

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 984 266**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.08.2014** **PCT/US2014/052317**

87 Fecha y número de publicación internacional: **25.02.2016** **WO16028316**

96 Fecha de presentación y número de la solicitud europea: **22.08.2014** **E 14761762 (5)**

97 Fecha y número de publicación de la concesión europea: **03.07.2024** **EP 3194612**

54 Título: **Métodos para el análisis genético cuantitativo del ADN libre de células**

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
29.10.2024

73 Titular/es:

RESOLUTION BIOSCIENCE, INC. (100.0%)
550 Kirkland Way, Suite 200
Kirkland, WA 98033, US

72 Inventor/es:

RAYMOND, CHRISTOPHER K.;
LIM, LEE P. y
ARMOUR, CHRISTOPHER D.

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

Véase nota informativa (Remarks, Remarques o
Bemerkungen) en el folleto original publicado por la
Oficina Europea de Patentes

ES 2 984 266 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 el análisis genético cuantitativo del ADN libre de células

5 **Antecedentes****Campo técnico**

10 **[0001]** La invención se refiere generalmente a métodos para el análisis genético cuantitativo del ADN libre de células (ADNlc). En particular, la presente invención se refiere a métodos para la caracterización genética y el análisis del ADNlc.

Descripción de la técnica relacionada

15 **[0002]** Cada vez está más claro que la mayoría, si no todos, de los cánceres humanos más comunes son enfermedades del genoma humano. El panorama emergente es que las mutaciones somáticas se acumulan durante la vida de un individuo, algunas de las cuales aumentan la probabilidad de que la célula en la que se encuentran se convierta en un tumor (Vogelstein y col., Science 339 (6127): 1546-1558 (2013)). Con la combinación equivocada de eventos mutacionales acumulados, un crecimiento precanceroso pierde las restricciones que mantienen bajo control la proliferación descontrolada y la masa celular resultante se convierte en cáncer. Las constelaciones de mutaciones que son necesarias y suficientes para causar cáncer frecuentemente se denominan colectivamente “mutaciones activadoras”. Uno de los temas que han surgido de los análisis moleculares recientes e intensivos es que el cáncer, que antes se consideraba una enfermedad única, específica de un tejido, es de hecho un grupo de enfermedades relacionadas, cada una con una patología molecular única. El proyecto del genoma humano sentó las bases para el análisis de los cánceres en todo el genoma.

25 **[0003]** Por ejemplo, la introducción de tecnologías de secuenciación de próxima generación (desde 2004 hasta la actualidad) ha acelerado el ritmo de descubrimiento de los factores genómicos causales que subyacen al diagnóstico del NSCLC, dejando claro que el NSCLC es realmente una familia de enfermedades relacionadas, cada una de las cuales puede responder a una terapia diana diferente.

30 **[0004]** La técnica carece de métodos de análisis molecular fiables y sólidos para el análisis de enfermedades genéticas. Tradicionalmente, el diagnóstico molecular ha consistido en pruebas basadas en anticuerpos (inmunodeficiencia), hibridación in situ con sondas de ADN (hibridación fluorescente in situ) y pruebas basadas en hibridación o PCR que consultan secuencias de nucleótidos específicas. Hasta hace poco, la secuenciación del ADN como herramienta de diagnóstico molecular se limitaba generalmente a los exones codificantes de uno o dos genes. Si bien la secuenciación del ADN se ha usado en el diagnóstico y el tratamiento de cánceres sólidos, uno de los inconvenientes más importantes de estos métodos es que requieren acceso directo a los tejidos tumorales. Este material es frecuentemente difícil de obtener de la biopsia inicial usada para diagnosticar la enfermedad y es prácticamente imposible de obtener en múltiples repeticiones a lo largo del tiempo. Similarmente, las biopsias no son posibles en pacientes con tumores inaccesibles y no son prácticas en personas que padecen una enfermedad metastásica.

45 **[0005]** Por lo tanto, el enorme potencial del diagnóstico molecular para las enfermedades genéticas; pruebas fetales; pruebas de paternidad; predecir la respuesta al tratamiento farmacológico; diagnosticar o controlar una afección médica; trastornos mendelianos; mosaicismo genético; detección de patógenos; elaboración de perfiles de microbiomas; y la monitorización de los trasplantes de órganos; aún no se ha realizado. Hasta la fecha, los enfoques de diagnóstico molecular existentes carecen de soluciones eficientes para clonar y amplificar moléculas de ADN individuales, así como de soluciones para dirigir eficientemente la secuenciación a loci genómicos específicos, con una sensibilidad suficiente para discriminar los resultados positivos verdaderos de las señales positivas falsas que surgen durante el procesamiento de las muestras.

50 **[0006]** La patente WO2014093330 describe un método para generar una genoteca de ADN etiquetado que comprende tratar el ADN fragmentado con enzimas reparadoras terminales para generar ADN reparado terminalmente fragmentado; y ligar una secuencia marcadora de ácido nucleico aleatoria, y opcionalmente una secuencia de código de muestra y/o una secuencia cebadora de PCR al ADN fragmentado con reparación terminal para generar la genoteca de ADN etiquetada.

Breve resumen

60 **[0007]** La invención es un método para el análisis genético diana del ADN libre de células (ADNlc) tal como se define en las reivindicaciones adjuntas. Las composiciones proporcionadas en la presente descripción tienen fines ilustrativos y no entran dentro del alcance de la invención reivindicada. Cualesquiera otros aspectos, configuraciones, instancias o realizaciones expuestas en la presente descripción que no estén dentro del alcance de las reivindicaciones es solo a título informativo.

[0008] La descripción se refiere generalmente a composiciones y métodos para mejorar las composiciones y métodos para el análisis genético del ADNlc.

[0009] De acuerdo con la invención, se proporciona un método para el análisis genético diana del ADN libre de células (ADNlc) tal como se indica en las reivindicaciones, que comprende: tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar clones de la genoteca de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci genéticos diana en los clones de la genoteca de ADNlc.

[0010] En una realización particular, el método comprende además aislar el ADNlc de una muestra biológica de un sujeto.

[0011] En una realización adicional, el ADNlc se aísla de una muestra biológica seleccionada del grupo que consiste en: líquido amniótico, sangre, plasma, suero, semen, líquido linfático, líquido cefalorraquídeo, líquido ocular, orina, saliva, heces, mucosa, y sudor.

[0012] En una cierta realización, el uno o más adaptadores comprende una pluralidad de especies de adaptadores.

[0013] En una realización particular, cada uno del uno o más adaptadores comprende un sitio de unión al cebador para la amplificación de la genoteca de ADNlc.

[0014] En una realización adicional, cada uno del uno o más adaptadores comprende uno o más códigos de lectura únicos.

[0015] En una realización adicional, cada uno del uno o más adaptadores comprende uno o más códigos de muestra para la multiplexación de muestras.

[0016] En otra realización, cada uno del uno o más adaptadores comprende una o más secuencias para la secuenciación del ADN.

[0017] En una realización particular, la qPCR se realiza en la genoteca de clones de ADNlc y la medición de la qPCR se compara con los estándares de equivalentes genómicos conocidos para determinar los equivalentes genómicos de la genoteca de clones de ADNlc.

[0018] En otra realización particular, la qPCR se realiza con un cebador que se une a una secuencia de Alu y un cebador que se une a una secuencia en un adaptador.

[0019] En una cierta realización, el análisis genético cuantitativo se realiza en una pluralidad de loci genéticos en los clones de la genoteca de ADNlc.

[0020] En una realización adicional, el análisis genético cuantitativo se realiza en una pluralidad de loci genéticos en una pluralidad de genotecas de clones de ADNlc.

[0021] El análisis genético cuantitativo comprende hibridar una o más sondas de captura con un locus genético diana para formar complejos de sonda de captura y clon de ADNlc.

[0022] El análisis genético cuantitativo comprende aislar los complejos de clon sonda de captura-ADNlc.

[0023] El análisis genético cuantitativo comprende la amplificación de la secuencia clónica de ADNlc en los complejos aislados de sonda de captura hibridados y clon de ADNlc.

[0024] El análisis genético cuantitativo comprende la secuenciación del ADN para generar una pluralidad de lecturas de secuenciación.

[0025] El análisis genético cuantitativo comprende el análisis bioinformático de la pluralidad de lecturas de secuenciación.

[0026] En una realización particular, se usa el análisis bioinformático: para cuantificar el número de equivalentes genómicos analizados en la genoteca de clones de ADNlc; para detectar variantes genéticas en un locus genético diana; para detectar mutaciones dentro de un locus genético diana; para detectar fusiones genéticas dentro de un locus genético diana; y para medir las fluctuaciones del número de copias dentro de un locus genético diana.

[0027] En una realización adicional, el sujeto no tiene una enfermedad genética.

[0028] En una cierta realización, al sujeto no se le ha diagnosticado una enfermedad genética.

[0029] En otra cierta realización, al sujeto se le ha diagnosticado una enfermedad genética.

[0030] En otra realización, el análisis genético cuantitativo se usa para identificar o detectar una o más lesiones genéticas que causan o se asocian con la enfermedad genética.

[0031] En una cierta realización, la lesión genética comprende una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias, o una fusión de genes.

[0032] En una realización particular, la lesión genética comprende un reordenamiento genómico que fusiona la región codificante 3' del gen ALK con otro gen.

[0033] En una realización particular, la región codificante 3' del gen ALK se fusiona con el gen EML4.

[0034] En otra realización, la enfermedad genética es el cáncer.

[0035] En una realización adicional, la sujeto está embarazada.

[0036] En una realización adicional, el análisis genético cuantitativo se usa para identificar o detectar una o más variantes genéticas o lesiones genéticas de uno o más loci genéticos diana en el ADNlc fetal.

[0037] En una realización particular, el sujeto es un receptor de trasplante.

[0038] En una realización adicional, el análisis genético cuantitativo se usa para identificar o detectar el ADNlc del donante en el sujeto.

[0039] Con fines ilustrativos, se proporciona un método (no reivindicado) para predecir, diagnosticar o monitorizar una enfermedad genética en un sujeto, que comprende: aislar u obtener ADNlc de una muestra biológica de un sujeto; tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar una genoteca de clones de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci genéticos diana asociados con la enfermedad genética en la genoteca de clones de ADNlc, en donde la identificación o detección de una o más lesiones genéticas en uno o más loci genéticos diana sirve para pronosticar, diagnosticar o monitorizar la progresión de la enfermedad genética.

[0040] En una realización adicional, el ADNlc se aísla de una muestra biológica seleccionada del grupo que consiste en: líquido amniótico, sangre, plasma, suero, semen, líquido linfático, líquido cefalorraquídeo, líquido ocular, orina, saliva, heces, mucosa, y sudor.

[0041] En una cierta realización, la lesión genética comprende una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias, o una fusión de genes.

[0042] En una realización particular, la lesión genética comprende un reordenamiento genómico que fusiona la región codificante 3' del gen ALK con otro gen.

[0043] En una realización adicional, la región codificante 3' del gen ALK se fusiona con el gen EML4.

[0044] En una realización particular, la enfermedad genética es el cáncer.

[0045] Con fines ilustrativos, se proporciona un diagnóstico complementario (no reivindicado) para una enfermedad genética que comprende: aislar u obtener ADNlc de una muestra biológica de un sujeto; tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar una genoteca de clones de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más biomarcadores asociados con la enfermedad genética en la genoteca de clones de ADNlc, en donde la detección o la falta de detección de al menos uno del uno o más biomarcadores indica si el sujeto debe recibir tratamiento para la enfermedad genética.

[0046] En una realización particular, el ADNlc se aísla de una muestra biológica seleccionada del grupo que consiste en: líquido amniótico, sangre, plasma, suero, semen, líquido linfático, líquido cefalorraquídeo, líquido ocular, orina, saliva, heces, mucosa y sudor.

[0047] En una realización adicional, el biomarcador es una lesión genética.

[0048] En una realización particular, la lesión genética comprende una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias, o una fusión de genes.

[0049] En una realización adicional, la lesión genética comprende un reordenamiento genómico que fusiona la región codificante 3' del gen ALK con otro gen.

[0050] En una realización adicional, la región codificante 3' del gen ALK se fusiona con el gen EML4.

[0051] En una cierta realización, la enfermedad genética es el cáncer.

Breve descripción de las varias vistas de los dibujos

[0052] La **Figura 1** muestra las frecuencias variantes esperadas frente a las observadas como una función de la dilución del aditivo en ausencia de un filtrado de lectura único. En ausencia de un filtrado de lectura único, se produjeron cambios de base aleatorios en estas cuatro posiciones seleccionadas con frecuencias medibles, distintas de cero; por lo tanto, demuestra una falta de sensibilidad para detectar las variantes particulares de un solo nucleótido (SNV).

La **Figura 2** muestra el filtrado de lectura único realizado en los datos generados en la Figura 1. El panel de la izquierda muestra los datos de la Figura 1 del BRAF I326T SNV sin un filtro de lectura único. El panel de la derecha muestra que el uso de un filtrado de lectura único de los mismos datos aumentó la sensibilidad del ensayo y permitió la discriminación de la señal verdadera del ruido propenso a errores.

La **Figura 3** muestra el rendimiento de la sonda de captura como una función de la longitud y la temperatura de lavado. El eje y muestra el número total de lecturas asociadas a cada sonda de captura. Las barras del gráfico de barras se dividen en dos categorías: las barras abiertas corresponden a las lecturas sobre el diana que se alinean con las dianas de la sonda de captura previstos y las barras sólidas muestran las lecturas fuera del diana que están asociadas a una sonda de captura pero que se asignan a regiones no deseadas del genoma. Generalmente, las sondas de captura de 40 y 60 mer funcionan sustancialmente de la misma manera con lavados a 44 °C y 47 °C. Con el lavado a 50 °C, las sondas de captura de 40 mer funcionan de forma irregular. Estos datos validan el uso de sondas de captura de 40 mer a temperaturas de lavado en el intervalo de aproximadamente 44 °C a aproximadamente 47 °C.

La **Figura 4** muestra un esquema de la secuenciación dirigida y orientada del intrón 19 del gen ALK. A) En la secuencia de referencia “silvestre”, las sondas de captura ALK orientadas al antisentido identifican las secuencias del intrón 19. B) En el caso de genes de fusión patogénica, algunas sondas de captura de ALK identificarán las secuencias de unión asociadas con el evento de fusión de genes.

La **Figura 5** muestra un esquema para la colocación de la sonda de captura de alta densidad para la secuenciación completa de las regiones diana. Cada sonda de captura entonces captura una colección de secuencias que proporcionan una cobertura acumulativa en cada posición base. En este caso, la cobertura está representada por una línea, y la amplitud de la línea indica la profundidad de cobertura derivada de una sonda de captura particular. La cobertura superpuesta de las sondas de captura adyacentes proporciona una secuenciación completa de las regiones diana en ambas direcciones posibles. Además, la colocación cabeza a cabeza de las sondas de captura de cadena opuesta asegura que todos los sitios de unión de las sondas de captura estén secuenciados.

La **Figura 6** muestra un ejemplo representativo de la distribución de tamaños del ADN fragmentado usado en la construcción de genotecas.

La **Figura 7** muestra el rendimiento de sondas de captura de 40 mer de alta densidad en un experimento representativo. El eje y muestra el número total de lecturas, que se desglosan en lecturas dentro del diana, lecturas fuera del diana y lecturas no mapeables. El eje x enumera cada una de las 105 sondas de captura usadas en este experimento para la captura de secuencias.

La **Figura 8** muestra un ejemplo representativo de la cobertura acumulativa de una región diana usando sondas de captura de 40 mer de alta densidad. En la presente descripción se muestra la cobertura acumulada de los exones que codifican TP53.

La **Figura 9A** muestra un ejemplo representativo de la distribución de tamaños de genotecas de ADN libre de células (ADNlc). La banda dominante es consistente con una colección de fragmentos de 170+10 bp ligados a 90 bp de adaptadores. La **Figura 9B** muestra una imagen en gel publicada de ADNlc y una genoteca representativa de ADNlc generada usando los métodos descritos y/o contemplados en la presente descripción. La apariencia cualitativa de “escalera” se conserva en la genoteca, pero la genoteca se desplaza a una masa mayor mediante la adición de 90 pb de secuencias adaptadoras. La **Figura 9C** muestra un ejemplo representativo de genotecas de ADNlc genómicas

derivadas del plasma de pacientes con cáncer de ovario (OvC, por sus siglas en inglés) y “donantes sanos” (HD, por sus siglas en inglés).

La Figura 10 muestra los recuentos de lectura únicos en ocho genotecas de ADNlc derivadas de cuatro muestras de plasma. La fragmentación (frag) antes de la construcción de la genoteca con esta muestra 23407 aumentó el rendimiento de la genoteca en más de dos veces.

La Figura 11 muestra la cobertura de lectura representativa del ADNlc en una región del gen TP53. Se eligieron al azar 24 lecturas de 131 pb capturadas por la sonda de captura “TP53_NM_000546_CHR 17:7579351:region_3:280 nt: 41:80:r” (Id. de sec. N.º 201) y se alinearon usando el algoritmo BLAT en el navegador del genoma de la UCSC. Veintiún lecturas mapean la región diana y lo hacen siguiendo un patrón de cobertura superpuesta. La sonda usada para capturar estas lecturas está marcada con una flecha.

La Figura 12 muestra una visión general de la secuenciación de ADN dirigida de las regiones codificantes del gen TP53 a partir de una genoteca genómica de ADNlc. La cobertura (eje horizontal) se extiende a lo largo de las 10 regiones codificantes e incluye regiones intrónicas implicadas en el empalme del ARNm. La profundidad de secuenciación (eje vertical) alcanza un máximo de 4851 y es uniforme en todos los exones codificantes.

La Figura 13 muestra un gráfico de los recuentos de lectura únicos frente a los equivalentes genómicos estimados por qPCR en un ensayo basado en ACA2. Las mediciones de qPCR se muestran en el eje X frente a los recuentos de lectura en el eje Y. La concordancia perfecta entre estas medidas se muestra en la diagonal. Existe una correlación muy deficiente, si es que la hay, entre las mediciones, especialmente con entradas genómicas más bajas. Estos datos muestran que el ensayo qPCR basado en ACA2 subestima crónicamente la complejidad de la genoteca y es inadecuado para medir los equivalentes del genoma.

La Figura 14 muestra un esquema de los elementos centrales de un ensayo de medición equivalente del genoma de qPCR que acopla un cebador específico de repetición genómica (p. ej., Alu) y un cebador específico de adaptador largo. (A) Amplificación de genotecas estándar usando un único cebador de 25 nt denominado ACA2 (cebador 1). (B) Las versiones más largas de 58 nt del cebador ACA2 (cebador 2) no amplifican las genotecas genómicas debido a la supresión del bucle troncal. (C) Los cebadores directos e inversos dirigidos a un elemento de repetición Alu humano consensuado (cebadores 3 y 4) reconocen miles de loci y amplifican fácilmente el ADN genómico. (D) Un único cebador Alu solo, directo o inverso (cebador 3 o cebador 4), junto con el cebador ACA2 largo (cebador 2) no amplifica el ADN genómico. (E) El mismo par de cebadores que en (D) amplifica fácilmente los clones de la genoteca genómica de ADNlc que contienen secuencias de Alu.

La Figura 15 muestra datos de prueba de concepto para un ensayo de qPCR basado en un adaptador Alu Plus de equivalentes del genoma. (A) Amplificación de 10 pg de una genoteca genómica estándar con varios cebadores de PCR. El eje x especifica los cebadores de PCR usados para la amplificación y el eje Y (escala logarítmica) indica la señal de PCR medida en unidades de fg/μl. El cebador ACA2 estándar produjo una señal fuerte, como se esperaba. El cebador largo ACA2 no pudo producir señal debido a la supresión de la PCR. Los dos pares de cebadores Alu produjeron señal al 1 % de la cantidad de ACA2, lo que sugiere que el 1 % de los clones poseen una secuencia Alu amplificable. La combinación de cualquier cebador Alu con el cebador ACA2 largo también produjo señal en ~ 1 % de los clones. (B) Validación contra 10 pg de ADN genómico (cuatro muestras de la izquierda) o 10 pg de ADN de la genoteca (cuatro clones de la derecha). Los pares de cebadores Alu amplifican una señal comparable del ADN genómico o de una genoteca genómica. Por el contrario, los pares de cebadores que consisten en un cebador Alu y un cebador ACA2 largo amplifican el ADN genómico de manera deficiente (L+A1F) o no amplifican en absoluto (L+A1R). Estos mismos pares muestran una amplificación de la genoteca que supera ligeramente la señal de los pares de cebadores Alu.

La Figura 16 muestra una comparación directa del ensayo de qPCR con cebador ACA2 con el ensayo de qPCR con cebador largo Alu-ACA2. El ensayo de qPCR con cebador largo Alu ACA2 muestra un aumento de 8 veces en los equivalentes genómicos detectables, lo que es más coherente con los recuentos de lectura únicos derivados de los datos de secuenciación.

La Figura 17 muestra un ejemplo representativo de la estructura y función del adaptador para ensayos genéticos cuantitativos de alta sensibilidad, que proporcionan determinaciones precisas de los equivalentes genómicos analizados. (A) Estructura fina de la cadena de ligación del adaptador. Los detalles relacionados con cada elemento numerado se proporcionan en el Ejemplo 4. (B) El dúplex formado entre las cadenas de ligación de 45 nt y la cadena oligo asociada de 12 nt crea un sustrato de ligación de extremos romos compatible con los fragmentos de ADNlc reparados en los extremos (barras sólidas). (C) Tras la ligación, el complemento de la cadena de ligación se crea mediante una reacción de relleno mediada por la ADN polimerasa.

La Figura 18 muestra un ejemplo representativo de la distribución de tamaños de dos muestras de ADN (NA06994 y NCI-H2228) procesadas para imitar el ADNlc.

La Figura 19 muestra un ejemplo representativo de la sensibilidad de detección de la mutación puntual Q331 * de TP53 en una muestra de ADN tumoral (H2228) mezclada con ADN genómico normal (N). La detección más sensible corresponde a ~1 copia mutante de TP53 entre 1000 copias normales del gen.

5 **La Figura 20** muestra la determinación precisa de la secuencia de unión para el gen de fusión EML4-ALK albergado en la línea celular NCI-H2228 usando las composiciones y métodos contemplados en la presente descripción.

10 **La Figura 21** muestra la detección de la muestra de ADN tumoral del gen de fusión EML4-ALK (H2228) mezclada con ADN genómico normal (N). Como la fusión está presente en forma de heterocigoto en la línea celular NCI-H2228, la detección más sensible corresponde a la fusión de un gen entre aproximadamente 100 copias cromosómicas normales del gen ALK (50 equivalentes genómicos).

15 **La Figura 22** muestra la detección de la amplificación del gen MYCN en mezclas de la línea celular NCI-H69 (H69) diluida en ADN humano normal (N). El valor umbral de dos copias diploides normales se muestra como una línea roja discontinua.

20 **La Figura 23** muestra las mutaciones de ADN detectadas en el gen TP53 de tres pacientes con cáncer diferentes. El modelo genético canónico se muestra en la parte superior de la Figura. Los picos representan la cobertura de la secuencia de ADN (eje X) y la profundidad (eje Y) de la secuencia de ADN. La profundidad de secuenciación fue de >4000 equivalentes de genoma para todas las muestras analizadas. Una vista ampliada del exón 7 debajo del modelo genético muestra dónde se localizaron todas las mutaciones detectadas. Se muestra la frecuencia de detección de mutantes en el ADNlc (plasma), el tejido tumoral y el tejido adyacente normal, cuando esté disponible (NA: no disponible). OVA1 y OVA2 son pacientes con cáncer de ovario; Los CRC406 y CRC407 son pacientes con cáncer colorrectal. No se encontraron mutaciones en TP53 en ninguna de las muestras de OVA1.

