencia, hipermetilación de una región objetivo variable de hipermetilación y fragmentación anormal de una región objetivo variable de fragmentación;

(b) el método comprende además determinar una puntuación de recurrencia del cáncer basada al menos en parte en la fracción de ADN tumoral, en donde una fracción de ADN tumoral mayor o igual a un valor predeterminado en el rango de 10^{-11} a 1 o de 10^{-10} a 1 es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer, tal como en donde:

- (i) una fracción de ADN tumoral mayor o igual a un valor predeterminado en el rango de 10^{-10} a 10^{-9} , de 10^{-9} a 10^{-8} , de 10^{-5} a 10^{-1} , de 10^{-7} a 10^{-6} , de 10^{-6} a 10^{-5} , de 10^{-5} a 10^{-4} , de 10^{-4} a 10^{-3} , de 10^{-3} a 10^{-2} o de 10^{-2} a 10^{-1} es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer; y/o
- (ii) el valor predeterminado está en el rango de 10^{-5} a 10^{-6} o es 10^{-7} ;

y/o

(c) la fracción de ADN tumoral se determina como mayor o igual que el valor predeterminado si la probabilidad acumulada de que la fracción de ADN tumoral sea mayor o igual que el valor predeterminado es al menos 0,5, 0,75, 0,9, 0,95, 0,98, 0,99, 0,995 o 0,999, opcionalmente en donde la probabilidad acumulada es:

- (i) al menos 0,95; o
- (ii) está en el rango de 0,98-0,995 o es 0,99;

y/o

(IV) el conjunto de información de secuencia comprende secuencias de región objetivo de secuencia variable y secuencias de región objetivo epigenética, y determinar la puntuación de recurrencia del cáncer comprende

determinar una primera subpuntuación indicativa de la cantidad de SNV, inserciones/deleciones, CNV y/o fusiones presentes en secuencias de región objetivo de secuencia variable y una segunda subpuntuación indicativa de la cantidad de lecturas de secuencia anormales en secuencias de región objetivo epigenética, y combinar la primera y la segunda subpuntuaciones para proporcionar la puntuación de recurrencia del cáncer, opcionalmente en donde la combinación de la primera y la segunda subpuntuaciones comprende aplicar un umbral a cada subpuntuación de forma independiente (por ejemplo, mayor que un número predeterminado de mutaciones (por ejemplo, > 1) en regiones objetivo de secuencia variable, y mayor que una fracción predeterminada de lecturas anormales (por ejemplo, tumor) en regiones objetivo epigenéticas), o entrenar un clasificador de aprendizaje automático para determinar el estado en función de una pluralidad de muestras de entrenamiento positivas y negativas, tal como en donde un valor para la puntuación combinada en el rango de -4 a 2 o -3 a 1 es suficiente para que la puntuación de recurrencia del cáncer se clasifique como positiva para la recurrencia del cáncer.

15. El método de cualquiera de las reivindicaciones 10-14, en el que:

- (i) el uno o más puntos temporales preseleccionados se seleccionan del siguiente grupo que consiste en 1 mes, 2 meses, 3 meses, 4 meses, 5 meses, 6 meses, 7 meses, 8 meses, 9 meses, 10 meses, 11 meses, 1 año, 1,5 años, 2 años, 3 años, 4 años y 5 años después de la administración del uno o más tratamientos previos contra el cáncer;
- (ii) el cáncer es cáncer colorrectal;
- (iii) el uno o más tratamientos previos contra el cáncer comprenden cirugía;
- (iv) el uno o más tratamientos previos contra el cáncer comprenden la administración de una composición terapéutica; y/o
- (v) el uno o más tratamientos previos contra el cáncer comprenden quimioterapia.

16. El método de cualquiera de las reivindicaciones 1-15, en el que:

(i) las moléculas de ADN correspondientes al conjunto de regiones objetivo de secuencia variable se capturan del segundo grupo con un rendimiento de captura mayor que las moléculas de ADN correspondientes al conjunto de regiones objetivo epigenéticas, opcionalmente en el que las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable:

- (a) se secuencian a una profundidad de secuenciación al menos 2 veces mayor que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas;
- (b) se secuencian a una profundidad de secuenciación al menos 3 veces mayor que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas;
- (c) se secuencian a una profundidad de secuenciación de 4 a 10 veces mayor que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas; o
- (d) se secuencian a una profundidad de secuenciación de 4 a 100 veces mayor que las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas;

(ii) las regiones objetivo variables de secuencia se secuencian con una cobertura de al menos 1.000X, opcionalmente en donde las regiones objetivo variables de secuencia se secuencian con una mayor cantidad de cobertura que las regiones objetivo epigenéticas;