25 **La Figura 24** muestra la secuenciación del ADN de un panel más grande de trece genes (en caja). La secuenciación identificó una mutación KRAS en el ADNlc y el tumor de la paciente con cáncer de ovario OVA1.

30 **La Figura 25** muestra la secuenciación del ADN de un panel más grande de doce genes. La secuenciación identificó una amplificación del gen ERBB2 en el plasma del paciente con cáncer colorrectal CRC407.

Descripción detallada

A. Visión general

35 **[0053]** La presente invención proporciona un método para el análisis genético diana del ADN libre de células (ADNlc) tal como se define en el conjunto de reivindicaciones. También se describen en la presente descripción con fines ilustrativos composiciones para el análisis genético cuantitativo del estado genético de un individuo usando ADN libre de células (ADNlc); no se reivindican tales composiciones.

40 **[0054]** Como se usa en la presente descripción, el término “estado genético” se refiere a la secuencia de una o más secuencias del genoma diana en el genoma en relación con una secuencia normal no causal o en relación con una secuencia que es causal de una afección o enfermedad genética. En una realización, el análisis del estado genético se refiere a identificar, cuantificar o monitorizar una variante genética en un locus genético diana, en donde la variante varía con respecto a una secuencia de referencia (p. ej., una secuencia normal o mutada). Los presentes inventores han proporcionado soluciones a los problemas de diagnóstico molecular de las afecciones o enfermedades genéticas asociadas con la falta de sensibilidad para discriminar los verdaderos positivos de los falsos positivos, la clonación y amplificación ineficaces de moléculas de ADN individuales y la secuenciación dirigida ineficaz a loci genómicos específicos. Las soluciones contempladas en la presente descripción comprenden composiciones y métodos para un análisis genético cuantitativo fiable y robusto con una sensibilidad suficiente para discriminar los verdaderos resultados positivos de las pruebas de las señales falsas positivas que surgen durante el procesamiento de la muestra.

50 **[0055]** La tecnología de secuenciación de próxima generación ha brindado la oportunidad de agregar amplios estudios genómicos al diagnóstico molecular en una variedad de escenarios que incluyen cánceres, diagnósticos fetales, pruebas de paternidad, detección de patógenos y monitorización de trasplantes de órganos. En el contexto de las enfermedades genéticas, la información de secuenciación de próxima generación se usa en un entorno clínico para identificar las mutaciones en los genes que pueden alterar la función de los genes, identificar la ganancia o la pérdida de material genético dentro de las células e identificar los reordenamientos genómicos que no se encuentran en las células sanas y normales. Los resultados de estas amplias encuestas diagnósticas se usan con frecuencia para guiar el tratamiento de los pacientes.

60 **[0056]** Sin embargo, los beneficios potenciales de la secuenciación del ADN en el diagnóstico y el tratamiento del estado genético de un individuo o de las afecciones o enfermedades genéticas se ven superados por la necesidad de acceder directamente a los tejidos afectados para obtener muestras. Este material es frecuentemente difícil de obtener de la biopsia inicial usada para diagnosticar la enfermedad y es prácticamente imposible de obtener en múltiples repeticiones a lo largo del tiempo. Del mismo modo, en pacientes con cáncer, las biopsias no son posibles en pacientes

con tumores inaccesibles y no son prácticas en personas que padecen una enfermedad metastásica. Por el contrario, el enfoque de los presentes inventores se deriva del hecho de que todos los tejidos requieren acceso a la vasculatura para sobrevivir y, como consecuencia, estas masas depositan ADN en los fluidos corporales. Un depósito importante de fluido corporal en el que se encuentra el ADN de las células enfermas es el plasma de la sangre humana.

[0057] A diferencia de los métodos de prueba que se basan en una cobertura superficial de secuencias de todo el genoma, el diagnóstico molecular contemplado en la presente descripción para determinar el estado genético de un individuo; enfermedades genéticas; trastornos mendelianos; mosaicismo genético; pruebas fetales; pruebas de paternidad; predecir la respuesta al tratamiento farmacológico; diagnosticar o controlar una afección médica; detección de patógenos; elaboración de perfiles de microbiomas; y la monitorización de los trasplantes de órganos aprovechan la disponibilidad del ADNlc para proporcionar una cobertura de secuencia profunda de genes diana seleccionados. Además, los diagnósticos del cáncer basados en el ADNlc contemplados en la presente descripción poseen la capacidad de detectar una variedad de cambios genéticos, incluidas las variaciones de la secuencia somática que alteran la función de las proteínas, los reordenamientos cromosómicos a gran escala que crean fusiones génicas quiméricas y la variación del número de copias que incluye la pérdida o la ganancia de copias genéticas. Usando las composiciones y los métodos contemplados, estos cambios son detectables y cuantificables ante una dilución o mezcla significativa de secuencias normales dentro del ADNlc, a las que contribuyen los procesos normales de renovación que tienen lugar en los tejidos sanos. Las composiciones y los métodos contemplados en la presente descripción también abordan con éxito los principales desafíos asociados con la detección de cambios genéticos raros causantes de la enfermedad; es decir, que el ADNlc está muy fragmentado, que los niveles de ADNlc varían sustancialmente entre diferentes individuos y que el grado de mezcla de secuencias enfermas y normales es muy variable entre los pacientes, incluso entre los individuos que padecen la misma enfermedad molecular y el mismo estadio.

[0058] En diversas realizaciones, las composiciones (no reivindicadas) y los métodos para el análisis genético comprenden interrogar la fracción de ADN en muestras de fluidos biológicos y muestras de heces. Los métodos contemplados en la presente descripción proporcionan un marco novedoso y completo que aborda el análisis genético molecular usando ADNlc disponible a partir de una variedad de fuentes biológicas. La clonación de ADNlc purificado introduce secuencias de ADNlc etiquetadas que informan el análisis posterior y permiten la amplificación de las genotecas de clones resultantes. La captura híbrida con oligonucleótidos específicos del diana se usa para recuperar secuencias específicas para su posterior análisis. Se aplican mediciones independientes del número de genomas presentes en la genoteca a cada muestra, y estos ensayos proporcionan un medio para estimar la sensibilidad del ensayo. Los ensayos contemplados en la presente descripción proporcionan métodos fiables, reproducibles y robustos para el análisis, la detección, el diagnóstico o la monitorización de estados genéticos, afecciones o enfermedades.

[0059] La práctica de realizaciones particulares de la invención empleará, a menos que se indique específicamente lo contrario, métodos convencionales de química, bioquímica, química orgánica, biología molecular, microbiología, técnicas de ADN recombinante, genética, inmunología, y biología celular que están dentro de la técnica, muchos de los cuales se describen más abajo a título ilustrativo. Tales técnicas se explican completamente en la literatura. Véase, p. ej., Sambrook, y col., *Molecular Cloning: A Laboratory Manual* (3.ª edición, 2001); Sambrook, y col., *Molecular Cloning: A Laboratory Manual* (2.ª edición, 1989); Maniatis y col., *Molecular Cloning: A Laboratory Manual* (1982); Ausubel y col., *Current Protocols in Molecular Biology* (John Wiley and Sons, actualizado julio de 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); *Transcripción y traducción* (B. Hames y S. Higgins, editores, 1984); Perbal, *A Practical Guide to Molecular Cloning* (1984); y Harlow y Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998).

B. Definiciones

[0060] A menos que se defina de cualquier otra manera, todos los términos técnicos y científicos usados en la presente descripción tienen el mismo significado que se entiende comúnmente por los expertos en la técnica a la que pertenece la invención. Aunque cualquier método y material similar o equivalente a los descritos en la presente descripción puede usarse en la práctica o prueba de la presente invención, en la presente descripción se describen realizaciones preferidas de composiciones, métodos y materiales. Para los fines de la presente invención, los siguientes términos se definen a continuación.

[0061] Los artículos “un”, “una” y “el/la” se usan en la presente descripción para referirse a uno o más de uno (es decir, al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa un elemento o más de un elemento.

[0062] El uso de la alternativa (por ejemplo, “o”) debe entenderse que significa una, ambas o cualquier combinación de las alternativas.

[0063] El término “y/o” debe entenderse que significa una o ambas alternativas.

[0064] Como se usa en la presente, el término “aproximadamente” o “alrededor de” se refiere a una cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud que varía tanto como 15 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 % o 1 % con respecto a una cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud de referencia. En una realización, el término “aproximadamente” o “alrededor de” se refiere a un intervalo de cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud $\pm 15 \%$, $\pm 10 \%$, $\pm 9 \%$, $\pm 8 \%$, $\pm 7 \%$, $\pm 6 \%$, $\pm 5 \%$, $\pm 4 \%$, $\pm 3 \%$, $\pm 2 \%$ o $\pm 1 \%$ sobre una cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud.

[0065] A lo largo de esta descripción, a menos que el contexto lo requiera de cualquier otra manera, se entenderá que las palabras “comprende”, “comprenden” y “que comprende” implican la inclusión de una etapa o elemento o grupo de etapas o elementos establecidos, pero no la exclusión de cualquier otra etapa o elemento o grupo de etapas o elementos. En realizaciones particulares, los términos “incluye”, “tiene”, “contiene” y “comprende” se usan como sinónimos.

[0066] Por “que consiste en” se entiende que incluye, y se limita a, lo que siga a la frase “que consiste en”. Por lo tanto, la frase “que consiste en” indica que los elementos enumerados son obligatorios u obligatorios y que no pueden estar presentes otros elementos.

[0067] Se entiende que “consiste esencialmente en” incluye todos los elementos mencionados después de la frase, y se limita a otros elementos que no interfieren o no contribuyen a la actividad o acción especificada en la descripción de los elementos mencionados. Por lo tanto, la frase “que consiste esencialmente en” indica que los elementos enumerados son obligatorios u obligatorios, pero que no están presentes otros elementos que afecten materialmente la actividad o acción de los elementos enumerados.

[0068] La referencia en esta descripción a “una realización”, “una realización particular”, “una realización relacionada”, “una cierta realización”, “una realización adicional” u “otra realización” o combinaciones de las mismas significa que una característica, estructura o rasgo en particular descrita en relación con la realización se incluye en al menos una realización de la presente invención. Por lo tanto, las apariciones de las frases anteriores en diversos lugares a lo largo de esta especificación no se refieren todas necesariamente a la misma realización. Además, los rasgos, estructuras, o características particulares se pueden combinar de cualquier manera adecuada en una o más realizaciones.

[0069] Tal como se usa en la presente descripción, el término “aislado” significa material que está sustancial o esencialmente exento de los componentes que normalmente lo acompañan en su estado nativo. En realizaciones particulares, el término “obtenido” o “derivado” se usa como sinónimo de aislado.

[0070] Tal como se usa en la presente descripción, el término “ADN” se refiere al ácido desoxirribonucleico. En diversas realizaciones, el término ADN se refiere a ADN genómico, ADN recombinante, ADN sintético, o ADNc. En una realización, el ADN se refiere al ADN genómico o al ADNc. En realizaciones particulares, el DNA comprende una “región diana”. Las genotecas de ADN contempladas en la presente descripción incluyen genotecas de ADN genómico y genotecas de ADNc construidas a partir de ARN, *p. ej.*, una genoteca de expresión de ARN. En diversas realizaciones, las genotecas de ADN comprenden una o más secuencias y/o etiquetas de ADN adicionales.

[0071] Un “locus genético diana” o “región diana de ADN” se refiere a una región de interés dentro de una secuencia de ADN. En diversas realizaciones, los análisis genéticos dirigidos se realizan en el locus genético diana. En realizaciones particulares, la región diana del ADN es una región de un gen que está asociada con un estado genético, una afección genética, enfermedades genéticas particulares; pruebas fetales; mosaicismo genético, pruebas de paternidad; predecir la respuesta al tratamiento farmacológico; diagnosticar o controlar una afección médica; elaboración de perfiles de microbiomas; detección de patógenos; o monitorización de trasplantes de órganos.

[0072] Tal como se usan en la presente descripción, los términos “ADN circulante”, “ADN libre de células circulantes” y “ADN libre de células” frecuentemente se usan indistintamente y se refieren al ADN que es ADN extracelular, al ADN que se ha extruido de las células o al ADN que se ha liberado de células necróticas o apoptóticas.

[0073] Un “sujeto”, “individuo” o “paciente”, tal como se usa en la presente descripción, incluye cualquier animal que presente un síntoma de una afección que pueda detectarse o identificarse con las composiciones contempladas en la presente descripción. Los sujetos adecuados incluyen animales de laboratorio (como ratones, ratas, conejos, o conejillos de Indias), animales de granja (como caballos, vacas, ovejas, cerdos) y animales domésticos o mascotas (como un gato o un perro). En realizaciones particulares, el sujeto es un mamífero. En ciertas realizaciones, el sujeto es un primate no humano y, en realizaciones preferidas, el sujeto es un ser humano.

C. Métodos de análisis genético del ADN libre de células

[0074] La invención proporciona un método para el análisis genético del ADNlc, tal como se indica en el conjunto de reivindicaciones.

[0075] El método comprende: generar y amplificar una genoteca de ADNlc, determinar el número de equivalentes genómicos en la genoteca de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci diana genómicos.

[0076] En particular, el método comprende tratar el ADNlc con una o más enzimas reparadoras de extremos para generar un ADNlc reparado en los extremos y ligar uno o más adaptadores a cada extremo del ADNlc reparado en los extremos para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar clones de la genoteca de ADNlc; determinar el número de equivalentes genómicos de los clones de la genoteca de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci genéticos diana en los clones de la genoteca de ADNlc.

1. Generar una genoteca de ADNlc

[0077] Los métodos de análisis genético contemplados en la presente descripción comprenden generar una genoteca de ADNlc que comprende tratar el ADNlc con una o más enzimas reparadoras de extremos para generar ADNlc reparado en los extremos y ligar uno o más adaptadores a cada extremo del ADNlc reparado en los extremos para generar la genoteca de ADNlc.

(a) ADN libre de células (ADNlc)

[0078] Los métodos y composiciones contemplados en la presente descripción se diseñan para analizar, detectar, diagnosticar y/o monitorizar de manera eficiente los estados genéticos, las afecciones genéticas, las enfermedades genéticas, el mosaicismo genético, el diagnóstico fetal, las pruebas de paternidad, la elaboración de perfiles de microbiomas, la detección de patógenos y la monitorización del trasplante de órganos usando ADN libre de células (ADNlc) como analito. La distribución de tamaños de los fragmentos de ADNlc oscila entre aproximadamente 150 bp y aproximadamente 180 bp. La fragmentación puede ser el resultado de la actividad endonucleolítica y/o exonucleolítica y presenta un desafío formidable para el análisis preciso, confiable y sólido del ADNlc. Otro desafío para analizar el ADNlc es su corta vida media en el torrente sanguíneo, del orden de unos 15 minutos. Sin pretender ceñirse a ninguna teoría en particular, la presente invención contempla, en parte, que el análisis del ADNlc es como una “biopsia líquida” y es una instantánea en tiempo real de los procesos biológicos actuales.

[0079] Además, dado que el ADNlc no se encuentra dentro de las células y puede obtenerse de varias fuentes adecuadas que incluyen, entre otras, fluidos biológicos y muestras de heces, no está sujeto a las limitaciones existentes que afectan al análisis de secuenciación de próxima generación, como el acceso directo a los tejidos que se están analizando.

[0080] Los ejemplos ilustrativos de fluidos biológicos que son fuentes adecuadas de los que aislar el ADNlc en casos particulares incluyen, pero no se limitan a, el líquido amniótico, la sangre, el plasma, el suero, el semen, el líquido linfático, el líquido cefalorraquídeo, el líquido ocular, la orina, la saliva, la mucosa, y el sudor.

[0081] En realizaciones particulares, el fluido biológico es sangre o plasma sanguíneo.

[0082] En ciertas realizaciones, los kits disponibles en el mercado y otros métodos conocidos por el experto en la materia pueden usarse para aislar el ADNlc directamente de los fluidos biológicos de un paciente o de una muestra biológica previamente obtenida y opcionalmente estabilizada, p. ej., mediante la congelación y/o la adición de agentes quelantes enzimáticos que incluyen, pero no se limitan a, EDTA, EGTA, u otros agentes quelantes específicos para cationes divalentes.

(b) Generación de ADNlc reparado en los extremos

[0083] La generación de una genoteca de ADNlc comprende la reparación final del ADNlc aislado. El ADNlc fragmentado es procesado por enzimas reparadoras de extremos para generar ADNlc reparado en los extremos con extremos romos, salientes de 5' o salientes de 3'. En algunas realizaciones, las enzimas reparadoras terminales pueden producir, por ejemplo. En algunas realizaciones, el ADNlc reparado en los extremos contiene extremos romos. En algunas realizaciones, el ADNlc reparado en los extremos se procesa para que contenga extremos romos. En algunas realizaciones, los extremos romos del ADNlc reparado en los extremos se modifican adicionalmente para contener un único par de bases saliente. En algunas realizaciones, el ADNlc reparado en los extremos que contiene extremos romos puede procesarse adicionalmente para que contenga un saliente de adenina (A)/timina (T). En algunas realizaciones, el ADNlc reparado en los extremos que contiene extremos romos puede procesarse adicionalmente para que contenga adenina (A)/timina (T) saliente como el único saliente de par de bases. En algunas realizaciones, el ADNlc reparado en los extremos tiene salientes en 3' sin plantilla. En algunas realizaciones, el ADNlc reparado en los extremos se procesa para contener salientes en 3'. En algunas realizaciones, el ADNlc reparado en los extremos se procesa con la transferasa terminal (TdT) para contener salientes en 3'. En algunas realizaciones, TdT puede añadir una cola en G. *En algunas realizaciones, el ADNlc reparado en los extremos se procesa para que contenga los extremos salientes mediante digestión parcial con cualquier enzima de restricción conocida (p. ej., con la enzima Sau3A y lo similar.*

(c) Ligación de moléculas adaptadoras a ADNlc reparado en los extremos

[0084] La generación de una genoteca de ADNlc comprende ligar uno o más adaptadores a cada extremo del ADNlc reparado en los extremos. La presente invención contempla, en parte, un módulo adaptador diseñado para alojar un gran número de equivalentes genómicos en genotecas de ADNlc. Los módulos adaptadores están configurados para medir el número de equivalentes del genoma presentes en las genotecas de ADNlc y, por extensión, la sensibilidad de los ensayos de secuenciación usados para identificar las mutaciones de secuencia.

[0085] Tal como se usa en la presente descripción, el término “módulo adaptador” se refiere a un polinucleótido que comprende al menos cinco elementos: (i) un primer elemento que comprende un sitio de unión al cebador de la PCR para la amplificación de la genoteca de un único cebador; (ii) un segundo elemento que comprende un código de lectura de 5 nucleótidos que actúa para identificar de forma única cada lectura de secuenciación; (iii) un tercer elemento que comprende un código de muestra de 3 nucleótidos para identificar diferentes muestras y permitir la multiplexación de muestras dentro de una secuencia de secuenciación; (iv) un cuarto elemento que comprende una secuencia de anclaje de 12 nucleótidos que permite la calibración de las llamadas de bases adecuadas en las lecturas de secuenciación y actúa como un ancla para la hibridación con un oligonucleótido asociado; y (v) un quinto elemento que comprende los dos nucleótidos terminales 3' del elemento 4 (figura 17 y tablas 12-16). El módulo adaptador se hibrida con un oligonucleótido asociado que es complementario al elemento 4 para formar un adaptador adecuado para ligarse a los extremos del ADNlc, opcionalmente un ADNlc de extremos romos reparado en los extremos.

[0086] En realizaciones particulares, un módulo adaptador comprende una o más secuencias cebadoras de PCR, uno o más códigos de lectura, uno o más códigos de muestra, una o más secuencias de anclaje, y dos o más nucleótidos 3' que son sustratos de ligación eficientes. En realizaciones adicionales, el módulo adaptador comprende además uno o más sitios de unión al cebador de secuenciación.

[0087] En realizaciones particulares, un módulo adaptador comprende un primer elemento que comprende una o más secuencias de unión a cebadores de PCR para la amplificación con un único cebador de una genoteca de ADNlc. En una realización, la secuencia de unión al cebador de la PCR es de aproximadamente 12 a aproximadamente 40 nucleótidos, de aproximadamente 18 a aproximadamente 40 nucleótidos, de aproximadamente 20 a aproximadamente 35 nucleótidos o de aproximadamente 20 a aproximadamente 30 nucleótidos. En otra realización, la secuencia de unión al cebador de la PCR es de aproximadamente 12 nucleótidos, aproximadamente 13 nucleótidos, aproximadamente 14 nucleótidos, aproximadamente 15 nucleótidos, aproximadamente 16 nucleótidos, aproximadamente 17 nucleótidos, aproximadamente 18 nucleótidos, aproximadamente 19 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 21 nucleótidos, aproximadamente 22 nucleótidos, aproximadamente 23 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 25 nucleótidos, aproximadamente 26 nucleótidos, aproximadamente 27 nucleótidos, aproximadamente 28 nucleótidos, aproximadamente 29 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 31 nucleótidos, aproximadamente 32 nucleótidos, aproximadamente 33 nucleótidos, aproximadamente 34 nucleótidos, aproximadamente 35 nucleótidos, aproximadamente 36 nucleótidos, aproximadamente 37 nucleótidos, aproximadamente 38 nucleótidos, aproximadamente 39 nucleótidos o aproximadamente 40 nucleótidos o más.

[0088] En una realización, a secuencia de unión al cebador de la PCR es de aproximadamente 25 nucleótidos.

[0089] En realizaciones particulares, un módulo adaptador comprende un segundo elemento que comprende una o más secuencias de código de lectura. Tal como se usa en la presente descripción, el término “leer código” se refiere a un polinucleótido que se usa para identificar lecturas de secuenciación únicas. En una realización, el código de lectura es una secuencia aleatoria de nucleótidos. En una realización, el código de lectura es de aproximadamente 1 nucleótido, aproximadamente 2 nucleótidos, aproximadamente 3 nucleótidos, aproximadamente 4 nucleótidos, aproximadamente 5 nucleótidos, aproximadamente 6 nucleótidos, aproximadamente 7 nucleótidos, aproximadamente 8 nucleótidos, aproximadamente 9 nucleótidos, aproximadamente 10 nucleótidos, o más.

[0090] A modo de ejemplo no limitativo, un código de 5 nucleótidos consiste en 256 posibles secuencias únicas en las que cada código elegido es 2 nucleótidos diferente de cualquier otro código del conjunto. Esta función permite diferenciar las lecturas únicas y distintas de las lecturas que parecen ser únicas debido a un error de secuenciación en la región del código. En realizaciones particulares, los códigos que se ha determinado empíricamente que interfieren con la función del adaptador, debido a combinaciones de secuencias particulares, pueden excluirse del uso, *p. ej.*, siete códigos de los 256 tenían una sobrerrepresentación de nucleótidos G y se excluyeron.

[0091] En otras realizaciones, cada código de lectura de 5, 6, 7, 8, 9, 10 o más nucleótidos puede diferir en 2, 3, 4 o 5 nucleótidos de cualquier otro código de lectura.

[0092] En una realización, el código de lectura es de aproximadamente 5 nucleótidos y difiere de cualquier otro código de lectura en 2 nucleótidos.

[0093] En realizaciones particulares, un módulo adaptador comprende un tercer elemento que comprende una o más secuencias de código de muestra. Tal como se usa en la presente descripción, el término “código de muestra” se refiere a un polinucleótido que se usa para identificar la muestra. El código de muestra también es útil para

establecer reacciones de secuenciación múltiple porque cada código de muestra es único para la muestra y, por lo tanto, se puede usar para identificar una lectura de una muestra particular dentro de una reacción de secuenciación multiplexada.