(iii) las regiones objetivo variables de secuencia se secuencian con una cantidad de cobertura en el rango de 1.000X-20.000X, opcionalmente en donde las regiones objetivo variables de secuencia se secuencian con una mayor cantidad de cobertura que las regiones objetivo epigenéticas;

(iv) las regiones objetivo epigenéticas se secuencian con una cobertura de al menos 1.000X, opcionalmente en donde las regiones objetivo variables de secuencia se secuencian con una mayor cantidad de cobertura que las regiones objetivo epigenéticas;

(v) las regiones objetivo epigenéticas se secuencian con una cantidad de cobertura en el rango de 1.000X-10.000X, opcionalmente en donde las regiones objetivo variables de secuencia se secuencian con una mayor cantidad de cobertura que las regiones objetivo epigenéticas; y/o

(vi) el primer conjunto de regiones objetivo se agrupa con el segundo conjunto de regiones objetivo o la segunda pluralidad de conjuntos de regiones objetivo antes de la secuenciación;

opcionalmente en donde las moléculas de ADN capturadas del conjunto de regiones objetivo de secuencia variable y las moléculas de ADN capturadas del conjunto de regiones objetivo epigenéticas se secuencian en la misma célula de secuenciación.

17. El método de cualquiera de las reivindicaciones 1-16, en donde:

(i) el ADN se amplifica antes de la captura, opcionalmente en donde el método comprende además ligar adaptadores que contienen código de barras al ADN cuando o antes de que se amplifique el ADN;

(ii) capturar el segundo conjunto de regiones objetivo de ADN o la segunda pluralidad de conjuntos de regiones objetivo de ADN comprende poner en contacto el ADN con sondas de unión a la objetivo específicas para un conjunto de regiones objetivo de secuencia variable y sondas de unión a la objetivo específicas para un conjunto

de regiones objetivo epigenéticas, opcionalmente en donde las sondas de unión a la objetivo específicas para el conjunto de regiones objetivo de secuencia variable están:

- (a) presentes en una concentración más alta que las sondas de unión a la objetivo específicas para el conjunto de regiones objetivo epigenéticas;
- (b) presente en una concentración al menos 2 veces mayor que las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas; o
- (c) presente en una concentración al menos 4 o 5 veces mayor que las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas;

opcionalmente en donde las sondas de unión al objetivo específicas para el conjunto de regiones objetivo de secuencia variable tienen una afinidad de unión al objetivo mayor que las sondas de unión al objetivo específicas para el conjunto de regiones objetivo epigenéticas;

(iii) el conjunto de regiones objetivo epigenéticas tiene una huella que es al menos 2 veces mayor que el tamaño del conjunto de regiones objetivo de secuencia variable, opcionalmente en donde la huella del conjunto de regiones objetivo epigenéticas es al menos 10 veces mayor que el tamaño del conjunto de regiones objetivo de secuencia variable;

(iv) el conjunto de regiones objetivo de secuencia variable tiene una huella de al menos 25 kB o 50 kB;

(v) el ADN obtenido del sujeto de prueba se divide en al menos 2 fracciones en función del nivel de metilación, y los pasos posteriores del método se realizan en cada fracción, opcionalmente en donde:

(a) el paso de división comprende poner en contacto el ADN recolectado con un reactivo de unión a metilo inmovilizado en un soporte sólido, opcionalmente en donde el reactivo de unión a metilo comprende un dominio de unión a metilo o una proteína de unión a metilo; y/o

(b) las al menos 2 fracciones comprenden una fracción hipermetilada y una fracción hipometilada, y el método comprende además etiquetar diferencialmente la fracción hipermetilada y la fracción hipometilada o secuenciar por separado la fracción hipermetilada y la fracción hipometilada, tal como en donde la fracción hipermetilada y la fracción hipometilada están etiquetadas diferencialmente y el método comprende además agrupar las fracciones hipermetilada e hipometilada etiquetadas diferencialmente antes de un paso de secuenciación;

(vi) el método comprende además determinar si las moléculas de ADN correspondientes al conjunto de regiones objetivo de secuencia variable comprenden mutaciones asociadas al cáncer;

(vii) el método comprende además determinar si las moléculas de ADN correspondientes al conjunto de regiones objetivo epigenéticas comprenden o indican modificaciones epigenéticas asociadas al cáncer o variaciones en el número de copias (por ejemplo, amplificaciones focales), opcionalmente en donde el método comprende determinar si las moléculas de ADN correspondientes al conjunto de regiones objetivo epigenéticas comprenden o indican modificaciones epigenéticas asociadas al cáncer y variaciones en el número de copias (por ejemplo, amplificaciones focales), opcionalmente en donde las modificaciones epigenéticas asociadas al cáncer comprenden:

- (a) hipermetilación en una o más regiones objetivo variables de hipermetilación;
- (b) una o más perturbaciones de la unión de CTCF; y/o
- (c) una o más perturbaciones de los sitios de inicio de la transcripción;

y/o

(viii) los conjuntos capturados de moléculas de ADN se secuencian utilizando secuenciación de alto rendimiento, pirosecuenciación, secuenciación por síntesis, secuenciación de una sola molécula, secuenciación basada en nanoporos, secuenciación de semiconductores, secuenciación por ligación, secuenciación por hibridación, RNA-Seq (Illumina), expresión génica digital (Helicos), secuenciación de próxima generación (NGS), secuenciación de una sola molécula por síntesis (SMSS) (Helicos), secuenciación masiva paralela, matriz de una sola molécula clonal (Solexa), secuenciación shotgun, Ion Torrent, Oxford Nanopore, Roche Genia, secuenciación Sanger, secuenciación Maxam Gilbert, primer walking, secuenciación utilizando PacBio, SOLiD, Ion Torrent o una plataforma Nanopore.

18. El método de cualquiera de las reivindicaciones 1 a 17, en el que la salida del método se proporciona en el formato de un archivo de computadora.

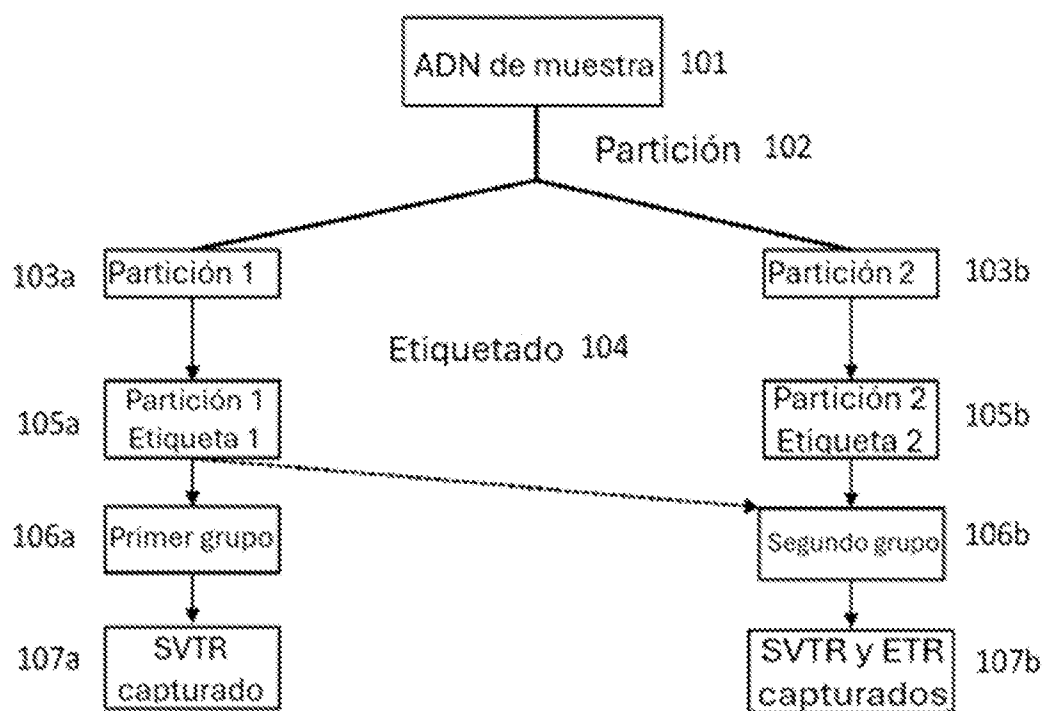


Fig. 1A

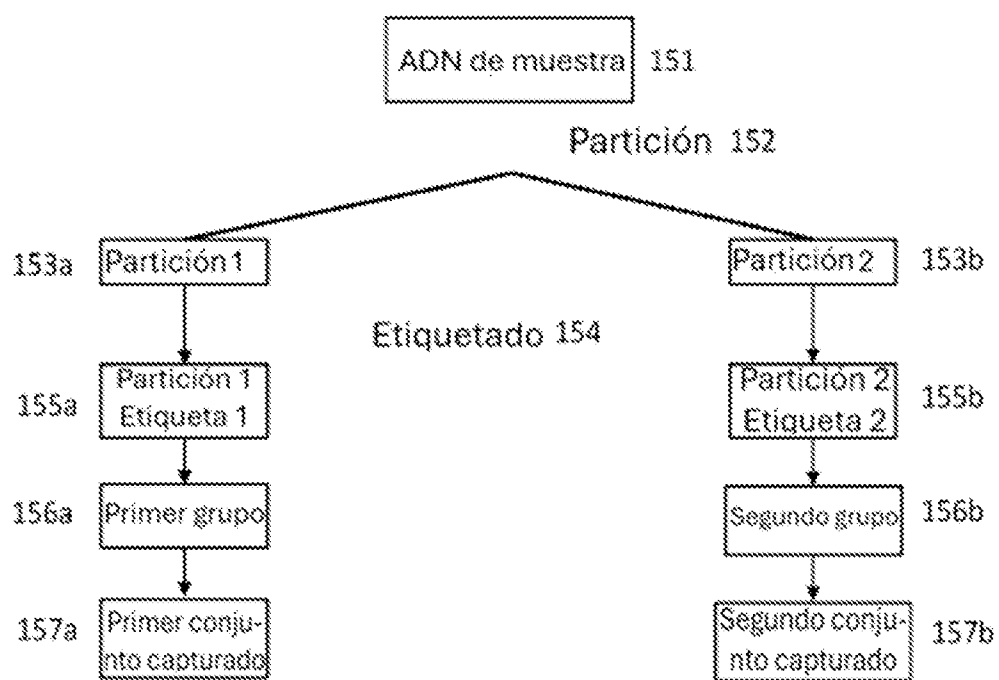


Fig. 1B