5 **[0094]** En una realización, el código de muestra comprende una secuencia que tiene aproximadamente 1, aproximadamente 2 nucleótidos, aproximadamente 3 nucleótidos, aproximadamente 4 nucleótidos o aproximadamente 5 nucleótidos, o más. En otra realización, cada código de muestra de 2, 3, 4, 5 o más nucleótidos puede diferir de cualquier otro código de muestra en 2, 3, 4, o 5 nucleótidos.

10 **[0095]** En una realización, el código de muestra tiene aproximadamente tres nucleótidos y difiere de cualquier otro código de muestra usado en otras muestras en dos nucleótidos.

15 **[0096]** En realizaciones particulares, un módulo adaptador comprende un cuarto elemento que comprende una o más secuencias de anclaje. Tal como se usa en la presente descripción, una “secuencia de anclaje” se refiere a una secuencia de nucleótidos de al menos 8 nucleótidos, al menos 10 nucleótidos, al menos 12 nucleótidos, al menos 14 nucleótidos o al menos 16 nucleótidos que se hibrida con un oligonucleótido asociado y que comprende las tres propiedades siguientes: (1) cada secuencia de anclaje es parte de una familia de cuatro secuencias de anclaje que representan colectivamente cada una de las cuatro posibles bases de ADN en cada sitio dentro de la extensión; esta característica, la representación de base equilibrada, es útil para calibrar la llamada de base adecuada en las lecturas de secuenciación en realizaciones particulares; (2) cada secuencia de anclaje está compuesta por solo dos de las cuatro bases posibles, y estas se eligen específicamente para que sean un número igual de A + C o un número igual de G + T; una secuencia de anclaje formada a partir de solo dos bases reduce la posibilidad de que la secuencia de anclaje participe en la formación de una estructura secundaria que impediría la función adaptadora adecuada; y (3) debido a que cada secuencia de anclaje está compuesta por números iguales de A+C o G+T, cada secuencia de anclaje comparte aproximadamente la misma temperatura de fusión y estabilidad dúplex que cualquier otra secuencia de anclaje en un conjunto de cuatro.

20 **[0097]** En realizaciones particulares, un módulo adaptador comprende un quinto elemento que está compuesto por los dos nucleótidos terminales 3' del elemento 4. Estas dos bases en el extremo 3' de cada anclaje se eligen basándose en una determinación empírica que muestra que estos dos nucleótidos son sustratos eficaces para la ligación al ADNlc. En realizaciones particulares, el elemento 5 comprende las secuencias seleccionadas del grupo que consiste en: AA, CC, TT y GG. En realizaciones particulares, el elemento 5 no comprende la combinación de dinucleótidos CG o TG, ya que los inventores han determinado que estas combinaciones no son sustratos de ligación eficientes.

30 **[0098]** En realizaciones particulares, una etapa de ligación comprende ligar un módulo adaptador al ADNlc reparado en los extremos para generar una genoteca de ADNlc “etiquetada”. En algunas realizaciones, se emplea un único módulo adaptador. En algunas realizaciones, se emplean dos, tres, cuatro o cinco módulos adaptadores. En algunas realizaciones, un módulo adaptador de secuencia idéntica se liga a cada extremo del ADN fragmentado con el extremo reparado.

35 **[0099]** En una realización, una pluralidad de especies adaptadoras se liga a una genoteca de ADNlc reparado en los extremos. Cada uno de la pluralidad de adaptadores puede comprender uno o más sitios de unión al cebador para la amplificación de la genoteca de ADNlc, una o más secuencias de código de lectura, una o más secuencias para la multiplexación de muestras y una o más secuencias para la secuenciación del ADN.

40 **[0100]** La ligación de uno o más adaptadores contemplados en la presente descripción puede llevarse a cabo mediante métodos conocidos por los expertos en la materia. En realizaciones particulares, uno o más adaptadores contemplados en la presente descripción se ligan al ADNlc reparado en los extremos que comprende extremos romos. En ciertas realizaciones, uno o más adaptadores contemplados en la presente descripción se ligan al ADNlc reparado en los extremos que comprende extremos complementarios apropiados para el método de ligación empleado. En ciertas realizaciones, uno o más adaptadores contemplados en la presente descripción se ligan al ADNlc reparado en los extremos que comprende un saliente de 3'.

2. Amplificación de la genoteca CFDNA

55 **[0101]** Los métodos de análisis genético contemplados en la presente descripción comprenden la amplificación de una genoteca de ADNlc para generar una genoteca de clones de ADNlc o una genoteca de clones de ADNlc. Cada molécula de la genoteca de ADNlc comprende un adaptador ligado a cada extremo de un ADNlc reparado en los extremos, y cada adaptador comprende uno o más sitios de unión al cebador de la PCR. En una realización, se ligan diferentes adaptadores a diferentes extremos del ADNlc reparado en los extremos.

60 **[0102]** En una realización preferida, el mismo adaptador se liga a ambos extremos del ADNlc. La ligación del mismo adaptador a ambos extremos del ADNlc reparado en los extremos permite la amplificación por PCR con una única secuencia de cebadores. En realizaciones particulares, una parte de la genoteca de ADNlc ligado a un adaptador se amplificará usando técnicas de PCR estándar con una única secuencia de cebadores que active la amplificación. En

una realización, la secuencia cebadora única es de aproximadamente 25 nucleótidos, opcionalmente con una T_m proyectada de $\geq 55^\circ\text{C}$ en condiciones de fuerza iónica estándar.

[0103] En realizaciones particulares, los picogramos de la genoteca inicial de ADNlc se amplifican en microgramos de clones de ADNlc, lo que implica una amplificación de 10.000 veces. La cantidad de producto amplificado se puede medir usando métodos conocidos en la técnica, *p. ej.*, la cuantificación en un instrumento Qubit 2.0 o Nanodrop.

3. Determinar el número de equivalentes del genoma

[0104] Un método para el análisis genético del ADNlc comprende determinar el número de equivalentes del genoma en la genoteca de clones de ADNlc. Tal como se usa en la presente descripción, el término “equivalente del genoma” se refiere al número de copias del genoma en cada genoteca. Un desafío importante al que se enfrentan las composiciones y métodos contemplados en la presente descripción es lograr una sensibilidad de ensayo suficiente para detectar y analizar mutaciones genéticas raras o diferencias en la secuencia genética. Para determinar el valor de sensibilidad del ensayo muestra por muestra, se miden los números de secuencias diferentes y distintas que están presentes en cada muestra, midiendo el número de equivalentes genómicos que están presentes en una genoteca de secuenciación. Para establecer la sensibilidad, se debe medir el número de equivalentes genómicos para cada genoteca de muestras.

[0105] El número de equivalentes del genoma se puede determinar mediante un ensayo de qPCR o mediante el uso de un recuento basado en bioinformática después de realizar la secuenciación. En el flujo del proceso de muestras clínicas, la medición por qPCR de los equivalentes del genoma se usa como un paso de control de calidad para las genotecas de ADNlc. Establece una expectativa de sensibilidad del ensayo antes del análisis de la secuencia y permite excluir una muestra del análisis si su genoteca de clones de ADNlc correspondiente carece de la profundidad requerida de equivalentes genómicos. En última instancia, el recuento de los equivalentes del genoma basado en la bioinformática también se usa para identificar los equivalentes del genoma y, por lo tanto, la sensibilidad del ensayo y las estimaciones falsas negativas para cada genoteca de clones de ADNlc determinada.

[0106] El ensayo empírico de qPCR y los ensayos de recuento estadístico deben estar bien correlacionados. En los casos en que la secuenciación no revele la profundidad de la secuencia en una genoteca de clones de ADNlc, puede ser necesario el reprocesamiento de la genoteca de clones de ADNlc y/o una secuenciación adicional.

[0107] En una realización, los equivalentes del genoma en una genoteca de clones de ADNlc se determinan usando un ensayo de PCR cuantitativa (qPCR, por sus siglas en inglés). En una realización particular, se usa una genoteca estándar de concentraciones conocidas para construir una curva estándar y las mediciones del ensayo de qPCR se ajustan a la curva estándar resultante y un valor para los equivalentes del genoma se deriva del ajuste. Sorprendentemente, los presentes inventores han descubierto que un ensayo de qPCR “basado en repeticiones” que comprende un cebador que se hibrida específicamente con una secuencia común del genoma, *p. ej.*, una secuencia repetida, y otro cebador que se une al sitio de unión del cebador en el adaptador, mide un aumento de 8 veces en los equivalentes del genoma en comparación con los métodos que usan solo el cebador específico del adaptador (presente en ambos extremos del clon de ADNlc). El número de equivalentes del genoma medido mediante los ensayos basados en repeticiones proporciona un rendimiento de genoteca a genoteca más consistente y una mejor alineación entre las estimaciones de qPCR de los equivalentes del genoma y los equivalentes de etiquetas contados bioinformáticamente en las ejecuciones de secuenciación.

[0108] Los ejemplos ilustrativos de repeticiones adecuadas para su uso en los ensayos equivalentes del genoma basados en repeticiones contemplados en la presente descripción incluyen, pero no se limitan a: elementos nucleares intercalados cortos (SINE), *p. ej.*, repeticiones Alu; elementos nucleares intercalados de forma prolongada (los LINE), *p. ej.*, LINE1, LINE2, LINE3; elementos de repetición de microsatélites, *p. ej.*, repeticiones cortas en tándem (STR), repeticiones de secuencia simple (SSR); y repeticiones intercaladas (miR) en todo mamífero.

[0109] En una realización, la repetición es una repetición Alu.

4. Análisis genético cuantitativo

[0110] Un método para el análisis genético del ADNlc comprende el análisis genético cuantitativo de uno o más loci genéticos diana de los clones de la genoteca de ADNlc. El análisis genético cuantitativo comprende uno o más de, o todas, las siguientes etapas: capturar clones de ADNlc que comprenden un locus genético diana; amplificación del locus genético diana capturado; secuenciación del locus genético diana capturado amplificado; y análisis bioinformático de las lecturas de secuencia resultantes.

(a) Captura del locus genético diana

[0111] La presente invención contempla, en parte, un módulo de sonda de captura diseñado para conservar la eficiencia y la fiabilidad de las sondas más grandes, pero que minimiza la generación de secuencias no informativas en una genoteca de clones de ADNlc. Un “módulo de sonda de captura” se refiere a un polinucleótido que comprende

una secuencia de sonda de captura y una secuencia de cola. En realizaciones particulares, la secuencia del módulo de la sonda de captura o una parte de la misma sirve como sitio de unión al cebador para uno o más cebadores de secuenciación.

[0112] En realizaciones particulares, un módulo de sonda de captura comprende una sonda de captura. Como se usa en la presente descripción, una “sonda de captura” se refiere a una región capaz de hibridarse con una región diana de ADN específica. Debido a que el tamaño promedio del ADNlc es de aproximadamente 150 a aproximadamente 170 pares de bases y está altamente fragmentado, las composiciones y métodos contemplados en la presente descripción comprenden el uso de sondas de captura de alta densidad y relativamente cortas para interrogar las regiones diana del ADN de interés.

[0113] Una preocupación particular con el uso de sondas de captura de alta densidad es que, generalmente, las sondas de captura se diseñan usando “reglas de secuencia” específicas. Por ejemplo, las regiones de secuencia redundante o que muestran sesgos extremos en la composición de bases generalmente se excluyen al diseñar las sondas de captura. Sin embargo, los presentes inventores han descubierto que la falta de flexibilidad en las reglas de diseño de la sonda de captura no afecta sustancialmente al rendimiento de la sonda. Por el contrario, las sondas de captura elegidas estrictamente por restricción posicional proporcionaron información sobre la secuencia diana; muestran muy poca captura de lectura fuera del diana y no se puede mapear; y producen lecturas uniformes, útiles y acordes con el diana, con solo unas pocas excepciones. Además, la alta redundancia con un espaciado estrecho entre las sondas compensa con creces las sondas de captura ocasionales de bajo rendimiento.

[0114] En realizaciones particulares, una región diana es el dirigida mediante una pluralidad de sondas de captura, en donde dos o más sondas de captura se diseñan para unirse a la región diana dentro de 10 nucleótidos entre sí, dentro de 15 nucleótidos entre sí, dentro de 25 nucleótidos entre sí, dentro de 30 nucleótidos entre sí, dentro de 35 nucleótidos entre sí, dentro de 45 nucleótidos entre sí, o dentro de 50 nucleótidos o más entre sí, así como todas las longitudes de nucleótidos intermedias.

[0115] En una realización, la sonda de captura tiene aproximadamente 25 nucleótidos, aproximadamente 26 nucleótidos, aproximadamente 27 nucleótidos, aproximadamente 28 nucleótidos, aproximadamente 29 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 31 nucleótidos, aproximadamente 32 nucleótidos, aproximadamente 33 nucleótidos, aproximadamente 34 nucleótidos, aproximadamente 35 nucleótidos, aproximadamente 36 nucleótidos, aproximadamente 37 nucleótidos, aproximadamente 38 nucleótidos, aproximadamente 39 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 41 nucleótidos, aproximadamente 42 nucleótidos, aproximadamente 43 nucleótidos, aproximadamente 44 nucleótidos, o aproximadamente 45 nucleótidos.

[0116] En una realización particular, la sonda de captura no tiene 60 nucleótidos.

[0117] En otra realización, la sonda de captura es sustancialmente más pequeña que 60 nucleótidos, pero se hibrida de manera comparable, tan bien, o mejor que una sonda de captura de 60 nucleótidos dirigida a la misma región diana de ADN.

[0118] En una cierta realización, la sonda de captura tiene 40 nucleótidos.

[0119] En ciertas realizaciones, un módulo de sonda de captura comprende una secuencia de cola. Tal como se usa en la presente descripción, el término “secuencia de cola” se refiere a un polinucleótido en el extremo 5' del módulo de la sonda de captura, que en algunas realizaciones particulares puede servir como sitio de unión al cebador. En realizaciones particulares, un cebador de secuenciación se une al sitio de unión del cebador en la región de la cola.

[0120] En realizaciones particulares, la secuencia de cola es de aproximadamente 5 a aproximadamente 100 nucleótidos, de aproximadamente 10 a aproximadamente 100 nucleótidos, de aproximadamente 5 a aproximadamente 75 nucleótidos, de aproximadamente 5 a aproximadamente 50 nucleótidos, de aproximadamente 5 a aproximadamente 25 nucleótidos, o de aproximadamente 5 a aproximadamente 20 nucleótidos. En ciertas realizaciones, la tercera región es de aproximadamente 10 a aproximadamente 50 nucleótidos, de aproximadamente 15 a aproximadamente 40 nucleótidos, de aproximadamente 20 a aproximadamente 30 nucleótidos o aproximadamente 20 nucleótidos, o cualquier número intermedio de nucleótidos.

[0121] En realizaciones particulares, la secuencia de cola es de aproximadamente 30 nucleótidos, aproximadamente 31 nucleótidos, aproximadamente 32 nucleótidos, aproximadamente 33 nucleótidos, aproximadamente 34 nucleótidos, aproximadamente 35 nucleótidos, aproximadamente 36 nucleótidos, aproximadamente 37 nucleótidos, aproximadamente 38 nucleótidos, aproximadamente 39 nucleótidos o aproximadamente 40 nucleótidos.

[0122] En diversas realizaciones, el módulo de sonda de captura comprende un miembro específico de un par de unión para permitir el aislamiento y/o la purificación de uno o más fragmentos capturados de una genoteca de ADNlc etiquetada y/o amplificada que se hibrida con la sonda de captura. En realizaciones particulares, el módulo de sonda de captura se conjuga con biotina u otro hapteno adecuado, p. ej., dinitrofenol, digoxigenina.

[0123] En diversas realizaciones, el módulo de sonda de captura se hibrida con una genoteca de ADNlc etiquetada y opcionalmente amplificada para formar un complejo. En algunas realizaciones, el módulo de sonda de captura multifuncional se hibrida sustancialmente con una región diana genómica específica en la genoteca de ADNlc.

[0124] Las condiciones de hibridación o hibridación pueden incluir cualquier condición de reacción en la que dos secuencias de nucleótidos formen un complejo estable; por ejemplo, la genoteca de ADNlc etiquetada y el módulo de sonda de captura forman un complejo estable de genoteca de ADNlc etiquetado y módulo de sonda de captura. Dichas condiciones de reacción son bien conocidas en la técnica y los expertos en la materia apreciarán que tales condiciones pueden modificarse según sea apropiado, *p. ej. p. ej.*, temperaturas de recocido reducidas con sondas de captura de menor longitud, y están dentro del alcance de la presente invención. Puede producirse una hibridación sustancial cuando la segunda región del complejo de sondas de captura presenta una identidad de secuencia, homología o complementariedad del 100 %, 99 %, 98 %, 97 %, 96 %, 95 %, 94 %, 93 %, 92 %, 91 %, 90 %, 89 %, 88 %, 85 %, 80 %, 75 % o 70 % de identidad de secuencia, homología o complementariedad con una región de la genoteca de ADNlc etiquetada.

[0125] En realizaciones particulares, la sonda de captura tiene aproximadamente 40 nucleótidos y tiene una temperatura de hibridación óptima de aproximadamente 44 °C a aproximadamente 47 °C.

[0126] En ciertas realizaciones, los métodos contemplados en la presente descripción comprenden aislar un complejo de módulos de sondas de captura y genoteca de ADNlc etiquetado. En realizaciones particulares, los expertos en la materia conocen bien los métodos para aislar complejos de ADN y cualquier método que un experto en la materia considere apropiado puede emplearse con los métodos de la presente invención (Ausubel y col., *Current Protocols in Molecular Biology*, 2007-2012). En realizaciones particulares, los complejos se aíslan usando técnicas de aislamiento de biotina-estreptavidina.

[0127] En realizaciones particulares, se contempla la eliminación de los extremos 3' monocatenarios del complejo módulo de sondas de captura y genoteca de ADNlc etiquetado aislado. En ciertas realizaciones, los métodos comprenden el procesamiento enzimático de la exonucleasa 3'-5' del complejo modular de sonda de captura multifuncional y genoteca de ADN marcado aislado para eliminar los extremos 3' monocatenarios.

[0128] En ciertas otras realizaciones, los métodos comprenden realizar la extensión 5'-3' de la ADN polimerasa de una sonda de captura multifuncional usando los fragmentos de la genoteca de ADN etiquetados aislados como plantilla.

[0129] En ciertas otras realizaciones, los métodos comprenden la creación de una molécula diana de ADNlc etiquetada híbrida aislada por una sonda de captura mediante la acción concertada de una endonucleasa 5' FLAP, la polimerización del ADN y el cierre por muesca mediante una ligasa de ADN.

[0130] Se pueden emplear una variedad de enzimas para el procesamiento enzimático de la exonucleasa 3'-5' del complejo modular de sonda de captura multifuncional de genoteca de ADNlc etiquetado aislado. Los ejemplos ilustrativos de enzimas adecuadas, que muestran actividad enzimática de la exonucleasa 3'-5', que pueden emplearse en realizaciones particulares incluyen, pero no se limitan a: T4 o exonucleasas I, III, V (véase también *Shevelev IV, Hübscher U.*, "The 3' 5' exonucleases", *Nat Rev Mol Cell Biol.* 5:364-76 (2002)). En realizaciones particulares, la enzima que comprende la actividad de la exonucleasa 3'-5' es la polimerasa T4. En realizaciones particulares, puede emplearse una enzima que muestre actividad enzimática de exonucleasa 3'-5' y que sea capaz de extender el molde del cebador, incluyendo, por ejemplo, la T4 o las exonucleasas I, III, V. Id.

[0131] En algunas realizaciones, los métodos contemplados en la presente descripción comprenden realizar la secuenciación y/o la PCR en el complejo procesado enzimáticamente de la exonucleasa 3'-5' discutido anteriormente y en otras partes de la *presente* descripción. En realizaciones particulares, se copia una parte de la cola de una molécula sonda de captura para generar una molécula de ácido nucleico híbrida. En una realización, la molécula de ácido nucleico híbrida generada comprende la región diana capaz de hibridarse con el módulo de la sonda de captura y el complemento de la secuencia de cola del módulo de la sonda de captura.

[0132] En una realización particular, el análisis genético comprende a) hibridar uno o más módulos de sondas de captura con uno o más loci genéticos diana en una pluralidad de clones de genotecas de ADNlc para formar uno o más complejos de clones de módulos de sondas de captura y genotecas de ADNlc; b) aislar uno o más complejos de clones de módulo de sonda de captura y genoteca de ADNlc de a); c) procesar enzimáticamente uno o más complejos aislados de clones de módulo de sonda de captura y genoteca de ADNlc de la etapa b); d) realizar una PCR en el complejo procesado enzimáticamente de c) en donde la porción de cola de la molécula de la sonda de captura se copia para generar moléculas de ácido nucleico híbridadas amplificadas, en donde las moléculas de ácido nucleico híbridadas amplificadas comprenden una secuencia diana en el locus genómico diana capaz de hibridarse con la sonda de captura y el complemento de la secuencia de cola del módulo de la sonda de captura; y e) realizar un análisis genético cuantitativo en las moléculas de ácido nucleico híbridadas amplificadas de d).

[0133] En una realización particular, se contemplan métodos para determinar el número de copias de un locus genético diana específico que comprenden: a) hibridar uno o más módulos de sondas de captura con uno o más loci genéticos diana en una pluralidad de clones de la genoteca de ADNlc para formar uno o más complejos de módulos de sonda de captura y clon de la genoteca de ADNlc; b) aislar uno o más complejos de clones de módulo de sonda de captura y genoteca de ADNlc de a); c) procesar enzimáticamente uno o más complejos aislados de clones de módulo de sonda de captura y genoteca de ADNlc de la etapa b); d) realizar una PCR en el complejo procesado enzimáticamente de c) en donde la porción de cola de la molécula de la sonda de captura se copia para generar moléculas de ácido nucleico híbridas amplificadas, en donde las moléculas de ácido nucleico híbridas amplificadas comprenden una secuencia diana en el locus genético diana capaz de hibridarse con la sonda de captura y el complemento de la secuencia de cola del módulo de la sonda de captura; e) realizar la amplificación por PCR de las moléculas de ácido nucleico híbridas amplificadas en d); y f) cuantificar la reacción de PCR en e), donde la cuantificación permite determinar el número de copias de la región diana específica.

[0134] En una realización, el procesamiento enzimático de la etapa c) comprende realizar el procesamiento enzimático de la exonucleasa 3'-5' en uno o más complejos de clones del módulo de sonda de captura y la genoteca de ADNlc de b) usando una enzima con actividad de exonucleasa 3'-5' para eliminar los extremos 3' monocatenarios; crear una o más moléculas clonadas híbridas de módulo de sonda de captura y genoteca de ADNlc mediante la acción concertada de una endonucleasa FLAP 5', la polimerización del ADN y el cierre por muesca mediante una ligasa de ADN; o realizar una extensión de ADN polimerasa en 5'-3' de la sonda de captura usando el clon de ADNlc aislado en el complejo como plantilla.

[0135] En una realización, el procesamiento enzimático de la etapa c) comprende realizar una extensión 5'-3' con ADN polimerasa de la sonda de captura usando el clon de ADNlc aislado en el complejo como plantilla.

[0136] En realizaciones particulares, la PCR se puede realizar usando cualquier condición de reacción de PCR estándar bien conocida por los expertos en la materia. En ciertas realizaciones, la reacción de PCR en e) emplea dos cebadores de PCR. En una realización, la reacción de PCR en e) emplea un primer cebador de PCR que se hibrida con una repetición dentro del locus genético diana. En una realización particular, la reacción de PCR en e) emplea un segundo cebador de PCR que se hibrida con las moléculas de ácido nucleico híbridas en la unión entre el locus y la cola del gen diana. En ciertas realizaciones, la reacción de PCR en e) emplea un primer cebador de PCR que se hibrida con el locus genético diana y un segundo cebador de PCR se hibrida con las moléculas de ácido nucleico híbridas amplificadas en la unión entre el locus genético diana y la cola. En realizaciones particulares, el segundo cebador se hibrida con la unión entre el locus genético diana y la cola de modo que al menos uno o más nucleótidos del cebador se hibridan con el locus genético diana y al menos uno o más nucleótidos del cebador se hibridan con la secuencia de cola.

[0137] En ciertas realizaciones, las moléculas de ácido nucleico híbridas amplificadas obtenidas en la etapa e) se secuencian y las secuencias se alinean horizontalmente, es decir, se alinean entre sí pero no se alinean con una secuencia de referencia. En realizaciones particulares, las etapas a) a e) se repiten una o más veces con uno o más módulos de sonda de captura. Los módulos de la sonda de captura pueden ser iguales o diferentes y estar diseñados para dirigirse a cualquier cadena de ADNlc de un locus genético diana. En algunas realizaciones, cuando las sondas de captura son diferentes, se hibridan en secuencias diana superpuestas o adyacentes dentro de un locus genético diana en la genoteca de clones de ADNlc etiquetados. En una realización, se usa una estrategia de sonda de captura de alta densidad en donde una pluralidad de sondas de captura se hibrida con un locus genético diana, y en donde cada una de la pluralidad de sondas de captura se hibrida con el locus genético diana dentro de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 puntos de referencia de cualquier otra sonda de captura que hibride con el locus genético diana en una genoteca de clones de ADNlc etiquetados, incluidas todas las distancias intermedias.

[0138] En algunas realizaciones, el método puede realizarse usando dos módulos de sonda de captura por locus genético diana, en donde uno se hibrida con la cadena "Watson" (cadena no codificante o plantilla) corriente arriba de la región diana y otro se hibrida con la cadena "Crick" (cadena codificante o no plantilla) corriente abajo de la región diana.

[0139] En realizaciones particulares, los métodos contemplados en la presente descripción pueden realizarse además varias veces con cualquier número de módulos de sonda de captura, por ejemplo 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más módulos de sonda de captura por locus genético diana, cualquier número de los cuales se hibrida con la cadena de Watson o Crick en cualquier combinación. En algunas realizaciones, las secuencias obtenidas pueden alinearse entre sí para identificar cualquiera de una serie de diferencias.

[0140] En ciertas realizaciones, se interroga una pluralidad de loci genéticos diana, p. ej., 1100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 10.000, 50.000, 100.000, 500.000 o más en una sola reacción, usando uno o más módulos de sonda de captura.

(b) Secuenciación

[0141] El análisis genético cuantitativo comprende secuenciar una pluralidad de moléculas de ácido nucleico híbridas, como se ha descrito en otra parte del presente documento, anteriormente, para generar profundidades de secuenciación suficientes para obtener una pluralidad de lecturas de secuenciación únicas. Una lectura única se define como la única lectura de consenso de una “familia” de lecturas que comparten el mismo código de lectura y el mismo punto de inicio de secuencia dentro de ADNlc. Cada sonda de captura produce un conjunto de lecturas únicas que se destilan computacionalmente del total de lecturas agrupándolas en familias. Las lecturas únicas para una muestra determinada se calculan después como el promedio de todas las lecturas únicas observadas sonda por sonda. Los casos en los que hay un cambio evidente en el número de copias se excluyen del conjunto de datos usado para calcular el promedio. Las lecturas únicas son importantes porque cada lectura única debe derivarse de un clon de ADNlc único. Cada lectura única representa la entrada y el análisis de un equivalente haploide del ADN genómico. La suma de las lecturas únicas es la suma de los genomas haploides analizados. El número de genomas analizados, a su vez, define la sensibilidad del ensayo de secuenciación. A manera de ejemplo no limitativo, si el recuento promedio de lecturas únicas es de 100 equivalentes de genoma, entonces ese ensayo en particular tiene la sensibilidad de poder detectar una lectura mutante en 100, o el 1 %. Cualquier observación inferior a esta no es defendible.

[0142] En realizaciones particulares, el análisis genético cuantitativo comprende la secuenciación múltiple de moléculas de ácido nucleico híbridas derivadas de una pluralidad de muestras.

[0143] En diversas realizaciones, el análisis genético cuantitativo comprende obtener uno o más o una pluralidad de clones de genotecas de ADN etiquetados, comprendiendo cada clon una primera secuencia de ADN y una segunda secuencia de ADN, en donde la primera secuencia de ADN comprende una secuencia en un locus genético diana y la segunda secuencia de ADN comprende una secuencia de sonda de captura; realizar una reacción de secuenciación de extremos pareados en uno o más clones y obtener una o más lecturas de secuenciación o realizar una reacción de secuenciación en uno o más clones en donde se obtiene una única lectura de secuenciación larga de más de aproximadamente 100, 200, 300, 400, 500 o más nucleótidos, en donde la lectura es suficiente para identificar tanto la primera secuencia de ADN como la segunda secuencia de ADN; y ordenar o agrupar las lecturas de secuenciación de uno o más clones de acuerdo con las secuencias de sonda de las lecturas de secuenciación.

(c) Análisis bioinformático

[0144] El análisis genético cuantitativo comprende además el análisis bioinformático de las lecturas de secuenciación. El análisis bioinformático excluye cualquier análisis puramente mental realizado en ausencia de una composición o método de secuenciación. En ciertas realizaciones, el análisis bioinformático incluye, pero no se limita a: alineamientos de secuencias; análisis de equivalentes genómicos; análisis de variantes de un solo nucleótido (SNV); análisis de la variación del número de copias de genes (CNV); y detección de lesiones genéticas. En realizaciones particulares, el análisis bioinformático es útil para cuantificar el número de equivalentes genómicos analizados en la genoteca de clones de ADNlc; para detectar el estado genético de un locus genético diana; para detectar lesiones genéticas en un locus genético diana; y para medir las fluctuaciones del número de copias dentro de un locus genético diana.

[0145] Los alineamientos de secuencia se pueden realizar entre las lecturas de secuencia y una o más secuencias de ADN de referencia humanas. En realizaciones particulares, las alineaciones de secuenciación pueden usarse para detectar lesiones genéticas en un locus genético diana que incluyen, pero no se limitan a, la detección de una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias, o una fusión de genes. La detección de lesiones genéticas que son indicadores causales o pronósticos puede ser útil en el diagnóstico, el pronóstico, el tratamiento y/o la monitorización de una afección o enfermedad genética en particular.

[0146] También se contemplan en la presente descripción métodos para el análisis de alineamiento de secuencias que pueden realizarse sin la necesidad de alinearse con una secuencia de referencia, denominados en la presente descripción análisis de secuencia horizontal. Dicho análisis se puede realizar en cualquier secuencia generada por los métodos contemplados en la presente descripción o en cualquier otro método. En realizaciones particulares, el análisis de secuencias comprende realizar alineamientos de secuencia en las lecturas obtenidas mediante los métodos contemplados en la presente descripción.

[0147] En una realización, los equivalentes del genoma en una genoteca de clones de ADNlc se determinan mediante el recuento basado en bioinformática después de realizar la secuenciación. Cada lectura de secuenciación está asociada a una sonda de captura particular, y la colección de lecturas asignadas a cada sonda de captura se analiza en grupos. Dentro de un grupo, los conjuntos de lecturas individuales comparten el mismo código de lectura y la misma posición de inicio de la secuencia de ADN dentro de la secuencia genómica. Estas lecturas individuales se agrupan en una “familia” y una única representación consensuada de esta familia se presenta como una “lectura única”. Todas las lecturas individuales que constituían una familia se derivan de un solo evento de ligación y, por lo tanto, son “hermanos” derivados de la amplificación entre sí. Cada lectura única se considera un evento de ligación único y la suma de las lecturas únicas se considera equivalente al número de equivalentes del genoma analizados.

[0148] A medida que el número de clones únicos se acerca al número total de posibles combinaciones de secuencias, la probabilidad dicta que las mismas combinaciones de código y sitio de inicio se creen mediante eventos independientes y que estos eventos independientes se agrupen de manera inapropiada en familias individuales. El resultado neto será una subestimación de los equivalentes genómicos analizados, y las lecturas de mutantes raros pueden descartarse como errores de secuenciación porque se superponen con las lecturas silvestres que contienen los mismos identificadores.

[0149] En realizaciones particulares, para proporcionar un análisis preciso de las genotecas de clones de ADNlc, el número de equivalentes genómicos analizados es de aproximadamente 1/10, aproximadamente 1/12, aproximadamente 1/14, aproximadamente 1/16, aproximadamente 1/18, aproximadamente 1/20, aproximadamente 1/25 o menos del número de posibles clones únicos. Debe entenderse que el procedimiento descrito anteriormente es meramente ilustrativo y no limitativo.

[0150] En algunas realizaciones, puede ser necesario aumentar el número de equivalentes genómicos que van a analizarse. Para ampliar la profundidad de los equivalentes del genoma, se contemplan al menos dos soluciones. La primera solución es usar más de un juego de adaptadores por muestra. Al combinar adaptadores, es posible ampliar multiplicativamente el número total de clones posibles y, por lo tanto, ampliar los cómodos límites de la entrada genómica. La segunda solución es expandir el código leído en 1, 2, 3, 4 o 5 o más bases. El número de códigos de lectura posibles que difieren en al menos 2 bases de cualquier otro código leído se escala como $4^{(n-1)}$, donde n es el número de bases dentro de un código leído. Por lo tanto, en un ejemplo no limitativo, si un código leído es de 5 nucleótidos y $4^{(5-1)} = 256$; por lo tanto, la inclusión de bases adicionales amplía el repertorio disponible en un factor de cuatro por cada base adicional.

[0151] En una realización, el análisis genético cuantitativo comprende el análisis bioinformático de las lecturas de secuenciación para identificar variantes raras de un único nucleótido (SNV, por sus siglas en inglés).

[0152] La secuenciación de próxima generación tiene una tasa de error inherente de aproximadamente el 0,02 al 0,02 %, lo que significa que entre 1/200 y 1/500 llamadas base son incorrectas. Para detectar variantes y otras mutaciones que se producen a frecuencias inferiores a esta, por ejemplo, a frecuencias de 1 por 1000 secuencias, es necesario invocar estrategias de anotación molecular. A manera de ejemplo no limitativo, el análisis de 5000 moléculas únicas mediante tecnología de captura de secuencias específicas generaría, a profundidades de secuenciación suficientes de >50.000 lecturas, una colección de 5000 lecturas únicas, y cada lectura única pertenecería a una "familia" de lecturas que poseen todas el mismo código de lectura. Un SNV que se presenta dentro de una familia es candidato a ser una variante poco frecuente. Cuando esta misma variante se observa en más de una familia, se convierte en una candidata muy fuerte por ser una variante poco frecuente que existe dentro de la muestra inicial. Por el contrario, es probable que las variantes que se producen esporádicamente dentro de las familias sean errores de secuenciación y las variantes que se producen dentro de una y solo una familia son poco frecuentes o son el resultado de una alteración de la base que se produjo *ex vivo* (p. ej., la oxidación de una base de ADN o errores introducidos por la PCR).

[0153] En una realización, los métodos de detección de las SNV comprenden introducir 10 veces más información genómica (genomas o equivalentes del genoma) como la sensibilidad diana deseada del ensayo. En un ejemplo no limitativo, si la sensibilidad deseada es del 2 % (2 en 100), entonces el diana experimental es una entrada de 2000 genomas.

[0154] En realizaciones particulares, el análisis bioinformático de los datos de secuenciación se usa para detectar o identificar la SNV asociado con un estado, afección o enfermedad genética, el mosaicismo genético, las pruebas fetales, las pruebas de paternidad, la predicción de la respuesta al tratamiento farmacológico, el diagnóstico o la monitorización de una afección médica, la elaboración de perfiles del microbioma, la detección de patógenos, y la monitorización de los trasplantes de órganos.

[0155] En diversas realizaciones, se proporciona un método para el análisis de determinación del número de copias que comprende obtener uno o más o una pluralidad de clones, comprendiendo cada clon una primera secuencia de ADN y una segunda secuencia de ADN, en donde la primera secuencia de ADN comprende una secuencia en un locus genético diana y la segunda secuencia de ADN comprende una secuencia de sonda de captura. En realizaciones relacionadas, se lleva a cabo una reacción de secuenciación de extremos pareados en el uno o más clones y se obtienen una o más lecturas de secuenciación. En otra realización, se lleva a cabo una reacción de secuenciación en el uno o más clones en donde se obtiene una única lectura de secuenciación larga de más de aproximadamente 100 nucleótidos, en donde la lectura es suficiente para identificar tanto la primera secuencia de ADN como la segunda secuencia de ADN. Las lecturas de secuenciación de uno o más clones pueden ordenarse o agruparse de acuerdo con la secuencia de sonda de las lecturas de secuenciación.

[0156] Los análisis del número de copias incluyen, pero no se limitan a, los análisis que examinan el número de copias de un gen o mutación en particular que se produce en una muestra de ADN genómico determinada y pueden incluir además la determinación cuantitativa del número de copias de un gen determinado o de las diferencias de secuencia en una muestra determinada. En realizaciones particulares, el análisis del número de copias se usa para

detectar o identificar la amplificación génica asociada con estados, afecciones o enfermedades genéticas, las pruebas fetales, el mosaicismo genético, las pruebas de paternidad, la predicción de la respuesta al tratamiento farmacológico, el diagnóstico o la monitorización de una afección médica, la elaboración de perfiles del microbioma, la detección de patógenos y la monitorización de los trasplantes de órganos.

[0157] En realizaciones particulares, el análisis bioinformático de los datos de secuenciación se usa para detectar o identificar una o más secuencias o lesiones genéticas en un locus diana que incluye, pero no se limita a, la detección de una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias o una fusión de genes. La detección de lesiones genéticas que son indicadores causales o pronósticos puede ser útil en el diagnóstico, el pronóstico, el tratamiento y/o la monitorización de una afección o enfermedad genética en particular. En una realización, las lesiones genéticas están asociadas con estados, afecciones o enfermedades genéticas, las pruebas fetales, el mosaicismo genético, las pruebas de paternidad, la predicción de la respuesta al tratamiento farmacológico, el diagnóstico o la monitorización de una afección médica, la elaboración de perfiles del microbioma, la detección de patógenos y la monitorización de los trasplantes de órganos.

D. Aplicaciones clínicas del análisis genético cuantitativo

[0158] Solo con fines ilustrativos y no mencionados en el conjunto de reivindicaciones, se proporciona un método para detectar, identificar, predecir, diagnosticar o controlar una afección o enfermedad en un sujeto.

[0159] En realizaciones particulares, un método para detectar, identificar, predecir, diagnosticar o monitorizar un estado, afección o enfermedad genéticos en un sujeto comprende realizar un análisis genético cuantitativo de uno o más loci genéticos diana en una genoteca de clones de ADNlc para detectar o identificar un cambio en la secuencia en uno o más loci genéticos diana.

[0160] En una realización, un método para detectar, identificar, predecir, diagnosticar o controlar un estado, afección o enfermedad genéticos comprende aislar u obtener ADNlc a partir de una muestra biológica de un sujeto; tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar una genoteca de clones de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci genéticos diana en una genoteca de clones de ADNlc para detectar o identificar un cambio en la secuencia en uno o más loci genéticos diana.

[0161] En realizaciones particulares, un método para detectar, identificar, predecir, diagnosticar o controlar un estado genético o una afección o enfermedad genética seleccionada del grupo que consiste en: enfermedades genéticas; mosaicismo genético; pruebas fetales; pruebas de paternidad; pruebas de paternidad; predecir la respuesta al tratamiento farmacológico; diagnosticar o controlar una afección médica; elaboración de perfiles de microbiomas; detección de patógenos; y la monitorización del trasplante de órganos que comprende aislar u obtener ADNlc de una muestra biológica de un sujeto; tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar una genoteca de clones de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más loci genéticos diana en una genoteca de clones de ADNlc para detectar o identificar una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias o una fusión génica en la secuencia en uno o más loci genéticos diana.

[0162] Los ejemplos ilustrativos de enfermedades genéticas que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, entre otros, el cáncer, la enfermedad de Alzheimer (APOE1), la enfermedad de Charcot-Marie-Tooth, la neuropatía óptica hereditaria de Leber (LHON), el síndrome de Angelman (UBE3A, ubiquitina-proteína ligasa E3A), el síndrome de Prader-Willi (región del cromosoma 15), β -Talasemia (HBB, β -globina), enfermedad de Gaucher (tipo I) (GBA, glucocerebrosidasa), fibrosis quística (canal de cloruro epitelial CFTR), anemia drepanocítica (HBB, β -globina), enfermedad de Tay-Sachs (HEXA, hexosaminidasa A), fenilcetonuria (HAP, fenilalanina hidrolasa), hipercolesterolemia familiar (LDLR, receptor de lipoproteínas de baja densidad), poliquistosis renal en adultos (PKD1, policistina), enfermedad de Huntington (HDD, huntingtina), neurofibromatosis tipo I (NF1, gen supresor de tumores NF1), distrofia miotónica (DM, miotonina), esclerosis tuberosa (TSC1, tuberina), acondroplasia (FGFR3, receptor del factor de crecimiento de fibroblastos), síndrome del cromosoma X frágil (FMR1, proteína de unión al ARN), distrofia muscular de Duchenne (DMD, distrofina), hemofilia A (F8C, factor VIII de la coagulación sanguínea), síndrome de Lesch-Nyhan (HPRT1, hipoxantina guanina ribosiltransferasa 1) y adrenoleucodistrofia (ABCD1).

[0163] Los ejemplos ilustrativos de cánceres que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: cáncer de células B, p. ej., mieloma múltiple, melanomas, cáncer de mama, cáncer de pulmón (como el carcinoma pulmonar de células no pequeñas o NSCLC), cáncer de bronquios, cáncer colorrectal, cáncer de próstata, cáncer de

páncreas, cáncer de estómago, cáncer de ovario, cáncer de vejiga urinaria, cáncer del cerebro o del sistema nervioso central, cáncer del sistema nervioso periférico, cáncer de esófago, cáncer de cuello uterino o endometrial, cáncer de cavidad oral o faringe, cáncer de hígado, cáncer de riñón, cáncer testicular, cáncer de vías biliares, cáncer de intestino delgado o apéndice, cáncer de glándulas salivales, tiroides, cáncer de glándulas, cáncer de glándulas suprarrenales, osteosarcoma, condrosarcoma, cáncer de tejidos hematológicos, adenocarcinomas, tumores miofibroblásticos inflamatorios, tumor del estroma gastrointestinal (GIST), cáncer de colon, mieloma múltiple (MM), síndrome mielodisplásico (SMD), trastorno mieloproliferativo (MPD), leucemia linfocítica aguda (LLA), leucemia mielocítica aguda (LMA), crónica leucemia mielocítica (LMC), leucemia linfocítica crónica (LLC), policitemia vera, linfoma de Hodgkin, linfoma no hodgkiniano (LNH), sarcoma de tejidos blandos, fibrosarcoma, mixosarcoma, liposarcoma, osteogénico sarcoma, cordoma, angiosarcoma, endoteliosarcoma, linfangiosarcoma, linfangioendoteliosarcoma, sinovioma, mesotelioma, tumor de Ewing, leiomiomasarcoma, rhabdomyosarcoma, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma de glándulas sebáceas, carcinoma papilar, adenocarcinomas medulares carcinoma, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de vías biliares, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, carcinoma de vejiga, carcinoma epitelial, glioma, astrocitoma, meduloblastoma, craneofaringioma, ependimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, linfoma folicular, linfoma difuso de células B grandes, linfoma de células del manto, carcinoma hepatocelular, cáncer de tiroides, cáncer gástrico, cáncer de cabeza y cuello, cánceres de células pequeñas, trombocitemia esencial, metaplasia mieloide agnógena, síndrome hipereosinofílico, mastocitosis sistémica, hipereosinofilia familiar, leucemia eosinofílica crónica, cánceres neuroendocrinos, tumores carcinoides y lo similar.

[0164] En una realización, la lesión genética es una lesión anotada en la base de datos Cosmic (las lesiones y los datos de secuencia se pueden descargar de cancer.sanger.ac.uk/cosmic/census) o una lesión anotada en el Atlas del genoma del cáncer (las lesiones y los datos de secuencia se pueden descargar de tcga-data.nci.nih.gov/tcga/TCGAdownload.jsp).

[0165] Los ejemplos ilustrativos de genes que albergan una o más lesiones genéticas asociadas con el cáncer que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a, ABCB1, ABCC2, ABCC4, ABCG2, ABL1, ABL2, AKT1, AKT2, AKT3, ALDH4A1, ALK, APC, AR, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRAF, BRCA1, BRCA2, C1orf144, CARD11, CBL, CCND1, CCND2, CCND3, CCNE1, CDH1, CDH2, CDH20, CDH5, CDK4, CDK6, CDK8, CDKN2A, CDKN2B, CDKN2C, CEBPA, CHEK1, CHEK2, CRKL, CRLF2, CTNNB1, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DNMT3A, DOT1L, DPYD, EGFR, EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, EPHX1, ERBB2, ERBB3, ERBB4, ERCC2, ERG, ESR1, ESR2, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FCGR3A, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, FOXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GSTP1, GUCY1A2, HOXA3, HRAS, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, ITPA, JAK1, JAK2, JAK3, JUN, KDR, KIT, KRAS, LRP1B, LRP2, LTK, MAN1B1, MAP2K1, MAP2K2, MAP2K4, MCL1, MDM2, MDM4, MEN1, MET, MITF, MLH1, MLL, MPL, MRE11A, MSH2, MSH6, MTHFR, MTOR, MUTYH, MYC, MYCL1, MYCN, NF1, NF2, NKX2-1, NOTCH1, NPM1, NQO1, NRAS, NRP2, NTRK1, NTRK3, PAK3, PAX5, PDGFRA, PDGFRB, PIK3CA, PIK3R1, PKHD1, PLCG1, PRKDC, PTCH1, PTEN, PTPN11, PTPRD, RAF1, RARA, RB1, RET, RICTOR, RPTOR, RUNX1, SLC19A1, SLC22A2, SLC01B3, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMO, SOD2, SOX10, SOX2, SRC, STK11, SULT1A1, TBX22, TET2, TGFB2, TMRSS2, TNFRSF14, TOP1, TP53, TPMT, TSC1, TSC2, TYMS, UGT1A1, UMPS, USP9X, VHL, y WT1.

[0166] En realizaciones particulares, la lesión genética comprende una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, un cambio en el número de copias, o una fusión de genes.

[0167] En una realización, la lesión genética es una fusión génica que fusiona la región codificante 3' del gen ALK con otro gen.

[0168] En una realización, la lesión genética es una fusión génica que fusiona la región codificante 3' del gen ALK con el gen EML4.

[0169] Los ejemplos ilustrativos de afecciones adecuadas para las pruebas fetales que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: Síndrome de Down (trisomía 21), síndrome de Edwards (trisomía 18), síndrome de Patau (trisomía 13), síndrome de Klinefelter (XXY), síndrome de Triple X, síndrome XYY, trisomía 8, trisomía 16, síndrome de Turner (XO), translocación robertsoniana, síndrome de DiGeorge y síndrome de Wolf-Hirschhorn.

[0170] Los ejemplos ilustrativos de alelos adecuados para las pruebas de paternidad que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a, 16 o más de: D20S1082, D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D4S2364, D9S1122, D2S1776, D10S1425, D3S3053, D5S2500, D1S1627,

D3S4529, D2S441, D17S974, D6S1017, D4S2408, D9S2157, amelogenina, D17S1301, D1GATA113, D18S853, D20S482, y D14S1434.