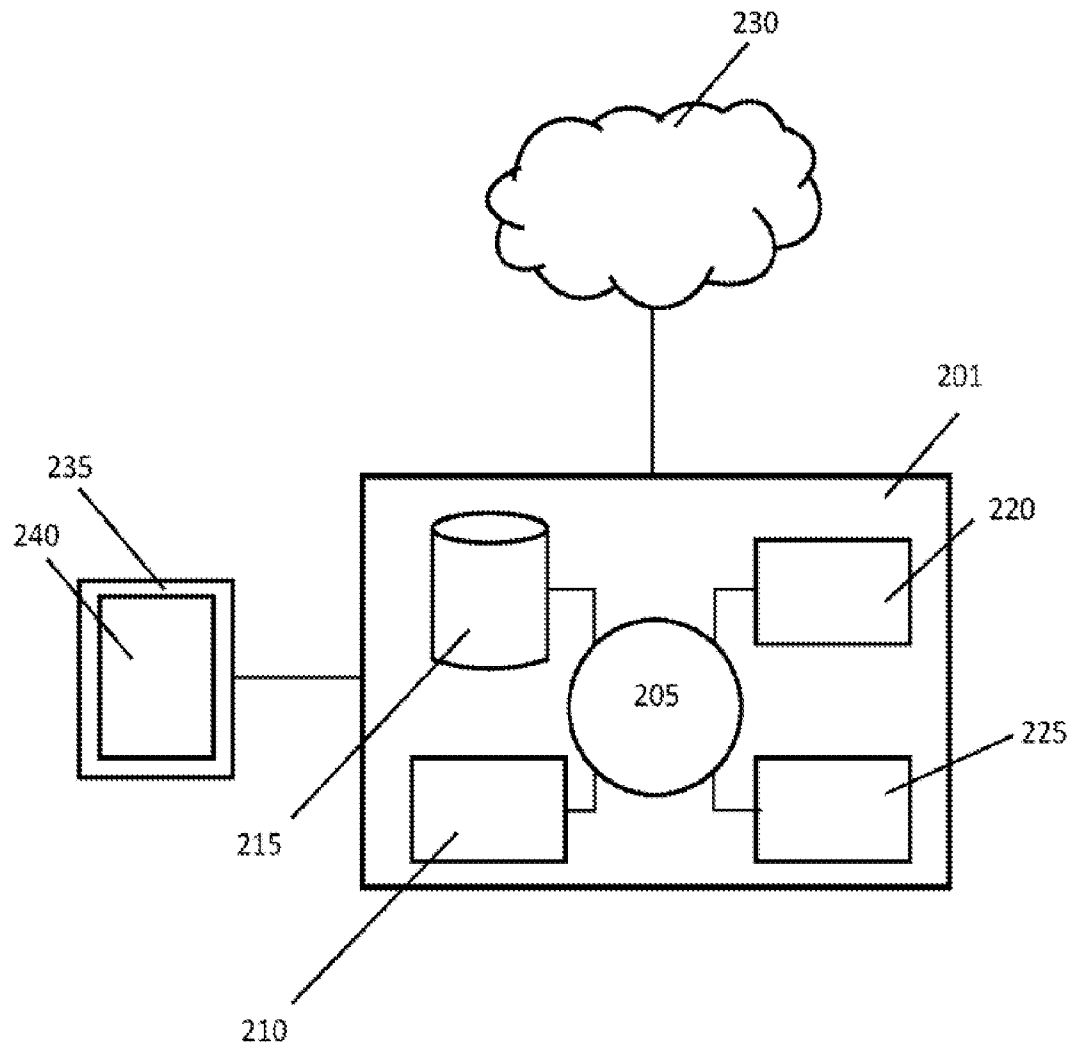


Fig. 2

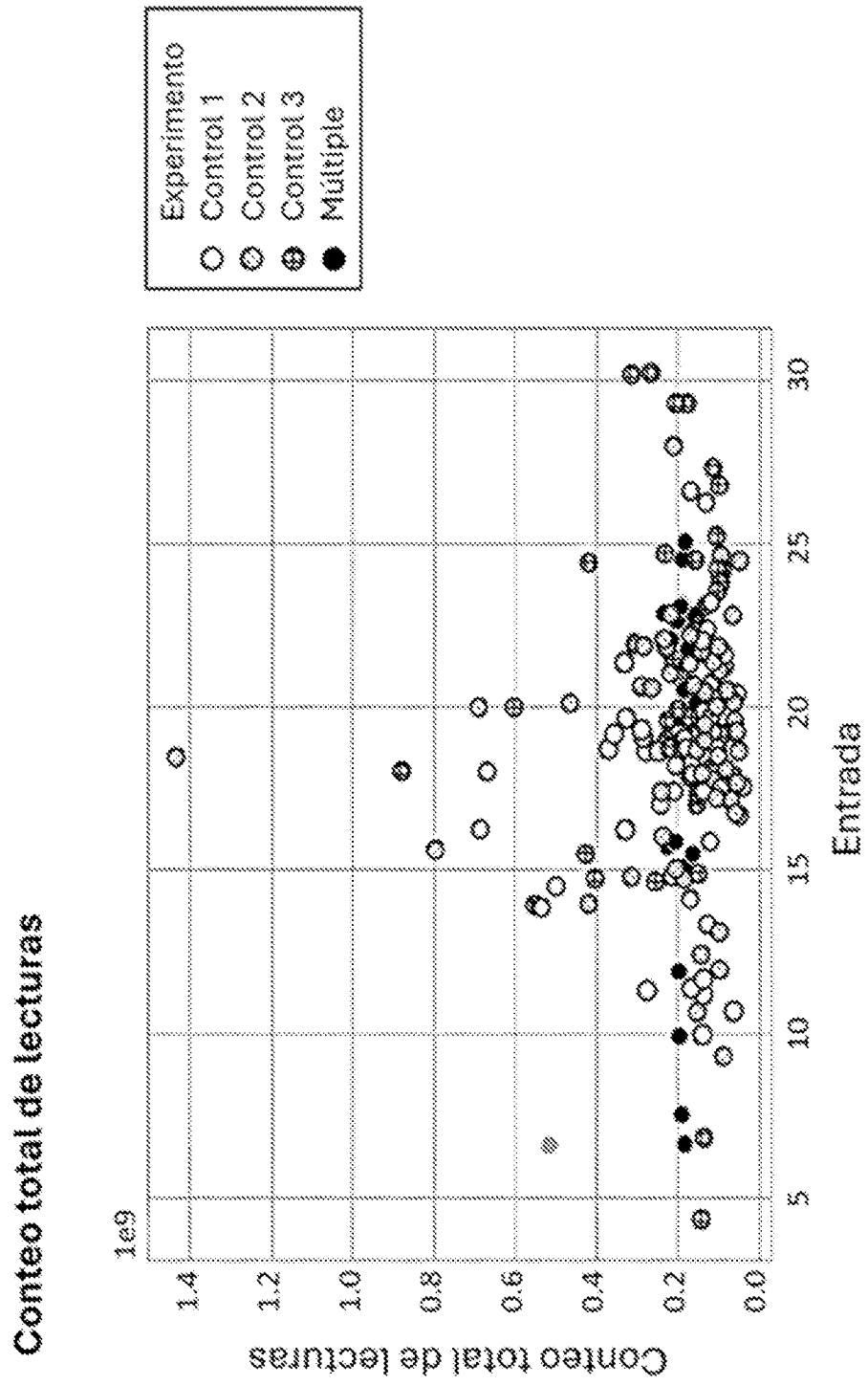


Fig. 3A

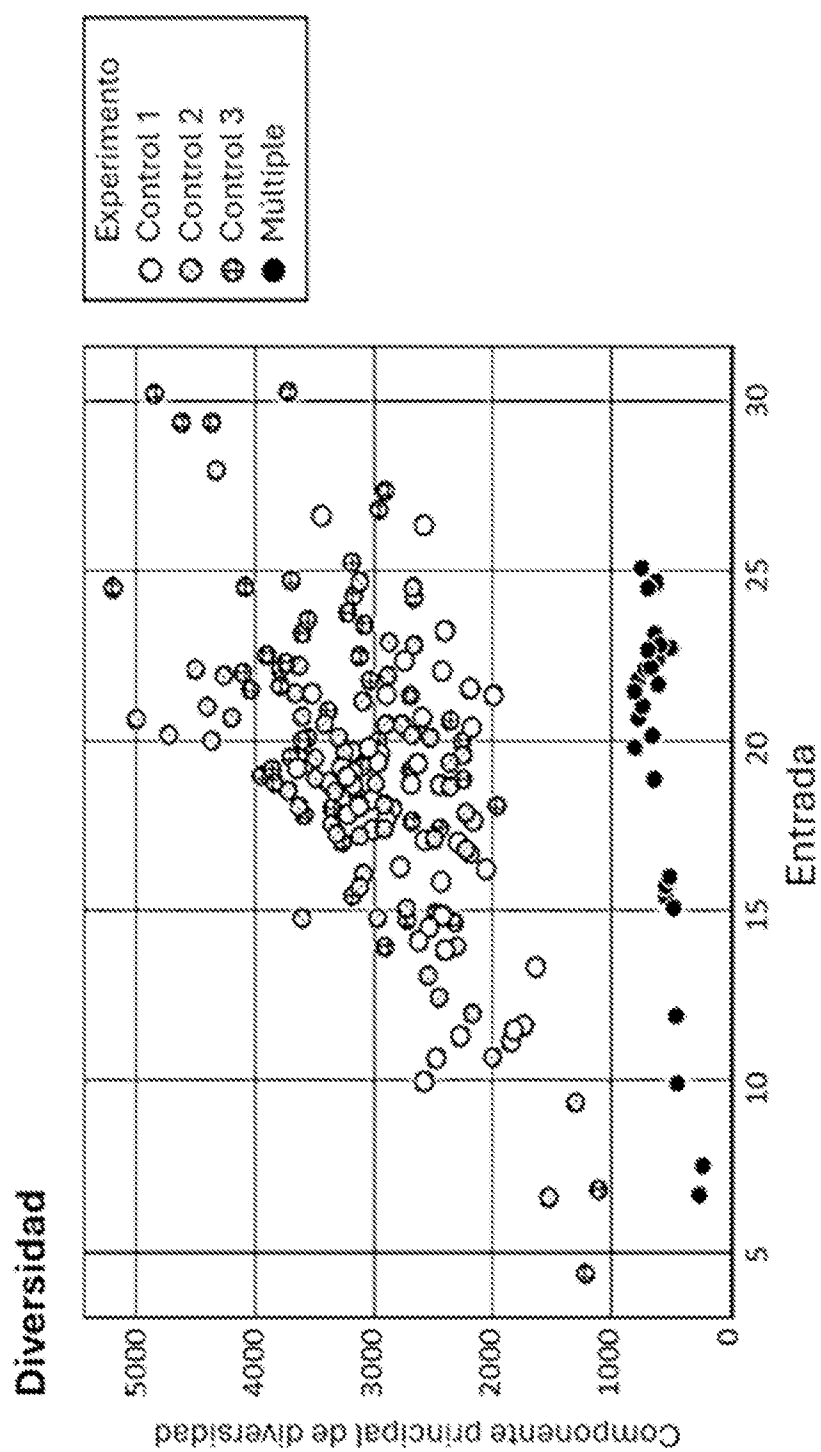


Fig. 3B

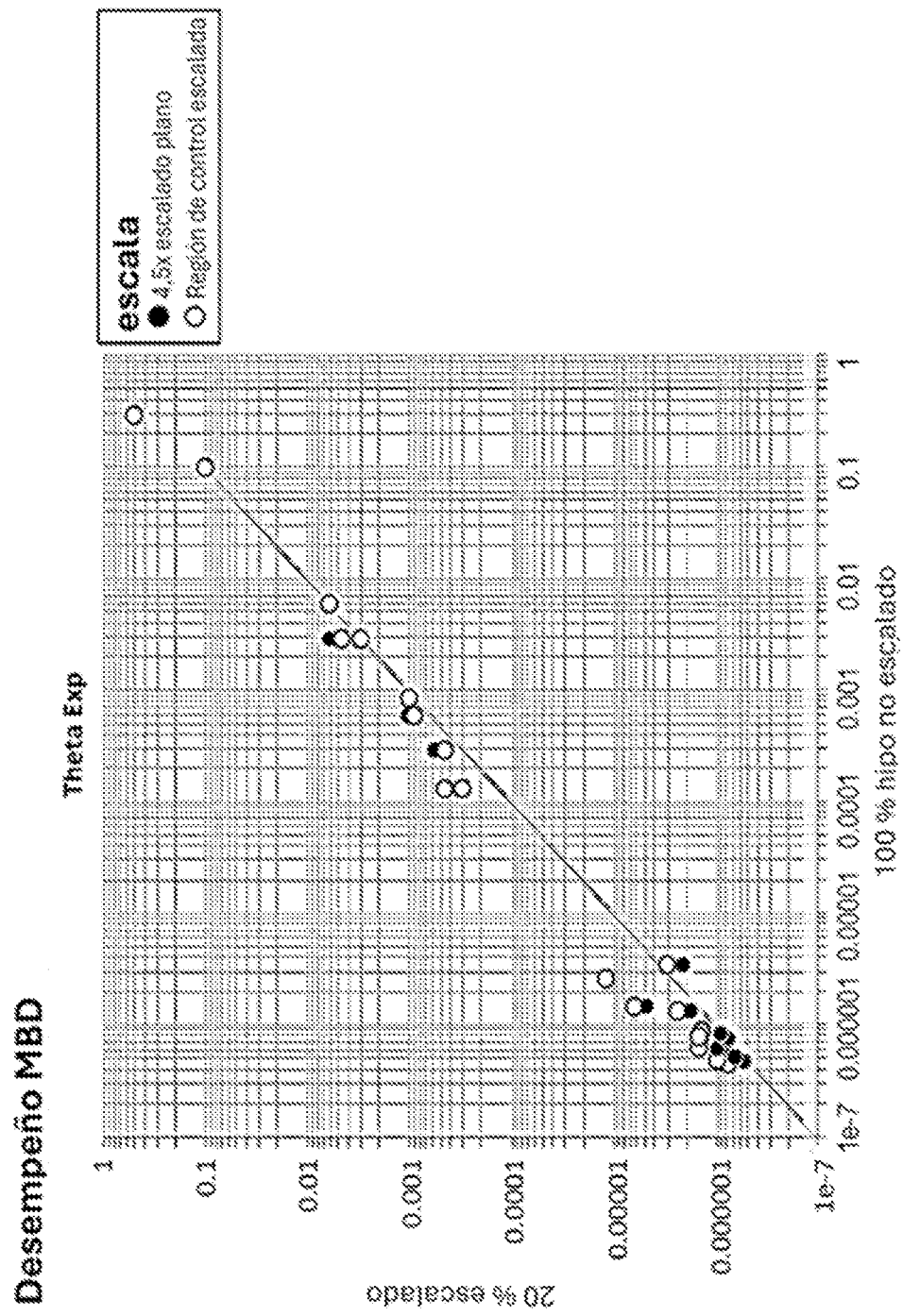


Fig. 3C

Llamada somática

Llamadas somáticas y MAF máx. para muestras de dilución CRC de reproducibilidad hib múltiples