[0171] Los ejemplos ilustrativos de genes adecuados para predecir la respuesta al tratamiento farmacológico que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a, uno o más de los siguientes genes: ABCB1 (casete de unión a ATP, subfamilia B (MDR/TAP), miembro 1), ECA (enzima convertidora de angiotensina I), ADH1A (alcohol deshidrogenasa 1A (clase I), polipéptido alfa), ADH1B (alcohol deshidrogenasa 1B (clase I), polipéptido beta), ADH1C (alcohol deshidrogenasa 1C (clase I), polipéptido gamma), ADRB1 (receptor adrenérgico beta-1), ADRB2 (receptor adrenérgico beta-2 de superficie), AHR (receptor de aril hidrocarburos), ALDH1A1 (familia de la aldehído deshidrogenasa 1, miembro A1), ALOX5 (araquidonato 5-lipoxigenasa), BRCA1 (cáncer de mama 1, aparición temprana), COMT (catecol-O-metiltransferasa), CYP2A6 (citocromo P450, familia 2, subfamilia A, polipéptido 6), CYP2B6 (citocromo P450, familia 2, subfamilia B, polipéptido 6), CYP2C9 (citocromo P450, familia 2, subfamilia C, polipéptido 9), CYP2C19 (citocromo P450, familia 2, subfamilia C, polipéptido 19), CYP2D6 (citocromo P450, familia 2, subfamilia D, polipéptido 6), CYP2J2 (citocromo P450, familia 2, subfamilia J, polipéptido 2), CYP3A4 (citocromo P450, familia 3, subfamilia A, polipéptido 4), CYP3A5 (citocromo P450, familia 3, subfamilia A, polipéptido 5), DPYD (dihidropirimidina deshidrogenasa), DRD2 (receptor dopaminérgico D2), F5 (factor de coagulación V), GSTP1 (glutación S-transferasa pi), HMGCR (3-hidroxi-3-metilglutaril-Coenzima A reductasa), KCNH2 (canal de potasio activado por voltaje, subfamilia H (eag-relacionado), miembro 2), KCNJ11 (canal rectificador interno de potasio, subfamilia J, miembro 11), MTHFR (5,10-metilentetrahidrofolato reductasa (NADPH)), NQO1 (NAD(P)H deshidrogenasa, quinona 1), P2RY1 (receptor purinérgico P2Y, acoplado a proteína G, 1), P2RY12 (receptor purinérgico P2Y, acoplado a proteína G, 12), PTGIS (prostaglandina I2 (prostaciclina) sintasa), SCN5A (canal de sodio, activado por voltaje, tipo V, alfa (síndrome de QT largo 3)), SLC19A1 (familia de transportadores de solutos 19 (transportador de folato), miembro 1), SLC01B1 (familia de transportadores de aniones orgánicos, miembro 1B1), SULT1A1 (familia de sulfotransferasas, citosólica, 1A, preferente al fenol, miembro 1), TPMT (tiopurina S-metiltransferasa), TYMS (timidilato sintetasa), UGT1A1 (familia UDP glucuronosiltransferasa 1, polipéptido A1), VDR (receptor de vitamina D (1,25- dihidroxivitamina D3)), VKORC1 (complejo vitamina K epóxido reductasa, subunidad 1).

[0172] Los ejemplos ilustrativos de afecciones médicas que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: accidente cerebrovascular, ataque isquémico transitorio, lesión cerebral traumática, enfermedad cardíaca, ataque cardíaco, angina, aterosclerosis e hipertensión arterial.

[0173] Los ejemplos ilustrativos de patógenos que se pueden detectar con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: bacterias, hongos y virus.

[0174] Los ejemplos ilustrativos de especies bacterianas que pueden seleccionarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: un *Mycobacterium* spp., un *Pneumococcus* spp., un *Escherichia* spp., un *Campylobacter* spp., un *Corynebacterium* spp., un *Clostridium* spp., un *Streptococcus* spp., un *Staphylococcus* spp., un *Pseudomonas* spp., un *Shigella* spp., un *Treponema* spp., o un *Salmonella* spp.

[0175] Los ejemplos ilustrativos de especies fúngicas que pueden seleccionarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: *Aspergillus* spp., un *Blastomyces* spp., un *Candida* spp., un *Coccidioides* spp., un *Cryptococcus* spp., dermatophytes, un *Tinea* spp., un *Trichophyton* spp., un *Microsporum* spp., un *Fusarium* spp., un *Histoplasma* spp., un *Mucoromycotina* spp., un *Pneumocystis* spp., un *Sporothrix* spp., un *Exserophilum* spp., o un *Cladosporium* spp.

[0176] Los ejemplos ilustrativos de virus que pueden detectarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a: Influenza A como el H1N1, H1N2, H3N2 y H5N1 (gripe aviar), influenza B, virus de la influenza C, virus de la hepatitis A, virus de la hepatitis B, virus de la hepatitis C, virus de la hepatitis D, virus de la hepatitis E, rotavirus, cualquier virus del grupo de los virus de Norwalk, adenovirus entéricos, parvovirus, virus del dengue, viruela del mono, mononegavirales, virus del lisavirus como virus de la rabia, virus de los murciélagos, virus de Mokola, virus de Duvenhage, virus de los murciélagos europeos 1 y 2 y virus de los murciélagos australianos, efemerovirus, vesiculovirus, virus de la estomatitis vesicular (VSV), herpesvirus como el virus del herpes simple de tipo 1 y 2, varicela zóster, citomegalovirus, virus de Epstein-Bar (EBV), herpesvirus humano (HHV), herpesvirus humano de tipo 6 y 8, virus de la leucemia murina de Moloney (m-MuLV), virus del sarcoma murino de Moloney (MoMSV), virus del sarcoma murino de Harvey (HaMuSV), virus del tumor mamario murino (MumTV), virus de la leucemia del simio gibón (GaLV), el virus de la leucemia felina (FLV), el virus del espumavirus, el virus de la leucemia murina Friend, el virus de las células madre murinas (MSCV) y el virus del sarcoma de Rous (RSV), el VIH (virus de la inmunodeficiencia humana); incluidos el VIH tipo 1 y el VIH tipo 2), el virus visna-maedi (VMV), el virus de la artritis-encefalitis caprina (CAEV), el virus de la anemia infecciosa equina (EIAV), el virus de la inmunodeficiencia felina (FIV), el virus de la inmunodeficiencia bovina (BIV) y el virus de la inmunodeficiencia de los simios (SIV), el virus del papiloma, el gammaherpesvirus murino y los arenavirus como el virus de la fiebre hemorrágica argentina, el virus de la fiebre hemorrágica boliviana, el virus de la fiebre hemorrágica asociada a la Sabia, el virus de la fiebre

hemorrágica venezolana, el virus de la fiebre de Lassa, el virus de Machupo, el virus de la coriomeningitis linfocítica (LCMV), Bunyaviridae como el virus de la fiebre hemorrágica de Crimea-Congo, el hantavirus, la fiebre hemorrágica con el virus que causa el síndrome renal, el virus de la fiebre del Valle del Rift, los filoviridae (filovirus), incluidos la fiebre hemorrágica del Ébola y la fiebre hemorrágica de Marburgo, los flaviviridae, el virus de la enfermedad del bosque de Kaysanur, el virus de la fiebre hemorrágica de Omsk, el virus que causa encefalitis transmitida por garrapatas y Xoviridae como el virus Hendra y el virus Nipah, la variola mayor y la variola minor (viruela), alfavirus como el virus de la encefalitis equina venezolana, el virus de la encefalitis equina oriental, el virus de la encefalitis equina, el coronavirus asociado al SARS (SARS-CoV), el virus del Nilo Occidental, y cualquier virus que cause encefalitis.

[0177] Los ejemplos ilustrativos de genes adecuados para monitorizar un trasplante de órganos en un receptor de trasplante que pueden detectarse, identificarse, predecirse, diagnosticarse o monitorizarse con las composiciones y métodos contemplados en la presente descripción incluyen, pero no se limitan a, uno o más de los siguientes genes: HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DP y HLA-DQ.

[0178] En realizaciones particulares, se usa un análisis bioinformático para cuantificar el número de equivalentes genómicos analizados en la genoteca de clones de ADNlc; detectar variantes genéticas en un locus genético diana; detectar mutaciones dentro de un locus genético diana; detectar fusiones genéticas dentro de un locus genético diana; o medir las fluctuaciones del número de copias dentro de un locus genético diana.

E. Diagnósticos complementarios

[0179] Solo con fines ilustrativos y no mencionados en el conjunto de reivindicaciones, se proporciona un diagnóstico complementario para una enfermedad genética, que comprende: aislar u obtener ADNlc a partir de una muestra biológica de un sujeto; tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente; ligar uno o más adaptadores a cada extremo del ADNlc con los extremos reparados para generar una genoteca de ADNlc; amplificar la genoteca de ADNlc para generar una genoteca de clones de ADNlc; determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y realizar un análisis genético cuantitativo de uno o más biomarcadores asociados con la enfermedad genética en la genoteca de clones de ADNlc, en donde la detección o la falta de detección de al menos uno del uno o más biomarcadores indica si el sujeto debe recibir tratamiento para la enfermedad genética.

[0180] Tal como se usa en la presente descripción, el término “diagnóstico complementario” se refiere a una prueba de diagnóstico que está vinculada a una terapia anticancerosa particular. En una realización particular, los métodos de diagnóstico comprenden la detección de una lesión genética en un biomarcador asociado a una muestra biológica, lo que permite la identificación rápida de los pacientes que deben o no ser tratados con la terapia contra el cáncer.

[0181] La terapia contra el cáncer incluye, pero no se limita a, cirugía, radiación, productos quimioterapéuticos, fármacos contra el cáncer e inmunomoduladores.

[0182] Los ejemplos ilustrativos de fármacos contra el cáncer incluyen, pero no se limitan a: agentes alquilantes tales como tiotepa y ciclofosfamida (CYTOXAN™); alquilsulfonatos tales como busulfán, improsulfán y piposulfán; aziridinas tales como benzodopa, carboquona, meturedopa, y uredopa; se resumen las etileniminas y metilamelaminas, incluidas la altretamina, la trietilenmelamina, la trietilenfosforamida, la trietilenfosforamida y la trimetilolmelamina; mostazas de nitrógeno como clorambucilo, clornafazina, colofosfamida, estramustina, ifosfamida, mecloretamina, clorhidrato de óxido de mecloretamina, melfalán, novembicina, fenesterina, prednimustina, trofosfamida, mostaza de uracilo; nitrosureas tales como carmustina, clorozotocina, fotemustina, lomustina, nimustina, ranimustina; antibióticos como aclacinomisinas, actinomicina, autramicina, azaserina, bleomicinas, cactinomicina, caliqueamicina, carabicina, carminomicina, carzinofilina, cromomicinas, dactinomicina, daunorrubicina, detorubicina, 6-diazo-5-oxo-L-norleucina, doxorubicina y sus formulaciones pegiladas, epirubicina, esorubicina, idarrubicina, marcelomicina, mitomicinas, ácido micofenólico, nogalamicina, olivomicinas, peplomicina, potfiomicina, puromicina, quelamicina, rodorrubicina, estreptonigrina, estreptozocina, tubercidina, ubenimex, zinoestatina, zorubicina; antimetabolitos como metotrexato y 5-fluorouracilo (5-FU); análogos del ácido fólico tales como denopterina, metotrexato, pteropterina, trimetrexato; análogos de purina tales como fludarabina, 6-mercaptopurina, tiamiprina, tioguanina; análogos de pirimidina tales como ancitabina, azacitidina, 6-azauridina, carmofur, citarabina, didesoxiuridina, doxifluridina, enocitabina, floxuridina, 5-FU; andrógenos tales como calusterona, propionato de dromostanolona, epitioestanol, mepitioestano, testolactona; antisuipararrenales tales como aminoglutetimidina, mitotano, trilostano; reponedor de ácido fólico como el ácido frolínico; aceglatona; glucósido de aldofosfamida; ácido aminolevulínico; amsacrina; bestrabucilo; bisantreno; edatraxato; defofamina; demecolcina; diaziquna; elformitina; acetato de eliptinio; etoglúcido; nitrato de galio; hidroxurea; lentinano; lonidamina; mitoguazona; mitoxantrona; mopidamol; nitracrina; pentostatina; fenamet; pirarrubicina; ácido podofilínico; 2-etilhidrazida; procarbazona; PSK®; razoxano; sizofiran; espirogermanio; ácido tenuazónico; triaziquona; 2,2',2''-triclorotrietilamina; uretano; vindesina; dacarbazina; mannomustina; mitobronitol; mitolactol; pipobromán; gacitosina; arabinósido (“Ara-C”); ciclofosfamida; tiotepa; taxoides, p. ej., paclitaxel (TAXOL®, Bristol-Myers Squibb Oncology, Princeton, Nueva Jersey) y doxetaxel (TAXOTERE®, Rhne-Poulenc Rorer, Antony, Francia); clorambucilo; gemcitabina; 6-tioguanina; mercaptopurina; metotrexato; análogos de platino tales como cisplatino y carboplatino; vinblastina; platino; etopósido (VP-16); ifosfamida; mitomicina C; mitoxantrona; vincristina; vinorelbina; navelbina; novantrona; tenipósido; aminopterina;

xeloda; ibandronato; CPT-11; inhibidor de la topoisomerasa RFS 2000; difluorometilomitina (DMFO); derivados del ácido retinoico tales como Targretin™ (bexaroteno), Panretin™ (alitretinoína); ONTAK™ (denileucina diftitox); esperamicinas; capecitabina; y sales, ácidos o derivados farmacéuticamente aceptables de cualquiera de los anteriores. También se incluyen en esta definición los agentes antihormonales que actúan para regular o inhibir la acción hormonal en los cánceres, tales como los antiestrógenos, incluidos, por ejemplo, tamoxifeno, raloxifeno, 4(5)-imidazoles inhibidores de la aromatasa, 4-hidroxitamoxifeno, trioxifeno, keoxifeno, LY117018, onapristona, y toremifeno (Fareston); y antiandrógenos como flutamida, nilutamida, bicalutamida, leuprolida, y goserelina; y sales, ácidos o derivados farmacéuticamente aceptables de cualquiera de los anteriores.

[0183] Los ejemplos ilustrativos de inmunomoduladores incluyen, pero no se limitan a: ciclosporina, tacrolimus, tresperimus, pimecrolimus, sirolimus, verolimus, laflunimus, laquinimod e imiquimod, así como análogos, derivados, sales, iones y complejos de los mismos.

Ejemplos

Ejemplo 1

Detección precisa de mutaciones raras mediante tecnología de captura de secuencias dirigidas

Propósito

[0184] El propósito de este experimento era proporcionar una demostración directa de prueba de principio de la detección de variantes raras usando tecnología de captura de secuencias dirigidas.

Antecedentes

[0185] La tecnología de captura de secuencias diana proporciona un análisis genético cuantitativo basado en secuencias de los ácidos nucleicos y puede aprovecharse para realizar un análisis combinado de mutaciones y números de copias de los genes del metabolismo de los fármacos. Los presentes inventores usaron la tecnología de captura de secuencias dirigidas y el análisis genético posterior para detectar variantes de secuencia raras.

[0186] Las entradas de ADN genómico desempeñan un papel central en la detección de variantes raras, pero el análisis cuantitativo y el control de las entradas genómicas ponen límites a la sensibilidad estimada del análisis de variantes raras. Los presentes inventores usaron un ensayo de qPCR genómica para estimar las entradas genómicas.

[0187] Un diana experimental para el análisis de variantes raras es introducir 10 veces más información genómica como diana de sensibilidad del ensayo. En otras palabras, para medir variantes con una sensibilidad del 1 % (1 en 100), el diana experimental es introducir 1000 genomas. Después de la secuenciación, el análisis bioinformático revela el número de lecturas únicas, y esto tiene la cualidad deseable de ser una medida tanto ortogonal como más directa de las entradas genómicas.

Resumen

[0188] Se mezcló una línea celular (ZR75-30) con SNV conocidas con una muestra de ADN de la línea germinal (NA12878) en una serie de diluciones que varió de 1 a 1 a 1 a 1000. Las regiones diana correspondientes a las diferencias de secuencia conocidas se recuperaron usando tecnología de captura de secuencias dirigidas y se secuenciaron. Se detectaron variantes de secuencia que se producen con una frecuencia de menos de 1 por 1000 secuencias.

Métodos

Sondas de captura

[0189] La siguiente tabla muestra una colección de 62 sondas de captura que se usaron en este experimento.

Tabla 1. Secuencias de sondas de 60 bases usadas en el estudio de prueba de concepto de Admix

diana	Id. de sec. n.º:	Secuencia de la sonda
BRAF	1	TAAACATTGGAAAGGTTTCTAATTAACCAGGAGATCCAAAAGA AAGCGGTTCAGTAGCA
	2	GATCTCAGTTTTTTTGGTTAACTATGTATTTTGGTATATGAAGCT TCTGGGTTTTGCACA
MYCN	3	GACAGATAAGCATACATATTAACATGGATATATATGTGAATTC ATTCAAATGGTTCTCA

	4	AGCTCTTAGCCTTTGGGGGGATGACACTCTTGAGCGGACGTGGG GACGCCCTCGCTCTTTA
5	BRAF	5 AAGCCCCCACC GCCCCTCTTTCCAAAATAAACACCAGCCAGC CGCCGAGCCCGGAGTCG
	6	GCCTCCCTTCCCCCTCCCCGCCGACAGCGGCCGCTCGGGGCCCC GGCTCTCGGTTATAAG
10	CDH1	7 GGTGTGGCAGCCAGGGGGGCGCACTCTGCTCTGGCTGGGCCCC TTCTCCCATGTTTTCTT
	8	TTACACAACCTTTGGGCTTGGACAACACTTTGGGGTCCAAAGAA CCTAAGAGTCTTTCTG
15	EPHX2	9 TGATGAAACTTTGGGCTGGATGGGGCACAGGTAGGGTGCTTGTT GCTTTCAGTCAGATGAA
	10	AATGAAAGAAAAGGAGGCCAGATTGCTACTCCTGGTCCCTGCC ACACACTAGGIACCCCTA
20	BRCA1	11 ATTGACAATACCTACATAAAACTCTTTCCAGAATGTTGTTAAGT CTTAGTCATTAGGGAG
	12	GGATTTCCACCAACACTGTATTATGTACCCATTTTCTCTTAAC CTAACTTTATTGGTC
25	BRCA2	13 CAAAGGGGGAAAACCATCAGGACATTATTTAACAACGGAAATA TCTAACTGAAAAGGCAAA
	14	CAGGCAGACCAACCAAAGTCTTTGTTCCACCTTTTAAACTAAA TCACATTTTTCACAGAG
30	MYCN	15 CCCCAGCCAGCGGTCCGCAACCCTTGCCGCATCCACGAAACTTT GCCCATAGCAGCGGGC
	MYC_r1_F1	16 CGACTCATCTCAGCATTAAAGTGATAAAAAATAAAATTTAAAG GCAAGTGGACTTCGGTG
35	MYC_r1_R1	17 CTGTGGCGCGCACTGCGCGCTGCGCCAGGTTTCCGCACCAAGA CCCCTTTAACTCAAGAC
	MYC_r2_F1	18 TTCTACTGCGACGAGGAGGAGAACTTCTACCAGCAGCAGCAGC AGAGCGAGCTGCAGCCC
40	MYC_r2_F2	19 ACCGAGCTGCTGCGACGAGACATGCTGAACACAGATTTTCATCT GCG/ a continuación
	MYC_r2_F3	20 GCCGCGCGCTCAGAGTGCATCGACCCCTCGGTGGTCTTCCCCTA CCCTCTCAACGACAGC
45	MYC_r2_F4	21 GGCGGCTAGGGGACAGGGCGGGGTGGGCAGCAGCTCGAATTT CTTCCAGATATCCTCGC
	MYC_r2_R1	22 AGACGAGCTTGCGGCGGGCCGAGAAGCCGCTCCACATACAGTC CTGGATGATGATGTTTT
50	MYC_r2_R2	23 AGGAGAGCAGAGAATCCGAGGACGGAGAGAAGGCGCTGGAGT CTTGCGAGGCGCAGGACT
	MYC_r2_R3	24 TAAGAGTGGCCCGTTAAATAAGCTGCCAATGAAAATGGGAAAG GTATCCAGCCGCCACT
55	MYC_r2_R4	25 TTGTATTTGTACAGCATTAACTCTGGTAATTGATTATTTAATGTA ACCTTGCTAAAGGAG
	MYC_r3_F1	26 GAGGCCACAGCAAACCTCCTCACAGCCCACTGGTCCTCAAGAG GTGCCACGTCTCCACAC
60	MYC_r3_F2	27 AGAGGAGGAACGAGCTAAAACGGAGCTTTTTTGCCCTGCGTGA CCAGATCCCCGAGTIGG
	MYC_r3_F3	28 TCCAACCTTGACCCTCTTGGCAGCAGGATAGTCCTCCGAGTGGA GGGAGGCGCTGCGTAG
65	MYC_r3_R1	29 GCTTGGACGGACAGGATGTATGCTGTGGCTTTTTTAAGGATAAC TACCTTGGGGGCGCTTT
	MYC_r3_R2	30 GCATTTGATCATGCATTTGAAACAAGTTCATAGGTGATTGCTCA GGACATTTCTGTTAGA
	MYC_r3_R3	31 CGCCCCGCGCCCTCCAGCCGGGTCCAGCCGGAGCCATGGGGC CGGAGCCGCAGTGAGCA
	ERBB2rlr	32 CTCTGGCCCCGCGCGCCGGGACCTCGGCGGGGCATCCACAG GGCAGGGTCCCCGCGCT

5	ERBB2r2f	33	GGCATGACTTGGAGTGAGTTTGGATGGGGTGGCCAGGTCTGAG AAGGTCCCCGCCAGTG
	ERBB2r2r	34	GCAGGGCACCTTCTTCTGCCACCCACCTGTAAACAGAGGGCTCA GCCAGCTGGAGGCAG
	ERBB2r3f	35	CCCAAGATCTCCAAGTACTGGGGAACCCCAGGGAGGCCCTGGG GGGTGGCAGTGTTCTTA
	ERBB2r3r	36	CTAATGCACACAAAGCCTCCCCCTGGTTAGCAGTGGCCCTGGTC AGCTCTGAATAACCAA
	ERBB2r4f	37	CTGCTCCTCTTTTAGAAGGCAGGAGGGCCCCAAGGGAAGCAGA AGGTGACAGAAGGGGAA
10	ERBB2r4r	38	TGGGGCAGTGGCGGGCAGGCACTGGGTTGTAAGTTGGGAGTTT GCGGCTGGGGTCAGGCT
	ERBB2r5f	39	TCTGCTGCTGTTTGTGCCTCTCTCTGTTACTAACCCGTCCTCTCG CTGTTAGACATCTCT
	ERBB2r5r	40	CCCACCCCTCCCATGTACCTGTATGACACCTGCATTCCACCCG GCCCCAGCCCTCCCT
20	ERBB2r6f	41	TGGGCCAGGTAGTCTCCCTAGAAGGTGATGCTGATGAGGGTCT GGTGGCCAGGGCGCCAC
	ERBB2r6r	42	GGTGGCCACCCCTTGCACTCTGGGGGGTAGAGCACATTGGGCA CAAAGCAGAGGCACATA
	ERBB2r7f	43	CACCCTGCCTGGTACTGCCCTATTGCCCTGGCACACCAGGGCA AAACAGCACAGTGAAA
25	ERBB2r7r	44	CCATTTACAGAAACAAACCTCCCCACCAAAATGAGAAAAGTGT GTTTCTCCCTGGCACTC
	ERBB2r8f	45	TTATTCTTCTTGTGCCTGGGCACGGTAATGCTGCTCATGGTGGT GCACGAAGGGCCAGGG
	ERBB2r8r	46	GAAGGATAGGACAGGGTGGGCTGGGCCAGGCTGCATGCGCAG AGGGACAGGAAGTGCAGC
30	ERBB2r9f	47	GGGCCCCGACCCCTGATGCTCATGTGGCTGTTGACCTGTCCCGGT ATGAAGGCTGAGACGG
	ERBB2r9r	48	TCTGTCTCTGCCATCCCCAAGAGATGCTGCCACATCTGGATCC TCAGGACTCTGTCTGC
	TYMSr2f	49	TCACGTCCCAGGGCAGTTTTCTTCCCTGAAGAAAGTTGGATGGC ATGATCTGTCTTCCCA
40	TYMSr2r	50	GTGTTGAGAACAGACTACTGACTTCTAATAGCAGCGACTTCTTT ACCTTGATAAACACCA
	TYMSr3f	51	AAAAAAAGGATGGGTTCCATATGGGTGGTGTCAAGTGCCCACC TCCTAGCAAGTCAGCAG
	TYMSr3r	52	CCCTCACAAGGTCAAAGCTATACATCAGCTCCTGTGACATTGAC TCATCCCCCAGACCTT
45	TYMSr4f	53	AACCCACCGAGATCTGCAAACTTTGCAGGATGCACCAGATGTC TTGTAGCCATGGGTCAA
	TYMSr4r	54	TGCCTCCCTCAGGTGCCTCTGCACAAAACCAGATTGCTCCCTC TAAGAGTATGGTTAGT
	TYMSr5f	55	GTTTTACTTTGCCTTTAGCTGTGGTCTTTCAAACCACCATCCCTC CTTATCTTCTCTCTGC
50	TYMSr5r	56	CTCTGCAATTTGTTTTCCCATATTAAAGAACTGAAGAGCTCAGT GTGGTAGGCTGGCAAAG
	TYMSr6f	57	TTTTAAATGATGTTTTAAAGAATTGAACTAACATACTGTTCTG CTTTCTCCCCCGGGTT
	TYMSr6r	58	CCTGCCCACCACTTCTCCCTAAACTGAAGCCCCACATTTGGAGC AGTCATCTTTATCTTG
60	TYMSr7f	59	GGTTGCGCTCCAATCATGTTACATAACCTACGGCAAGGTATCGA CAGGATCATACTCCTG
	TYMSr7r	60	GCACAGTTACATTTGCCAGTGGCAACATCCTTAAAAATTAATAA CTGATAGGTACGGAC
	TYMSr1f	61	CGTCCCGCCGCGCCACTTGGCCTGCCTCCGTCCCGCCGCGCCAC TTCGCTGCCTCCGTC
65			

TYMSr1r	62	CTGTAAGGCGAGGAGGACGATGCGTCCCCTCCCTCGCAGGATT GAGGTTAGGACTAAACG
---------	----	--

[0190] Los módulos de las sondas de captura se agruparon a partir de placas madre, se combinaron con el oligo asociado n.º 138 (Id. de sec. n.º 63) (GTGAAAACCAGGATCAACTCCCGTGCCAGTCACAT/3BioTEG/) y se diluyeron hasta una concentración de trabajo final de 1 nM.

Muestras genómicas

[0191] El ADN genómico adquirido comercialmente de la muestra de la línea germinal NA12878 y la línea celular ZR75-30 se fragmentó a una concentración de 10-20 ng/μl hasta un tamaño de fragmento diana de 500 pb en un instrumento de sonicación Covaris. El ADN se purificó con una concentración 1:1 de perlas de purificación de ADN y se reparó en los extremos usando el kit Quick Blunt de New England Biolabs (NEB) a una concentración final de 15 a 30 ng/μl. Los ADN de la línea germinal y de la línea celular se mezclaron en proporciones de 1: 1, 10: 1, 100:1 y 1000: 1, respectivamente. Se construyeron, purificaron y cuantificaron genotecas. Los códigos de muestra, la cuantificación de la genoteca y las entradas usadas para la construcción de la genoteca se muestran en la Tabla 2.

Tabla 2. Adaptadores y análisis genómico de genotecas usadas como entradas

Admix	Código de adaptador	Genomas/μl	Núm. de genomas deseados	μl en PCR
1:1	NNNNNNNNCATGGCCGCAGG (Id. de sec. n.º: 64)	55	200	4
10:1	NNNNNNNNATCTTAGTGGCA (Id. de sec. n.º: 65)	66	200	3
100:1	NNNNNNNNCGGAAGCTCGGAG (Id. de sec. n.º: 66)	64	1000	16
1000:1	NNNNNNNNGACTCCGATCCC (Id. de sec. n.º: 67)	77	10000	130

[0192] Las genotecas genómicas se agruparon, se desnaturalizaron, se combinaron con la sonda, se hibridaron y se lavaron. Los **complejos** de genotecas genómicas marcados con sondas de captura lavados se amplificaron con **cebadores de longitud completa directos e inversos**, se purificaron y se seleccionaron por tamaño para fragmentos de 225 a 600 pb en un instrumento Pippin-Prep. Finalmente, el material capturado se secuenció usando un kit de secuenciación Illumina 150-V3.

Resultados

[0193] Las sondas de captura emparejadas que se dirigen a BRAF (en dos loci), MYCN y CDH1 se usaron para analizar las SVN en estos loci. Los resultados se muestran en la Tabla 3.

Tabla 3. Resumen bioinformático

Mutación	Muestra NA:ZR	Lecturas únicas por etiqueta y posición					Sin filtrado único				
		Número total de lecturas	Cambio a A	Cambio a C	Cambio a G	Cambio a T	Número total de lecturas	Cambio a A	Cambio a C	Cambio a G	Cambio a T
BRAF-I326T-A-G	1to1	186		0	55	0	34.728		6	11.308	10
BRAF-I326T-A-G	10to1	171		0	10	0	46.464		6	2615	2
BRAF-I326T-A-G	100to1	733		0	3	0	51.540		10	89	2
BRAF-I326T-A-G	1000to1	3020		0	3	0	54.565		6	19	7
BRAF-E26D-C-A	1to1	171	66		0	0	29.192	12.388		5	2

5	BRAF-E26D-C-A	10to1	183	12	0	1	38.085	2573	1	4
	BRAF-E26D-C-A	100to1	631	3	0	0	38.913	110	1	2
	BRAF-E26D-C-A	1000to1	2367	1	0	0	46.623	13	2	2
10	CDH1-E243*-G-T	1to1	180	0	0	54	23.846	2	1	8556
	CDH1-E243*-G-T	10to1	208	0	0	12	42.307	1	2	2658
	CDH1-E243*-G-T	100to1	728	0	0	4	42.440	6	0	129
20	CDH1-E243*-G-T	1000to	2727	0	0	0	53.632	5	1	7
	MYCN-E270Q-G-C	1to1	158	0	47	0	20302	0	5587	0
	MYCN-E270Q-G-C	10to1	186	0	14	0	35.733	1	2663	2
30	MYCN-E270Q-G-C	100to1	566	0	3	0	35.393	0	247	5
	MYCN-E270Q-G-C	1000to1	2101	0	2	0	37223	0	15	2

[0194] La columna 3 muestra el número total de recuentos de lectura únicos, que a su vez proporcionan límites en la sensibilidad del ensayo. Las entradas genómicas estimadas y medidas estaban muy dentro del rango entre sí. Los recuadros ligeramente sombreados resaltan la SVN donde la secuencia de la línea celular difería de la secuencia de la línea germinal. En ausencia de un filtro de lectura único, que se muestra en la parte derecha de la tabla, se produjeron cambios de base aleatorios en estas cuatro posiciones seleccionadas con frecuencias medibles distintas de cero. Figura 1. Al requerir que los cambios se produjeran dentro de familias de lectura únicas, fue posible clasificar la señal verdadera del ruido propenso a errores. Figura 2.

Ejemplo 2

Un diseño de sonda novedoso efectivo para la secuenciación completa de las regiones diana en un ADN altamente fragmentado

Propósito

[0195] El propósito de estos experimentos es desarrollar un sistema de ensayo para interrogar de forma fiable y reproducible los ADN circulantes.

Antecedentes

[0196] El análisis del ADN circulante de los fluidos corporales representa una oportunidad emocionante, pero aún no aprovechada, en el diagnóstico molecular. El ADN genómico está muy intacto. La literatura sugiere que el tamaño promedio del ADN circulante es de aproximadamente 150 pares de bases, lo que se correlaciona bien con el tamaño de los ADN envueltos alrededor de un único complejo de histonas nucleosómicas.

Resumen

[0197] Los parámetros técnicos de la tecnología de captura de secuencias dirigidas contemplada en la presente descripción se diseñaron para adaptarse al ADN altamente fragmentado y para conservar la capacidad de generar una cobertura de secuencia completa del ADN diana. La densidad de las sondas de captura aumentó y la longitud de las secuencias de las sondas de captura se redujo de 60 nucleótidos a 40 nucleótidos para minimizar la generación

de secuencias poco informativas en la genoteca de clones. El genoma humano está plagado de secuencias repetitivas y fluctuaciones drásticas en la composición de las bases, por lo que no se pudo admitir la idoneidad de implementar densidades de sonda de captura más altas y sondas de captura más cortas, pero se requirió la validación empírica del nuevo ensayo.

[0198] Se establecieron condiciones en las que las secuencias más cortas de la sonda de captura de 40 mer muestran un rendimiento de ensayo fiable y robusto. En una primera serie de experimentos, los ensayos se usaron para consultar dos grandes regiones: las regiones codificantes del gen supresor de tumores TP53 y el intrón 19 largo y contiguo del oncogén ALK, ambas fundamentales para el diagnóstico del cáncer. En un segundo conjunto de experimentos, se usaron varias sondas de captura por pares de alta densidad que poseen secuencias de sonda de captura de 40 nucleótidos más cortas para interrogar las SVN conocidas que residen en la línea celular NCH69.

[0199] Las nuevas sondas de captura más cortas de alta densidad se usaron con éxito para consultar ADN cortos y fragmentados y los resultados indicaron que el diseño del ensayo es adecuado para la secuenciación de los ADN circulantes que se encuentran en la fracción plasmática de la sangre.

Métodos: Sondas de captura de 40 mer modificadas

[0200] Las secuencias de las sondas de captura usadas para validar empíricamente el rendimiento de las sondas de captura de 40 mer se muestran en la Tabla 4.

Sondas de captura de 60 metros		
Nombre	Id. de sec. n.º:	Secuencia
PLP1_ex2_F	68	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGTTTGAGT GGCATGAGCTACCTACTGGATGTGCCTGACTGTTCCCTTCTTCT TCCC
PLP1_ex2_R	69	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTATCTCCAG GATGGAGAGAGGGGAAAAAAGATGGGTCTGTGTGGGAGGGCA7 0GGTACTT
PLP1_ex3_F	70	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAAAGAAGCC AGGTCTTCAATTATAAGATTCTCTGCTCTCTCTTCTTCTACCTGTT AATG a continuación
PLP1_ex3_M	71	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAGACTCGCG CCCAATTTCCCCCACCCTTGTATTGCCACAAAATCCTGAGGAT GATC
CYP2D6_F	72	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAAAGCACCTAG CCCCATTCTGCTGAGCAGGAGGTGGCAGGTACCCAGACTGGGA GGTAA
CYP2D6_R	73	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGTCGGTGGG GCCAGGATGAGGCCAGTCTGTTACACATGGCTGCTGCCTCTCA GCTCT
chrX_15_F	74	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCTGGCCCTC AGCCAGTACAGAAAGTCATTTGTCAAGGCCTTCAGTTGGCAGACG TGCTC
chrX_15_R	75	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGAAATTCATT GCCAGCTATAAATCTGTGGAAACGCTGCCACACAATCTTAGCACA CAAGA
chrX_69_F	76	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTACTTCCCTC CAGTTTGTGTGCTTGCAAAACAACAGAATCTTCTCTCCATGAAATC ATG
chrX_69_R	77	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCAGGGGTATC TATTATCCCCATTTTCTCACAAGGAAACCAAGATAAAAGGTTTA AATGG
KRAS_ex 1_F	78	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGTTACCTTTA AAAGACATCTGCTTTCTGCCAAAATTAATGTGCTGAACCTAAACT TACC
KRAS_ex 1_R	79	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTCCCAGTAA ATTACTCTTACCAATGCAACAGACTTTAAAGAAGTTGTGTTTAC AATGC

5	KRAS_ex 2_F	80	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCTACTAAATGACAT AACAGTTATGATTTTGCAGAAAACAGATCTGTATTTATTTCACTGT TACT
	KRAS_ex 2_R	81	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGACAGGTTTT GAAAGATATTTGTGTACTAATGACTGTGCTATAACTTTTTTTTCT TTCC
10	MYC_r2_F1	82	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGTGGCGCG CACTGCGCGCTGCGCCAGGTTTCCGCACCAAGACCCCTTAACTC AAGAC
	MYC_r2_R1	83	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGCGGCTAGG GGACAGGGGGCGGGTGGGCAGCAGCTCGAATTTCTTCCAGATATC CTCGC
15	MYC_r2_F3	84	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACACCGAGCTGC TGGGAGGAGACATGGTGAACCAGAGTTTCATCTGCGACCCGGAC GACGAG
	MYC_r2_R3	85	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGGAGAGCAG AGAATCCGAGGACGGAGAGAAAGGCGCTGGAGTCTTCCGAGGCGC AGGACT
20	SRY_r1_F	86	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGTAAGTTA TCGTAAAAAGGAGCATCTAGGTAGGTCTTTGTAGCCAATGTTACC CGATT
	SRY_r1_M3	87	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAAATGGCCATT CTTCCAGGAGGCACAGAAATTACAGGCCATGCACAGAGAGAAAT ACCCGA
25	VHL_r3_F	88	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTTGTTCTGTT CTTGTAAGTACGACCTAGTCTGCCACTGAGGATTTGGTTTTTGCC TTC
	VHL_r3_R	89	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATCAAGACTC ATCAGTACCATCAAAAGCTGAGATGAAACAGTGTAAGTTTCAACA GAAAT
30	UGT1A1_r_4F	90	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTGTGTCCAGC TGTAAGTCTAGAGATGTAAGTGTGACATCTCTCTATTTTGCAT CTCA
	UGT1A1_r_4R	91	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATTTGAAACA ATTTTATCATGAATGCCATGACCAAAGTATTCTTCTGTATCTTCT TCTT
35	TNFRSF1 4_r3_F	92	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTGATGGGTGG GCTCCCGAAGGGGCTCCCGCAGACTTGCGAAGTTCCCACTCTCT GGGCG
	TNFRSF1 4_r3_R	93	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCAGGGTGCGG GGGCATCCAGGCTGCCCAAGCGGAGGCTGGGCCGGCTGTGCTGG CCCTCT
40	RUNX1_r_4_F	94	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTTTTGAAATGT GGGTTTGTGCTATGAAACGTGTTCAAGCATAGTTTGTACAGAT AACG
	RUNX1_r_4_R	95	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTGCCCTAAAA GTGTATGTATAACATCCCTGATGTCTGCATTTGTCTTTGACTGGT GTTT
45	RHD_r5_F	96	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAACCCCTCGA GGCTCAGACCTTTGGAGCAGGAGTGTGATTCTGGCCAACCACCT CTCTG
	RHD_r5_R	97	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCATAAATATG TGTGCTAGTCTGTAGACCCAAGTGTGCCCCAAGGGCAGCGCCC TGCTC
50	PTEN_r5_F	98	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTACTTGTTAAT TAAAAATTCAAGAGTTTTTTTTCTTATTCTGAGGTTATCTTTTTAC CA
	PTEN_r5_R	99	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAAAATCTG TTTTCCAATAAATTCTCAGATCCAGGAAGAGGAAAGGAAAAACA TCAAAA

5	EP300_r1 8_F	100	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACATACTCCATCT CCCGTAAAAATAGTGAGACTTGAGTAATGTTTGATGTCACCTTGTC TTTC
	EP300_r1 8_R	101	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCAGTCACCAC TATATTATTCTAGGTATCCCAGAAAAGTTAAAGTCAAATCTGAAA CACAT
10	VHL_r1_F	102	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCGCCCCGCGT CCGACCCGCGGATCCCGCGGCGTCCGGCCCCGGGTGGTCTGGATCG CGGAG
	VHL_r1_R	103	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCCATACGGGC AGCACGACGCGCGGACTGCGATTGCAAGATGACCTGGGAGGG CTCGCG
15	VHL_r1_M1	104	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCTAGAGGGGCT TCAGACCGTGCTATCGTCCCTGCTGGGTGCGGCCCTAAGCGCCGGG CCCGT
	VHL_r1_M2	105	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGGCGCCGAGG AGGAGATGGAGGCCCGGCGGCCGCGCCCCGTGCTGCTGCTGGTG AACTCG
20	VHL_r2_F	106	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGGTGTGGGCC ACCGTGCCAGCCACCGGTGTGGCTCTTAACAACCTTTGCTTGTC CCGA
	VHL_r2_R	107	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACAAGTGGTCTA TCCGTACTTACCACAACAACCTTATCTTTTAAAAAGTAAAACGT CAGT
25	VHL_r3_F	108	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCTTGTTGTTTC CTTGTAAGTACCTAGTCTGCCACTGAGGATTGGTTTTTGCC TTC
	VHL_r3_R	109	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACATCAAGACTC ATCAGTACCATCAAAAGCTGAGATGAAACAGTGTAAGTTTCAACA GAAAT
40 mer			a continuación
35	Nombre	Id. de sec. n.º:	Secuencia
40	PLP1_ex2 F 40	110	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACACCTACTGGA TGTGCCTGACTGTTTCCCTTCTTCTTCCC
	PLP1_ex2 R 40	111	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGGGAAAAAAA AGATGGGTCTGTGTGGGAGGGCAGGTACTT
45	PLP1_ex3 F 40	112	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCTAATAAGAT TCCCTGGTCTCGTTTGTCTACCTGTTAATG
	PLP1_ex3 M 40	113	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCCCCACCCCTT GTTATTGCCACAAAATCCTGAGGATGATC
50	CYP2D6_F 40	114	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGCTGAGCAGG AGGTGGCAGGTACCCCAGACTGGGAGGTAA
	CYP2D6_R 40	115	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGGCCCAGTCT GTTACACATGGCTGCTGCCTCTCAGCTCT
55	chrX_15_F 40	116	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACGAAAGTCATT TGTC AAGGCCTTCAGTTGGCAGACGTGCTC
	chrX_15_R 40	117	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACAATCTGTGGA AACGCTGCCACACAATCTTAGCACACAAGA
60	chrX_69_F 40	118	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCTGCTTGCAAA ACAACAGAATCTTCTCTCCATGAAATCATG
	chrX_69_R 40	119	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACATTTTCTCACA AAGGAAACCAAGATAAAAAGTTTAAATGG
65	KRAS_ex 1_F_40	120	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCTGCTTTCTGCC AAAATTAATGTGCTGAACTTAACTTACC
	KRAS_ex 1_R_40	121	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACCCAATGCAAC AGACTTTAAAGAAGTTGTGTTTTACAATGC

5	KRAS_ex 2 F 40	122	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACATTTTGCAGA AAACAGATCTGTATTTATTTCACTGTTACT
	KRAS_ex 2 R 40	123	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGTGTTACTA ATGACTGTGCTATAACTTTTTTTCTTTCC
	MYC_r2_F1_40	124	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGCGCCAGGT TTCCGCACCAAGACCCCTTTAACTCAAGAC
	MYC_r2_R1_40	125	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACGGGGTGGGCA GCAGCTCGAATTTCTTCCAGATATCCTCGC
	MYC_r2_F3_40	126	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCATGGTGAAC CAGAGTTTCATCTGCGACCCGGACGACGAG
10	MYC_r2_R3_40	127	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGACGGAGAGA AGGCGCTGGAGTCTTGCAGGGCGCAGGACT
	SRY_r1_F 40	128	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAGCATCTAG GTAGGTCTTTGTAGCCAATGTTACCCGATT
	SRY_r1_M3_40	129	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCACAGAAAT TACAGGCCATGCACAGAGAGAAATACCCGA
20	VHL_r3_F 40	130	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGACCCTAGT CTGCCACTGAGGATTTGGTTTTTGCCCTTC
	VHL_r3_R 40	131	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCAAAAGCTG AGATGAAACAGTGTAAGTTTCAACAGAAAT
	UGT1A1_r_4F_40	132	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGAGATGTAA CTGCTGACATCCTCCCTATTTGCATCTCA
25	UGT1A1_r_4R_40	133	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAATGCCATG ACCAAAGTATTCTTCTGTATCTTCTTTCTT
	TNFRSF1_4_r3_F_40	134	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGCTCCCG CAGACTTGCGAAGTTCCCACTCTCTGGGCG
	TNFRSF1_4_r3_R_40	135	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCTGCCAAG CGGAGGCTGGGCGGCTGTGCTGGCCTCTT
35	RUNX1_r_4_F_40	136	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCCATGAAAC GTGTTTCAAGCATAGTTTGTACAGATAACG
	RUNX1_r_4_R_40	137	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAACATCCCTG ATGCTGCAATTTGCTCTTGACTGGTGTTT
	RHD_r5_F_40	138	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTTGGAGCAG GAGTGTGATTTCTGGCCAACCACCCTCTCTG
40	RHD_r5_R_40	139	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGTTAGACC CAAGTGCTGCCAAGGGCAGCGCCTGCTC
	PTEN_r5_F_40	140	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAAGAGTTTTTT TTTCTTATTCTGAGGTTATCTTTTACCA
	PTEN_r5_R_40	141	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAATTCTCAGA TCCAGGAAGAGGAAAGGAAAAACATCAAAA
50	EP300_r1_8_F_40	142	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATAGTGAGAC TTGAGTAATGTTTGATGTCACCTGTCTTTC
	EP300_r1_8_R_40	143	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTAGGTATCCC AGAAAAGTTAAAGTCAAATCTGAAACACAT
	VHL_r1_F 40	144	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGATCCCCGCG CGTCCGGCCCCGGTGGTCTGGATCGCGGAG
55	VHL_r1_R_40	145	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCGGACTGCG ATTGCAGAAGATGACCTGGGAGGGCTCGCG
	VHL_r1_M1_40	146	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTATCGTCCCT GCTGGGTGCGGCCTAAGCGCCGGGCCCCGT
	VHL_r1_M2_40	147	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGCCGGGCGG CCGCGGCCCCGTGCTGCGCTCGGTGAACTCG
60	VHL_r2_F 40	148	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCCACCGGTG TGGCTCTTTAAACAACCTTGCTTGTCCTGA
	VHL_r2_R_40	149	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACACCACAACAA CCTTATCTTTTAAAAAGTAAACGTCAGT

VHL_r3_F 40	150	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACAGACCCTAGT CTGCCACTGAGGATTTGGTTTTTGCCTTC
VHL_r3_R_40	151	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTTCACTCAAAAAGCTG AGATGAAACAGTGTAAGTTTCAACAGAAAT

[0201] El rendimiento de las sondas de captura de 40 mer se comparó con el de las sondas de captura de 60 mer. El 40 mer se diseñó a partir del 60 mer eliminando 20 nucleótidos del extremo 5' del 60 mer. Aunque el extremo 3' de ambos conjuntos de sondas de captura es idéntico con respecto a las secuencias que se copian de los clones genómicos capturados, la firma de la secuencia de la sonda (lectura 2 de la lectura de los extremos emparejados) es diferente entre los conjuntos de sondas de 60 y 40 mer. Este diseño es útil porque permite multiplexar las sondas de captura durante la secuenciación y analizar su rendimiento posteriormente durante la deconvolución bioinformática corriente abajo.