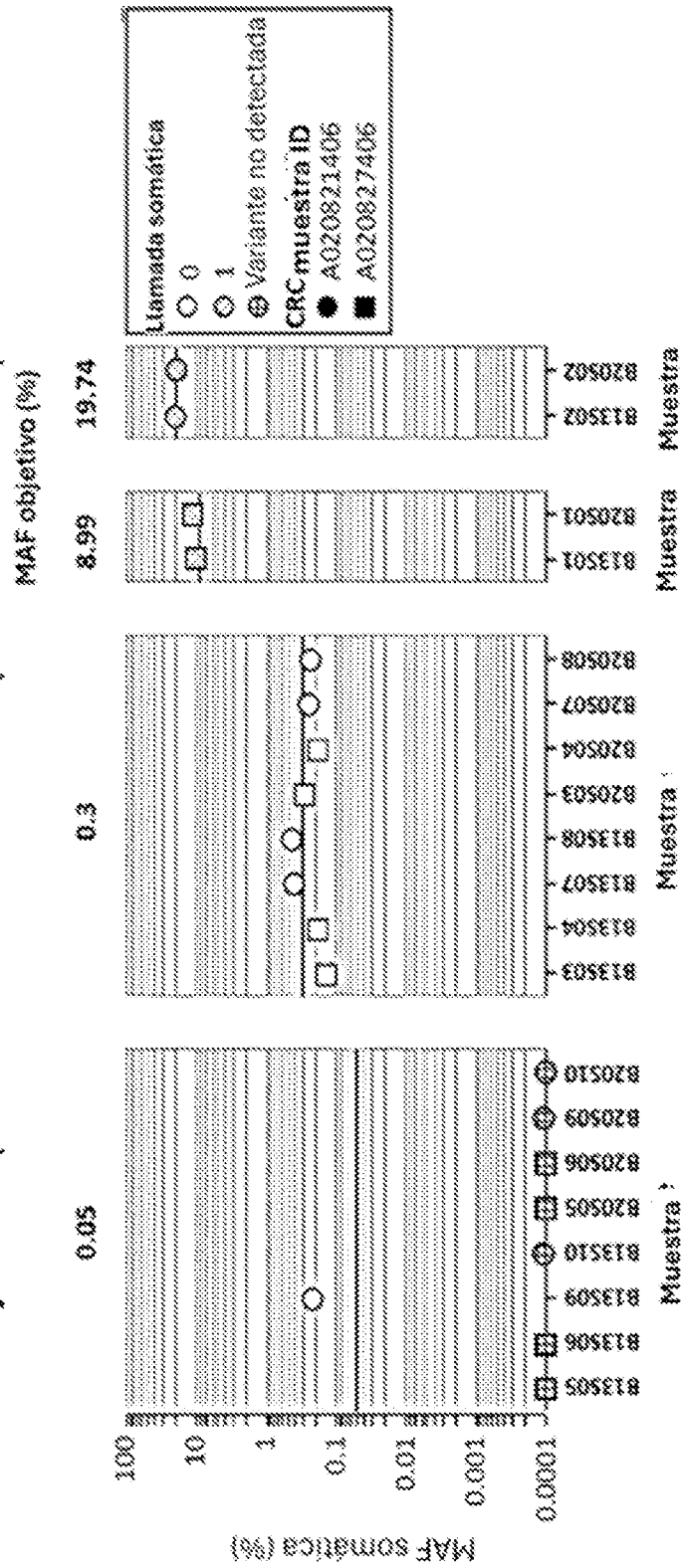


Fig. 4A

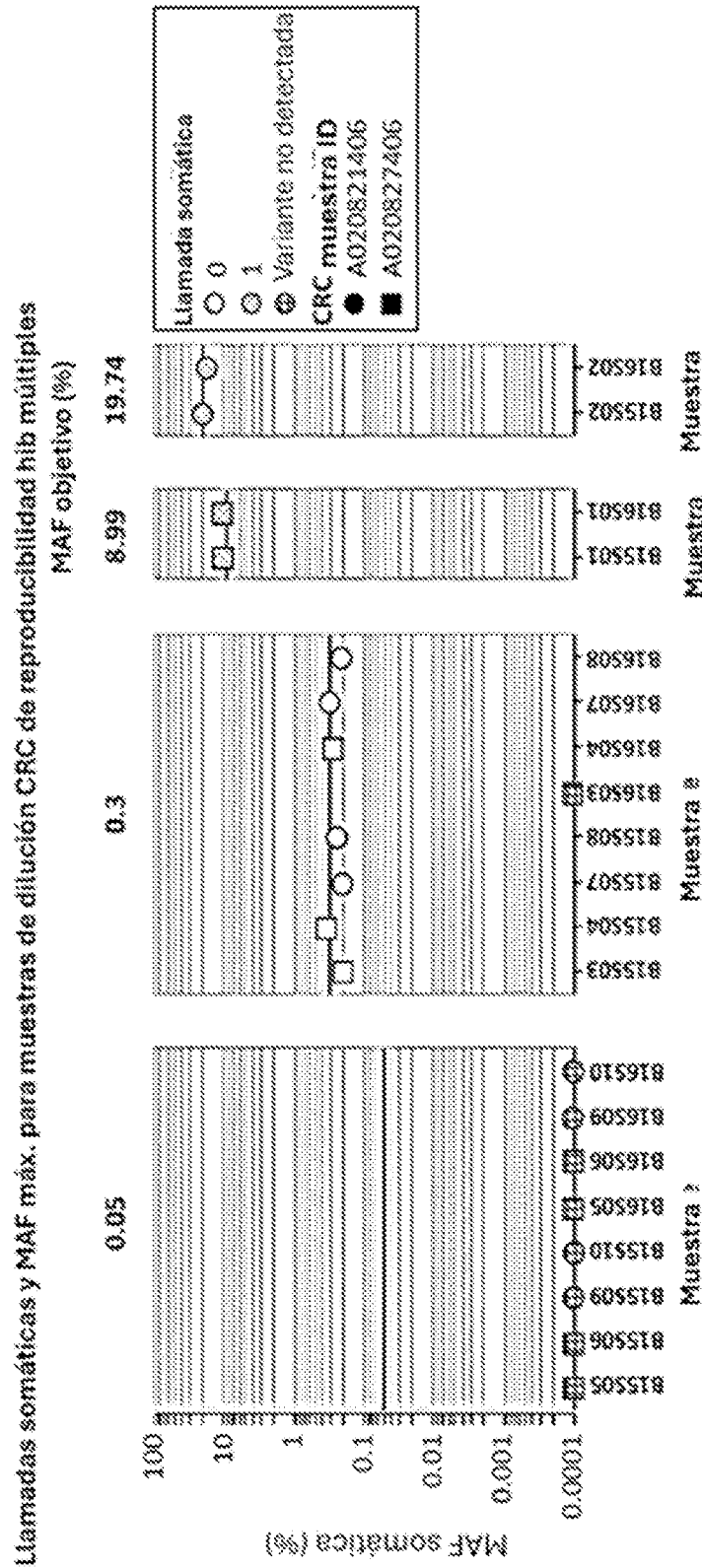


Fig. 4B