Muestras genómicas

[0202] Se usó un conjunto de 12 muestras de ADN genómico (elegidas de un panel humano de Coriell de 112 ADN genómico humano) como ADN diana. Las 12 muestras se dividieron en cuatro conjuntos de cuatro muestras cada uno, como se muestra en detalle en la Tabla 5.

Sondas de captura	Temperatura de lavado	Código de muestra	Muestras			
60mer	50C	AT	GM20291	M	Américas	ASCENDENCIA AFRICANA EN EL SUROESTE DE EE. UU.
		CTA	GM19373	M	Africana occidental	LUHYA EN WEBUYE, KENYA
		GCG	HG00428	F	Asiática del este	HAN, SUR DE CHINA
		TCC	HG01624	M	Euro	POBLACIONES IBÉRICAS EN ESPAÑA

40mer	50C	AAT	GM20291	M	Américas	ASCENDENCIA AFRICANA EN EL SUROESTE DE EE. UU.
		CTA	GM19373	M	Africana occidental	LUHYA EN WEBUYE, KENIA
		GCG	HG00428	F	Asiática del este	HAN, SUR DE CHINA
		TCC	HG01624	M	Euro	POBLACIONES IBÉRICAS EN ESPAÑA

60mer	47C	AGA	HG02489	M	Américas	AFROCARIBEÑA EN BARBADOS
		CCT	HG01108	F	Américas	PUERTORRIQUEÑA EN PUERTO RICO
		GAC	GM19011	F	Asiática del este	JAPONESA EN TOKIO, JAPÓN
		TTG	GM18946	F	Asiática del este	JAPONESA EN TOKIO, JAPÓN

40mer	47C	AGA	HG02489	M	Américas	AFROCARIBEÑA EN BARBADOS
		CCT	HG01108	F	Américas	PUERTORRIQUEÑA EN PUERTO RICO
		GAC	GM19011	F	Asiática del este	JAPONESA EN TOKIO, JAPÓN
		TTG	GM18946	F	Asiática del este	JAPONESA EN TOKIO, JAPÓN

60mer	44C	ATC	NA13783	F	NA13783	GM13783
		CAG	HG03700	F	Asiática del sur	PUNJABI EN LAHORE, PAKISTÁN
		GCA	HG03367	M	Africana occidental	ESAN DE NIGERIA
		TGT	NA22991	F	NA22991	GM22991

40mer	44C	ATC	NA13783	F	NA13783	GM13783
		CAG	HG03700	F	Asiática del sur	PUNJABI EN LAHORE, PAKISTÁN

	GCA	HG03367	M	Africana occidental	ESAN DE NIGERIA
	TGT	NA22991	F	NA22991	GM22991

5 Hibridación, lavado y secuenciación

[0203] Se usaron seis condiciones de hibridación diferentes para hibridar las sondas de 60 mer y 40 mer con el ADN genómico diana:

- 1) Sondas de 60meros lavadas a 50 °C.
- 2) Sondas de 40 mer lavadas a 50 °C.
- 3) Sondas de 60 mer lavadas a 47 °C.
- 4) Sondas de 40 mer lavadas a 47 °C.
- 5) Sondas de 60 mer lavadas a 44 °C.
- 6) Sondas de 40 mer lavadas a 44 °C.

Para cada experimento, los oligos de la sonda de captura se combinaron con el oligo asociado; la concentración final de la sonda de captura dúplex fue de 1 nM para cada sonda de captura.

[0204] Cada reacción de hibridación tenía ~2,5 µg de genoteca genómica en un volumen total de 40 µl. Cada muestra se calentó a 98 °C durante 2 minutos y después se enfrió en hielo. Se añadieron 20 µl de sonda de captura y 90 µl de regulador de hibridación y la mezcla de hibridación se incubó durante 24 horas comenzando a 80 °C y disminuyendo un grado cada 48 minutos hasta 50 °C. Los complejos se unieron a 20 µl de perlas de estreptavidina en 1 ml de volumen total de regulador TEzero + 0,05 % de Tween20 (TT). Las perlas se lavaron 3 veces, 5 minutos cada una con 200 µl de TT y una vez a 45 °C durante 5 minutos en regulador de lavado. A continuación, las perlas se lavaron con TEzero y cada reacción se resuspendió en 20 µl de TEzero. A continuación, los complejos se amplificaron por PCR con FLFP de longitud completa hacia adelante (ACA2); Id. de sec. n.º: 152 AATGATACGGCGACCACCGAGATCTACACGTCATGCAGGACCAGAGAATTC GAATACA) y reversa completa (CAC3_FLRP; Id. de sec. n.º: 153 Cebadores CAAGCAGAAGACGGCATACGAGATGTGACTGGCACGGGAGTTGATCCTGGT TTTCAC).

[0205] Tras la amplificación y purificación, se midieron las masas del producto resultante y se agruparon masas iguales para la secuenciación.

Resultados: Cebadores 40 mer modificados

[0206] El rendimiento de la sonda de captura en función de la longitud y la temperatura de lavado se muestra gráficamente en la Figura 3. Generalmente, las sondas de captura de 40 mer funcionaron tan bien como las sondas de captura de 60 mer con lavados a 44 °C y 47 °C. Con el lavado a 50 °C, las sondas de captura de 40 mer muestran un comportamiento esporádico. Estos datos validan empíricamente el uso de sondas de captura de 40 mer y temperaturas de lavado en el rango de 44 °C a 47 °C cuando se usan estos reactivos.

Métodos: 40 mer de alta densidad

[0207] Generalmente, las sondas de captura de secuencias se diseñan usando “reglas” específicas. Por ejemplo, generalmente se evitan las regiones de secuencia redundante o que muestran sesgos extremos en la composición de bases. Una implicación clave del requisito de una alta densidad de sondas y un espaciado reducido de las sondas a lo largo de las regiones clave es que hay poca o ninguna latitud para mover las sondas a fin de adaptarse a cualquiera de estas reglas de diseño de sondas. En este estudio, las sondas se diseñaron basándose únicamente en su posición relativa entre sí sin tener en cuenta las secuencias de unión de las sondas; Por lo tanto, el uso de este enfoque de alta densidad requería validar empíricamente que los métodos de hibridación y procesamiento se adaptarían a dicha colección de sondas.

[0208] El gen ALK humano codifica una proteína quinasa importante en el desarrollo temprano, pero la expresión normal del gen ALK es esencialmente indetectable en adultos normales. Las fusiones oncogénicas de ALK se crean cuando el intrón 19 de la ALK se somete a una recombinación ilegítima para fusionar la parte que codifica la quinasa de la ALK con el extremo 5' de otro gen. Tales fusiones de genes suelen provocar la expresión ectópica de la quinasa ALK, que a su vez es importante para impulsar la proliferación celular inapropiada observada en los tumores pulmonares. En el cáncer de pulmón, este “otro gen” con frecuencia es el EML4, pero también se han detectado otros socios de fusión. Para crear un ensayo que pueda detectar cualquier posible evento de fusión del gen ALK, se diseñaron sondas de 40 nucleótidos que se colocaron a intervalos de 80 nucleótidos en el intrón 19 de ALK. Estas sondas se orientaron de tal manera que no tienen sentido en relación con el gen (Figura 4). Esto significa que su

extremo de 3' se extiende y copia las regiones génicas que están en 5' hasta su sitio de hibridación. Cuando los genes de fusión están presentes, la extensión de la sonda desde las sondas posicionadas cerca de la unión de fusión copia las secuencias de unión. Las secuencias de ADN que resultan de estos clones de unión tienen secuencias de pareja de fusión en su extremo 5', la secuencia de unión de fusión y secuencias del intrón ALK 19 en sus extremos 3' (figura 4B).

[0209] Otro objetivo diagnóstico importante en el cáncer es el gen TP53. Codifica un supresor de tumores y, frecuentemente, se inactiva por mutaciones en los cánceres. Las mutaciones que pueden inactivar la función génica están dispersas por todo el gen y, por lo tanto, los ensayos concluyentes basados en secuencias para detectar las mutaciones inactivadoras del TP53 deben abordar toda la región codificante y las regiones no traducidas (UTR, por sus siglas en inglés) del gen. Debido a que los fragmentos de ADN circulantes son cortos, se usaron sondas de alta densidad para interrogar todas las regiones diana del gen TP53. A diferencia de ALK, las sondas para TP53 se colocan en las dos orientaciones posibles (Figura 5). A altas densidades de sonda, la cobertura acumulada de múltiples sondas proporciona una cobertura uniforme de secuenciación profunda de las regiones diana.

[0210] La colección de 105 sondas usadas en este estudio se muestra en la Tabla 6. Además de las sondas que se dirigen a la región propensa a la fusión de la ALK y a las regiones codificantes de TP53, también se incluyeron las sondas que cubren las SVN conocidas en el ADN de la línea celular.

Tabla 6.

Name_target region	Id. de sec. n.º:	Secuencia de la sonda
ALK_chr2:29446208_fusion_f	154	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCGAAT GAGGGTGATGTTTTCCGCGGCACCTCCITCAGGT
ALK_chr2:29446288_fusion_f	155	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGTTGT AGTCGGTCATGATGGTCGAGGTGCGGAGCTTGCTC
ALK_chr2:29446368_fusion_f	156	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCAGC a continuación TGTGATTCGCGGAGCTTGCTC
ALK_chr2:29446448_fusion_f	157	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTACA CAGGCCACTTCCTACAGGAAGCCTCCCTGGATCTC
ALK_chr2:29446528_fusion_f	158	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAAAT ACTAATAAAATGATTAAAGAAGGTGTGCTTTAAT
ALK_chr2:29446608_fusion_f	159	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTATAT GGAAAATAATTATTTGTATTATATAGGGCAGAGTC
ALK_chr2:29446688_fusion_f	160	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATTAG ACCCAATATGGTCTGCAGATTTTATTAGAAGAAAT
ALK_chr2:29446768_fusion_f	161	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGTGAA CCAGCAGACTGTGTTGCAAGTATAACCCACGTGA
ALK_chr2:29446848_fusion_f	162	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCCAT GGAGCCTAAGGAAGTTTCAGCAAGGCCCTAAGGGG
ALK_chr2:29446928_fusion_f	163	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAG GAATTGGCCTGCCTTAGTATTTCTGCTGTGCTCAG
ALK_chr2:29447008_fusion_f	164	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTTGA GGGTGCAGCTGGGATCTTGGTCAGTTGTGTTTCT
ALK_chr2:29447088_fusion_f	165	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACACACAT CATGAAAAGATCTCTGAATTGGTGTCTGGGGATCT
ALK_chr2:29447168_fusion_f	166	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGAGG ACCAGGTCACAGGACCTCTTTGGACTGCAGTTTCC
ALK_chr2:29447248_fusion_f	167	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTAACC ACTGCCACTCCCCACCCTCTAGGGTTGTCAATGAA
ALK_chr2:29447328_fusion_f	168	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACGAGCT CTACCAATGTGAGTGACCATTATCACTCTACATG
ALK_chr2:29447408_fusion_f	169	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAAAAAT TGTGATTCAGTGGGTAGATTCTGTGTGTAAGCCC
ALK_chr2:29447488_fusion_f	170	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTATGT GCTCAGTTCCTCTCTATGCAATGGACCGACCGT
ALK_chr2:29447568_fusion_f	171	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCGTGA AATTGCCGAGCACGTAGTAACCATGCAACAAGTGT

5	ALK_chr2:29447648_fusion_f	172	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGGGG ACACAGTGTGTGCTGCCATCTCCCTTCTACCGGCA
	ALK_chr2:29447728_fusion_f	173	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCAAAAGAG CCTTTCCCTCTGCCCTTTTCAAGCCTCTGCCCCATC
	ALK_chr2:29447808_fusion_f	174	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGACCA CACTGAGTTCTCTGTGACCTGCAGGTCAGCTCACC
	ALK_chr2:29447888_fusion_f	175	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTTCT ATCTCTCTGCCCTGGAGGGTGGTGGAGGGCTGGTT
	ALK_chr2:29447968_fusion_f	176	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCAAAACA GGAGCTGCCCGGTGGAAGCATGTGGGAGCTAGAA
10	ALK_chr2:29448048_fusion_f	177	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGACA CTGAAGGAGCTCCCCACCCCTGATCAGCCAGGAG
	ALK_chr2:29448128_fusion_f	178	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGAA CTGCAGCTGCTCTGGTGGGGGGAAGGTTGGGAGCT
	ALK_chr2:29448208_fusion_f	179	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACACCCA ATTCCAGGGACTAGCATAACGAAGTGACACCTTGG
20	ALK_chr2:29448288_fusion_f	180	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGC CCCCTGGGAGTCCCTGGGGCTCTGTGCACTCACC
	MYCNr1f_40	181	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGAAAG CACCCCCGGTATTAAACGAACGGGGCGGAAAGAA
	MYCNr1r_40	182	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTAAC AAAGGGGACCGGACCCGGGGTCCAGTGCCCCAGGG
25	MYCNr1f2_40	183	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGG GGGGACTGGGTGGCCTCACCCCCAACCCGGTCATC
	MYCNr1r2_40	184	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCGCGC a continuación CGCTCCACAT
	MYCNr1f3_40	185	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAC CCGGCCGCCGAGTGCGTGGATCCCGCCGTGGTCTT
30	MYCNr1r3_40	186	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGCA CGGGCGCTGGCTCGGCTGTTCACGGGAAAGGGG
	MYCNr2f_40	187	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCAACAT GGATATATATGTGAATTCATTCAAATGGTCTCTCA
	MYCNr2r_40	188	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTAAAC CAACAATCTTAATGTCAACACAATGTTTGTAAAA
40	MYCNr2f2_40	189	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTA CGTGGAGAGTGAGGATGCACCCCCACAGAAGAAGA
	MYCNr2r2_40	190	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATGAC ACTCTTGAGCGGACGTGGGGACGCCTCGCTCTTAA
	MYCNr2f3_40	191	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCTCA CCCTCAGGGACCACGTGCCGGAGTTGGTAAAGAAT
45	MYCNr2r3_40	192	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCAAGTG GCCTTTTCAAAATGACCACCTTGGCGGCCTCTC
	TP53_chr17:7579779: region_1:75nt:-59:-20:f	193	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCTAG GGGGCTGGGGTGGGGTGGGGTGGTGGGCCTGCC
	TP53_chr17:7579838: region_1:75nt:1:40:f	194	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCAGTT TCCATAGGTCTGAAAATGTTTCTGACTCAGAGGG
50	TP53_chr17:7579878: region_1:75nt:41:+5:r	195	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGCC ATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCC
	TP53_chr17:7579932: region_1:75nt:+20:+59:r	196	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCATG CTGGATCCCCACTTTTCTCTTGCAGCAGCCAGAC
	TP53_chr17:7579640: region_2:23nt:-59:-20:f	197	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGCCC CCCAGCCCTCCAGGTCCCCAGCCCTCCAGGTCCCC
60	TP53_chr17:7579741: region_2:23nt:+20:+59:r	198	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCAGA GACCTGTGGGAAGCGAAAATCCATGGGACTGACT
	TP53_chr17:7579252: region_3:280nt:-59:-20:f	199	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCAGG GGGATACGGCCAGGCATTGAAGTCTCATGGAAGCC

5	TP53_chr17:7579311: region_3:280nt:1:40:f	200	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCGTG CAAGTCACAGACTTGGCTGTCCCAGAATGCAAGAA
	TP53_chr17:7579351: region_3:280nt:41:80:r	201	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAGA AAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGC
	TP53_chr17:7579391: region_3:280nt:81:120:f	202	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAAGG GACAGAAGATGACAGGGGCCAGGAGGGGGCTGGTG
	TP53_chr17:7579431: region_3:280nt:121:16 0:r	203	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGTGGC CCCTGCACCAGCAGCTCCTACACCGGCGGCCCTG
	TP53_chr17:7579471: region_3:280nt:161:20 0:f	204	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGGG GAGCAGCCTCTGGCATTCTGGGAGCTTCATCTGGA
10	TP53_chr17:7579511: region_3:280nt:201:24 0:r	205	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCCG GACGATATTGAACAATGGTTCACTGAAGACCCAGG
	TP53_chr17:7579610: region_3:280nt:+20:+5 9:r	206	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGGG GGGCTGGGGGGCTGAGGACCTGGTCTCTGACTGC
	TP53_chr17:7578327: region_4:185nt:-43:-4:f	207	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGG GCAACCAGCCCTGTCGTCTCTCCAGCCCCAGCTGC
15	TP53_chr17:7578370: region_4:185nt:1:40:f	208	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCATC GCTATCTGAGCAGCGCTCATGGTGGGGGCAGCGCC
	TP53_chr17:7578410: region_4:185nt:41:80:r	209	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGCCAT CTACAAGCAGTCACAGCACATGACGGAGGTTGTGA
	TP53_chr17:7578450: region_4:185nt:81:120:f	210	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCATGG CCCGGACGCGGGTGCCGGGCGGGGGTGTGGAATCA
20	TP53_chr17:7578490: region_4:185nt:121:16 0:r	211	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTTGC a continuación iCAGCTGTGGGT
	TP53_chr17:7578574: region_4:185nt:+20:+5 9:r	212	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGCTTT ATCTGTTCACTTGTGCCCTGACTTTCAACTCTGT
	TP53_chr17:7578117: region_5:114nt:-59:-20:f	213	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAGGG CCACTGACAACCACCCTTAACCCCTCCTCCCAGAG
25	TP53_chr17:7578176: region_5:114nt:1:40:f	214	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTCA GGCGGCTCATAGGGCACACCACACTATGTCGAAA
	TP53_chr17:7578216: region_5:114nt:41:80:r	215	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGGAA ATTTGCGTGTGGAGTATTIGGATGACAGAAACACT
	TP53_chr17:7578292: region_5:114nt:+3 :+42:r	216	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAGG GTCCCCAGGCCTCTGATTCTCTACTGATTGCTCTT
30	TP53_chr17:7577439: region_6:111nt:-59: - 20:1	217	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAGGC AAGCAGAGGCTGGGGCACAGCAGGCCAGTGTGCAG
	TP53_chr17:7577498: region_6:111nt:1:40:f	218	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGG AGTCTTCCAGTGTGATGATGGTGAGGATGGGCCTC
	TP53_chr17:7577538: region_6:111nt:41:80:r	219	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTACTAC ATGTGTAACAGTTCCTGCATGGGCGGCATGAACCG
35	TP53_chr17:7577628: region_6:111nt:+20:+5 9:r	220	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTTGC CACAGGTCTCCCCAAGGCGCACTGGCCTCATCTTG
	TP53_chr17:7576974: region_7:138nt:44:-5:f	221	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTGCA CCCTTGGTCTCTCTCCACCGCTTCTTGCTCTGCTTG
	TP53_chr17:7577018: region_7:138nt:1:40:f	222	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTCG CTTAGTGCTCCCTGGGGGCAGCTCGTGGTGAGGCT
40	TP53_chr17:7577058: region_7:138nt:41:80:r	223	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGACCG GCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGG
	TP53_chr17:7577098: region_7:138nt:81:120:f	224	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCTCC CAGGACAGGCACAAACACGCACCTCAAAGCTGTTC
	TP53_chr17:7577138: region_7:138nt:121:+2 2:r	225	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCTCTT TTCCTATCTGAGTAGTGGAATCTACTGGGACG
45	TP53_chr17:7577175: region_7:138nt:+20:+5 9:r	226	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGACA GGTAGGACCTGATTTCTTACTGCCTCTTGCTTCT
	TP53_chr17:7576793: region_8:75nt:-59:-20:f	227	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGCAT TTTGAGTGTTAGACTGGAAACTTTCCACTTGATAA

5	TP53_chr17:7576852: region_8:75nt:1:40:f	228	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGA AGGGTGAAATATTCTCCATCCAGTGGTTTTCTTCTT
	TP53_chr17:7576892: region_8:75nt:41:5:r	229	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTAG CACTGCCCAACAACACCAGCTCCTCTCCCCAGCCA
	TP53_chr17:7576931: region_8:75nt:5:44:r	230	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGCCT CAGATTCACCTTTATCACCTTTTCCTTGCCCTTTTC
	TP53_chr17:7573867: region_9:108nt:59:20:f	231	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACATGGC TTTCCAACCTAGGAAGGCAGGGGAGTAGGGCCAGG
	TP53_chr17:7573926: region_9:108nt:1:40:f	232	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGG AGTGAGCCCTGCTCCCCCTGGCTCCTTCCCAGCC
10	TP53_chr17:7573966: region_9:108nt:41:80:r	233	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCCGA GAGCTGAATGAGGCCTTGGAACTCAAGGATGCCCA
	TP53_chr17:7574053: region_9:108nt:20:5 9:r	234	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCATCT TTAACTCAGGTAAGTGTATATACTTACTTCTCC
	TP53_chr17:7572867: region_10:83nt:59:20:f	235	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGCAG GGGAGGGAGAGATGGGGGTGGGAGGCTGTCAGTGG
20	TP53_chr17:7572926: region_10:83nt:1:40:f	236	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGTCAG TCTGAGTCAGGCCCTTCTGTCTTGAACATGAGTTT
	TP53_chr17:7572966: region_10:83nt:41:80:r	237	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCTGA AGTCCAAAAGGGTCAGTCTACCTCCCCGCCATAAA
	TP53_chr17:7573028: region_10:83nt:20:5 9:r	238	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGCAC AGACCTCTCACTCATGTGATGTCATCTCTCTCC
30	ALDH4A1_chr1:1919 9369_rs61757683:G:T:f	239	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCAGGG GCTTATGTGTCTCCTTGATGACCTGCGGCGACGTC
	ALDH4A1_chr1:1919 9488_rs61757683:G:T:r	240	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCATC ATCTCTCCCTTCCCCTTCTGCCCAGGCTGTTGCA
	BRCA1_chr17:412230 15_rs1799966:T:A,C:f	241	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAAATTC TGGCTTCTCCCTGCTCACACTTTCTTCCATTGCA
35	BRCA1_chr17:412231 34_rs1799966:T:A,C:r	242	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGTCAG CTCGTGTGGCAACATAACCATCTTCAACCTCTGCA
	BRCA1_chr17:412439 21_rs16942:T:C:f	243	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACTAATT TCTTGGCCCCCTCTTCGGTAACCCTGAGCCAAATGT
	BRCA1_chr17:412440 40_rs16942:T:C:r	244	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGTGA AATAAAGGAAGATACTAGTTTTGCTGAAAATGACA
40	BRCA2_chr13:329066 70_rs144848:A:C:f	245	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTCA TTTGTAATCTGAAGTGGAAACCAATGATACTGATCC
	BRCA2_chr13:329067 69_rs144848:A:C:r	246	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGTTG AGACCATTACAGGCCAAAGACGGTACAACCTCTCT
	CDKN2A_chr9:21970 837_rs3731249:C:T:f	247	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGAAAA TGAATGCTCTGAGCTTTGGAAGCTCTCAGGGTACA
50	CDKN2A_chr9:21970 956_rs3731249:C:T:r	248	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACGGGCC ATCGCGATGTCGCACGGTACCTGCGCGCGGCTGCG
	DPYD_chr1:97981316 _rs1801159:T:C:f	249	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCCCA TCCAGCTTCAAAAGCTCTTCAATCATTGATGTGC
	DPYD_chr1:97981435 _rs1801159:T:C:r	250	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTGCCA AGCCTGAACTACCCCTCTTTTCACTCCTATTGAT
55	EPHX1_chr1:2260263 27_rs2234922:A:G:f	251	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCCACC CTGACTGTGCTCTGTCCCCCAGGGCTGGACATCC
	EPHX1_chr1:2260264 46_rs2234922:A:G:r	252	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACAGTCA GGAGTGGGATGATCTTATAAACTCGTAGAAAGAG
	MYC_chr8:12875075 2_G123E:f	253	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCTTCG GGGAGACAACGACGGCGGTGGCGGGAGCTTCTCCA
60	MYC_chr8:12875087 1_G123E:r	254	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTACCATAC AGTCCTGGATGATGATGTTTTGATGAAGGTCTCG
	RB1_chr13:49039115 _rs121913297:G:T:f	255	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACTTTTAC TGTTCTCTCTCAGACATTCAAACGTGTTTTGATC

RB1_chr13:49039204 _rs121913297:G:T:r	256	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACGTGGA AGCATACTGCAAAATATTTGTTTTCAGTCTCTGCA
TNFRSF14_chr1:2491 227 _rs2234163:G:A:f	257	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACACGTA CCCCTCTCAGCCCCCTCTCTTGGACTCCAGCCATG
TNFRSF14_chr1:2491 346 _rs2234163:G:A:r	258	ATGTGACTGGCACGGGAGTTGATCCTGGTTTTCACGTGGC GTAAGCGCGGCACGCGGCGCAGTGGTCCCCGTCCT

Muestras genómicas

[0211] Se analizaron tres muestras de ADN genómico:

- 1) Muestra de línea germinal NA 06994: una muestra humana normal obtenida del depósito de Coriell;
- 2) Línea celular cancerosa NCI-H69: una línea celular conocida por albergar una mutación en el TP53, una amplificación del locus MYCN y SNV en el ALDH4A1, BRCA1, BRCA2, CDKN2A, DPYD, EPHX1, MYC, RB1 y TNFRSF14 que se incluyeron en el conjunto de sondas diana;
- 3) La línea celular cancerosa ZR-75-1, de la que se informó que albergaba un gen de fusión EML4-ALK (Lin y col., Mol. Cáncer Res. 7(9): 1466, 2009).

[0212] Las genotecas de secuenciación de ADN se construyen generalmente a partir de fragmentos de ADN cortados. La disrupción acústica se usó para generar fragmentos de ADN que oscilaban en tamaño entre 200 y >500 bp. La fragmentación enzimática del ADN fragmentado acústicamente se realizó en un esfuerzo por emular el ADN circulante, que tiene fama de estar compuesto por fragmentos nucleosómicos de ~150 pb. En resumen, el ADN a 20-40 ng/μl se soncó en un entorno de 200 pb, lo que produjo fragmentos que varían en tamaño entre 150 bp y 400 bp en un frotis ancho. El ADN se fragmentó adicionalmente mediante la adición de 0,01 y 0,02 μl de la enzima ADNsa (ADNsa bovina recombinante de New England Biolabs) a alícuotas de 50 μl de ADN en regulador de ADNsa (10 mM Tris pH 8,0, 2,5 mM MnCl₂, 0,5 mM CaCl₂). La reacción de la ADNsa se incubó a 37 °C durante 10 min y se detuvo con la adición de EDTA 0,5 M hasta una concentración final de 25 mM. El ADN con un tamaño promedio de 150 pb se purificó mediante selección de perlas “bilateral” añadiendo primero 0,9 volúmenes de perlas a 1 volumen de ADN. Las perlas se unen a fragmentos más grandes no deseados y se descartan, y se añaden 1,6 volúmenes adicionales de perlas al sobrenadante. A continuación, el material unido se purifica y cuantifica. En la Figura 6 se muestra un gel de agarosa del ADN corto, altamente fragmentado y resultante usado para la construcción de genotecas.

[0213] El ADN fragmentado se reparó terminalmente usando el kit Quick Blunt de NEB y se mezcló en las proporciones que se muestran en la Tabla 7. Después se ligaron diez nanogramos de ADN mezclado a adaptadores con las secuencias que se muestran en la Tabla 7. Para las mezclas 9 y 15, se realizaron dos reacciones de ligación con 10 ng cada una y, posteriormente, se combinaron. Para la mezcla 16, se realizaron cuatro reacciones. En la Tabla 7 también se muestra una estimación de las entradas genómicas en cada genoteca usando un ensayo de qPCR.

Tabla 7. Proporciones de muestras y mezclas

Muestra	Proporción de mezcla	Código	Id. de sec. n.º:	Entradas genómicas
1)NA06994 = GL	Puro	NNNNNAAGATCTTAGTGGCAC	259	202
2)NCI-H69 = N	Puro	NNNNNCGACAGAACTATTGCC	260	149
3)ZR-75-1 = Z	Puro	NNNNNACTATCTTAGTGGCAC	261	242
4)GL:N	1:1	NNNNNCTCCAGAACTATTGCC	262	200
5)GL:N	2:1	NNNNNAGCATCTTAGTGGCAC	263	83
6)GL:N	4:1	NNNNNCATCAGAACTATTGCC	264	186
7)GL:N	10:1	NNNNNATAATCTTAGTGGCAC	265	264
8)GL:N	20:1	NNNNNAAGAAGGTAGACCCTC	266	203
9)GL:N	100:1	NNNNNTTCTCTACTCGTGAC	267	436
10)GL:Z	1:1	NNNNNACTAAGGTAGACCCTC	268	297
11)GL:Z	2:1	NNNNNGAAGCTACGAGTATCC	269	224

Muestra	Proporción de mezcla	Código	Id. de sec. n.º:	Entradas genómicas
12)GL:Z	4:1	NNNNNAGCAAGGTAGACCCTC	270	73
13)GL:Z	10:1	NNNNNCATTGACGTCTAGAGC	271	181
14)GL:Z	20:1	NNNNNTCACTCTACTCGTGAC	272	224
15)GL:Z	100:1	NNNNNATAAAGGTAGACCCTC	273	580
16)GL:N:Z	500:1:1	NNNNNTACCTCTACTCGTGAC	274	1324

Secuenciación dirigida

[0214] Se combinó un microgramo de cada una de las dieciséis genotecas de ADN que se muestran en la Tabla 7 y se ajustó a un volumen final de 160 µl. Se desnaturalizaron ocho alícuotas idénticas de 20 µl a 98 °C, se enfriaron en hielo y se añadieron 20 µl de sondas (Tabla 6) a 1 nM/sonda y 50 µl de regulador CF hyb. Las muestras se hibridaron durante 24 horas de 80 °C a 50 °C, se lavaron y se amplificaron. Tras la amplificación de los fragmentos capturados y procesados resultantes, se seleccionó el tamaño de la genoteca de secuenciación final usando el instrumento Pippin Prep™ con una selección de tamaño de 175 a 400 pb. La genoteca se secuenció en un Illumina MiSeq usando un kit V3 de 150 lecturas.

Resultados

[0215] Se monitorizó el rendimiento de las sondas de captura de alta densidad que se eligieron en función de su posición con las secuencias diana. En la Figura 7 se muestra una representación gráfica del rendimiento de cada sonda de captura. Estos datos demuestran que:

- 1) todas las sondas de captura elegidas estrictamente por restricción posicional proporcionaron información sobre la secuencia diana;
- 2) la mayoría de las sondas de captura muestran muy poca captura de lectura fuera del diana y no se puede mapear; y
- 3) el rendimiento de lecturas útiles y adecuadas fue sustancialmente uniforme.

Las sondas de captura que capturaron una alta proporción de lecturas no mapeables y fuera del diana se analizaron más a fondo. Estas sondas de captura se posicionaron generalmente en regiones de baja complejidad de secuencia/alta redundancia de secuencia. Sin embargo, en este caso, tales sondas de captura no tuvieron un impacto determinante significativo en la profundidad de secuenciación porque el alto nivel de redundancia de las sondas (sondas de alta densidad) significa que todas las regiones están cubiertas por lecturas derivadas de varias sondas. El efecto neto fue una excelente uniformidad de cobertura. Véase, p. ej., la Figura 8, la cobertura de la sonda para el gen TP53 usando las sondas de captura de 40 mer.

Conclusiones

[0216] En conjunto, estos datos demuestran que la longitud de la sonda de captura se puede reducir de 60 nucleótidos a 40 nucleótidos con poca o ninguna pérdida perceptible del rendimiento de la sonda (una vez que se ajustan las temperaturas de lavado de captura). También muestran que el diseño de la sonda puede seguir restricciones posicionales y, generalmente, puede ignorar el contexto o la composición de la secuencia. Aunque esta metodología produce ocasionalmente sondas de bajo rendimiento, la alta redundancia con un espaciado estrecho entre las sondas compensa con creces las deficiencias individuales de las sondas.

Ejemplo 3

Análisis genético del ADN circulante

Propósito

[0217] El propósito de este ejemplo era comparar el análisis genético del ADNlc usando un procedimiento de clonación eficiente para el ADNlc y el sistema de recuperación de dianas.

Antecedentes

[0218] Si bien existe un enorme entusiasmo en la comunidad científica y sanitaria por las “biopsias líquidas” (análisis del ADN circulante (ADNlc) en busca de marcadores asociados con posibles estados patológicos, hay muy poca información práctica sobre este posible analito.

Resumen

[0219] Las muestras de plasma recogidas de donantes sanos y de personas que padecían cáncer de ovario o de colon se usaron para realizar el análisis genético del ADN circulante. La cantidad y el carácter general del ADNlc circulante pueden variar ampliamente de un individuo a otro. Sorprendentemente, los presentes inventores descubrieron que el ADNlc es fácilmente clonable con una eficacia indistinguible del ADN genómico altamente purificado y fragmentado; que el tamaño del fragmento era notablemente consistente, con un tamaño medio de inserción de clon de 170 ± 10 pb (en 7/8 muestras); y que la representación del genoma de dichas muestras era uniforme y comparable a la de los experimentos realizados con ADNg purificado. Además, se estableció que al contar lecturas únicas, la profundidad de representación en cada genoteca proporcionaba una estimación de la frecuencia de alelos menores para los marcadores tumorales presentes en el ADNlc de pacientes enfermos. Este estudio estableció que los sistemas de construcción y recuperación de dianas contemplados en la presente descripción se aplicaron eficazmente al análisis genético cuantitativo del ADNlc.

Métodos

Purificación de ADN

[0220] Se adquirieron ocho conjuntos de muestras de plasma de Proteogenex, Inc., Culver City, CA (Tabla 8). El ADN circulante se extrajo de las muestras (en dos ocasiones distintas) usando el kit de purificación de ácidos nucleicos circulantes de Qiagen. Las muestras se pasaron a través de minicolumnas de ADN mediante centrifugación. Los identificadores de las muestras y el rendimiento de ADN se muestran en la Tabla 8.

Tabla 8. Muestras de plasma y rendimientos de ADNlc
a continuación

ID de la muestra	Diagnóstico del paciente	Tipo de muestra	Volumen	Rendimiento de ADN (ng por ml de entrada)
D5930P	Donante sano	Plasma	4 ml	11
D5942P	Donante sano	Plasma	4 ml	68
023407P	Cáncer colorrectal	Plasma	4 ml	10
023406P	Cáncer colorrectal	Plasma	4 ml	63
023185P	Cáncer colorrectal	Plasma	4 ml	171
023149P	Cáncer colorrectal	Plasma	4 ml	36
032667P	Cáncer de ovario	Plasma	4 ml	24
032676P	Cáncer de ovario	Plasma	4 ml	13

Construcción de genotecas

[0221] El ADN purificado de 4 ml de plasma se recogió en 100 μ l de regulador de elución. Para las cuatro muestras recogidas de pacientes con cáncer de colon (CRC), el ADN se dividió por la mitad y se sonicó una alícuota de 50 μ l de cada paciente a 200 pb. Una alícuota de 50 μ l de ADNlc no tratado y un ADNlc fragmentado de 50 μ l de cada paciente (la muestra completa de cada paciente) se repararon finalmente añadiendo (por muestra):

- 6 μ l de solución amortiguadora 10 veces más rápida y contundente (New England Biolabs (NEB))
- 0,6 μ l de dNTP 10 mM
- 2,4 μ l de mezcla enzimática rápida y contundente
- 1,2 μ l de mezcla de enzimas PreCR.

[0222] Las muestras se incubaron a 20 °C durante 30 min y a 70 °C durante 10 min. Las ligaciones con adaptadores (Tabla 2) se realizaron combinando:

- 60 μ l de ADNlc reparado terminalmente
- Adaptador dúplex de 12 μ l (10 μ M)
- 10 μ l de regulador de ligasa 10X (NEB)

- 50 % de 15 µl PEG₈₀₀₀
- 3 µl de ADN ligasa HC T4

5 Tabla 9. Muestras y códigos usados para cuatro muestras de plasma de CCR

ID de Progenex	N.º de muestra	Tratamiento previo	Adaptador
23149	1	ninguna	NNNNNTTTTGTGTGTGTGTG (Id. de sec. n.º:275)
23407	2	ninguna	NNNNNACTACACACACACAC (Id. de sec. n.º:276)
23406	3	ninguna	NNNNNCTCGTGTGTGTGTGT (Id. de sec. n.º:277)
23185	4	ninguna	NNNNNGAACACACACACACA (Id. de sec. n.º:278)
23149	5	Frag.	NNNNNCATGTGTGTGTGTGT (Id. de sec. n.º:279)
23407	6	Frag.	NNNNNGTGCACACACACACA (Id. de sec. n.º:280)
23406	7	Frag.	NNNNNATAACACACACACAC (Id. de sec. n.º:281)
23185	8	Frag.	NNNNNTACTGTGTGTGTGTG (Id. de sec. n.º:282)

[0223] Las reacciones se incubaron a 22 °C durante una hora y a 65 °C durante 10 min. Los productos de ligación se purificaron mediante la adición de 100 µl de perlas, el lavado y la elución en 40 µl de TEzero. Los 40 µl del producto de ligación se amplificaron mediante PCR con el cebador ACA2 (Id. de sec. n.º: 283) y las muestras se combinaron en igual masa para la captura dirigida.

Captura y secuenciación de secuencias dirigidas

[0224] Las cuatro muestras de plasma de colon no fragmentadas y cuatro fragmentadas (Figura 9C) se hibridaron con nuestro conjunto de sondas de 40 nucleótidos de alta densidad que se dirigen a TP53, ALK, entre otras. Los complejos de captura se procesaron como se describió anteriormente en el Ejemplo 2.

Resultados

Aspecto de la genoteca

[0225] En la Figura 9A se muestra una imagen en falso color de un gel de agarosa al 2 % cargado con 50 ng de cada genoteca. El tamaño medio de los fragmentos estaba en un estrecho intervalo de 260 ± 20 bp. Estos datos indicaron que la fracción clonable de ADNlc está presente predominantemente como fragmentos nucleosómicos. Además, el tamaño de las genotecas de ADNlc tenía el mismo aspecto superficial básico que el ADNlc en pacientes en diálisis renal (Atamaniuk y col., Clinical Chemistry 52 (3):págs. 523-26 (2006)) excepto que las genotecas de ADNlc se desplazaron a una masa mayor mediante la adición de secuencias adaptadoras (Figura 9B). Por el contrario, las genotecas de ADNlc diferían drásticamente de las genotecas de ADNg sonicadas, que aparecen como manchas anchas.

[0226] Se construyeron cuatro conjuntos adicionales de genotecas de ADNlc a partir de las muestras de plasma de dos pacientes con cáncer de ovario y dos muestras de plasma de voluntarias sanas. Se repararon terminalmente alícuotas de 38 µl de ADNlc en un volumen total de 50 µl. Las ligaciones incluyeron 40 µl de fragmento con los extremos reparados, 16 µl de adaptador (10 µM), 8 µl de regulador de ligasa 10X, 16 µl de PEG al 50 % y 4 µl de ADN ligasa HC T4 en un volumen total de 80 µl. La reacción de ligación se incubó a 20 °C durante 1 hora y a 65 °C durante 10 min. Para la purificación, se añadieron 20 µl de TEzero y 150 µl de perlas. Los productos de ligación purificados se resuspendieron en 40 µl, todos los cuales se usaron en una posterior amplificación de la genoteca de 200 µl por PCR. Las genotecas amplificadas resultantes se muestran en la Figura 9C.

Análisis de datos de secuenciación

[0227] El recuento promedio de lecturas únicas observado en cada una de las ocho genotecas oscila entre aproximadamente 700 lecturas únicas y más de 3000 lecturas únicas, lo que define un rango de sensibilidades que oscila entre el -0,5 % y el ~0,03 %. Figura 10. Es probable que se observe una lectura de un mutante poco frecuente más de una vez, lo que significa que las sensibilidades mínimas son inferiores a las calculadas anteriormente. En realizaciones preferidas, las lecturas únicas proporcionan el límite inferior de las frecuencias de observación estadísticamente defendibles.

Eficiencia de clonación de ADNlc

[0228] La muestra 23407 se usó como punto de referencia. Se recuperaron 10 ng/ml de ADNlc de la muestra de plasma y se usaron 20 ng del ADNlc aislado en cada uno de los dos esfuerzos de construcción de la genoteca. Los recuentos de lecturas únicas indicaron que recuperamos un promedio de 700 lecturas únicas (equivalentes del genoma) del ADN no fragmentado (“23.407” en la Figura 10). Dado que cada genoma contiene 0,003 ng de ADNg, se recuperaron 2,1 ng de ADN de entrada en esta genoteca (10 % de eficiencia de clonación).

[0229] La fragmentación antes de la construcción de la genoteca con esta muestra aumentó el rendimiento de la genoteca en más de dos veces (“23.407 fragmentos” en la Figura 10). Esto indica que gran parte del ADN presente en la muestra 23.407 era ADN de alto peso molecular que requería fragmentación para poder clonarse. Por lo tanto, la eficiencia de clonación de la genoteca probablemente fue muy superior al 10 % y probablemente estuvo en el rango del 20 % para el ADNlc de entrada. Esta eficiencia de clonación es comparable a la del ADN genómico altamente purificado e indica que el ADNlc probablemente no se modificó de ninguna manera que sea perjudicial para los esfuerzos de clonación posteriores.

Cobertura de genotecas

[0230] Las genotecas de ADNlc se parecían a un conjunto de bandas discretas con una cobertura aleatoria de las regiones diana. La Figura 11 muestra un muestreo aleatorio de datos de secuencia. Un conjunto aleatorio de lecturas de la muestra 23407 que no se fragmentaron antes de la clonación (véase la Figura 10) y que fueron capturadas por la sonda TP53 “chr17:7579351:región_3:280 nt: 41:80:r” (Id. de sec. n.º: 201) se alinearon usando BLAT. Dada la forma en que se preparó la muestra, es probable que sean un reflejo de los fragmentos de ADNlc generalmente porque la parte izquierda de estas lecturas (los sitios de inicio de la lectura) se distribuye aleatoriamente en la región diana. Esta distribución aleatoria indica la rotura aleatoria del ADN genómico y demuestra que, a pesar de la apariencia en forma de banda de las genotecas de ADNlc, el resultado de la secuenciación fue una cobertura aleatoria de la región diana. La distribución aleatoria es importante para un análisis genético eficaz usando la tecnología contemplada en la presente descripción.

[0231] La Figura 12 proporciona una descripción general de mayor resolución de la secuenciación de la región codificante de TP53 para una genoteca de ADNlc típica. Los elementos de la secuenciación dirigida (cobertura en todas las regiones diana y profundidad uniforme en cada base secuenciada) son evidentes. A esta profundidad de >4000 lecturas únicas por base, y con el requisito de que los cambios legítimos en las bases candidatas se produzcan al menos dos veces, es posible estimar que la sensibilidad de detección de mutaciones para esta genoteca en particular fue de aproximadamente 1 mutación en 2000 secuencias, o el 0,05 %. Este nivel de sensibilidad representa un logro técnico sobresaliente, sorprendente e inesperado.

Conclusiones

[0232] El ADNlc se aisló y clonó a partir de clones de plasma con una eficacia comparable a la del ADNg altamente purificado aislado de líneas celulares (el estándar de referencia). Las genotecas de ADNlc se parecían a fragmentos de ADN del tamaño de un nucleosómico circulante o a adaptadores y los extremos poseían un carácter suficientemente aleatorio, lo que permitía un análisis genético eficiente. Además, la característica de tamaño altamente uniforme de las genotecas de ADNlc plasmático permite diseñar estrategias de captura y secuencias de sondas subyacentes para maximizar la cobertura confiable de las dianas a una distancia de hasta 120 pb (= 160-40) desde los extremos de las sondas.

Ejemplo 4

Medición de los equivalentes del genoma en las bibliotecas de adn circulantes

Propósito y antecedentes

[0233] Uno de los principales desafíos en el análisis del ADN circulante libre de células es lograr una sensibilidad de ensayo suficiente. Si no se logra una sensibilidad suficiente, el análisis de las genotecas de ADNlc se confunde: si se secuencia una muestra y no se detectan eventos mutacionales, ese resultado podría interpretarse en el sentido de que no hay mutaciones o que se pasaron por alto eventos significativos porque la profundidad de muestreo fue demasiado pequeña. La sensibilidad de un ensayo se define en términos estadísticos como la tasa de falsos negativos. En el contexto de la secuenciación del ADN circulante libre de células, un obstáculo importante es la detección de una secuencia rara que se mezcla en un gran exceso de secuencia de referencia.

[0234] Un método para determinar la sensibilidad del ensayo es medir las apariciones de la secuencia mutante en un conjunto de muestras donde la secuencia mutante se diluye progresivamente en una secuencia de referencia no mutante. La dilución en la que ya no se detectan las secuencias mutantes define la sensibilidad del ensayo. Este método es adecuado si se conocen tanto la identidad de la secuencia mutante como el grado de dilución. Desafortunadamente, las muestras clínicas generalmente no proporcionan ninguno de los parámetros. Con frecuencia, se desconoce la identidad de la secuencia mutante y el grado de dilución varía de una muestra a otra. En este contexto, la sensibilidad del ensayo se establece muestra por muestra.

[0235] Para asignar un valor de sensibilidad muestra por muestra, se miden los números de secuencias diferentes y distintas que están presentes en cada muestra, midiendo el número de equivalentes genómicos que están presentes en una genoteca de secuenciación. A manera de ejemplo no limitativo, si se sabe que una genoteca de secuenciación de ADN contiene 3 ng (3000 pg) de ADN genómico humano y cada genoma humano tiene una masa de 3 pg, entonces la genoteca posee $3000 \div 3 = 1000$ equivalentes genómicos de ADN. Si una secuencia de ADN mutante debe detectarse dos veces para que sea estadísticamente significativa, entonces una estimación de la mejor sensibilidad de detección posible para esta genoteca en particular es de 2 secuencias mutantes \div 1000 secuencias totales = 0,002 = 0,2 %. Para establecer la sensibilidad, se debe medir el número de equivalentes genómicos para cada genoteca de muestras.

Resumen

[0236] Se usaron dos métodos para medir los equivalentes del genoma. El primer método se basa en la PCR cuantitativa (qPCR). Se construyó una genoteca genómica usando la ligación de adaptadores a fragmentos genómicos y un par de cebadores de PCR, uno que es específico para una secuencia genómica común (p. ej., la repetición Alu I) y otro que es específico para el adaptador. Se midió la abundancia de secuencias de adaptadores: fragmentos ligados de estas genotecas de ADNlc. Se usó una genoteca estándar de concentraciones conocidas para construir una curva estándar y las mediciones se ajustaron a la curva estándar resultante y se obtuvo un valor para los equivalentes del genoma a partir del ajuste.

[0237] El segundo método para medir los equivalentes del genoma usó el recuento bioinformático después de realizar la secuenciación. Cada secuencia única en una genoteca se identificó por su etiqueta de secuencia aleatoria y el nucleótido inicial de la secuencia genómica. Además, cada secuencia única debe derivarse de un genoma independiente. Por lo tanto, la suma de secuencias únicas presentes en los datos de secuencia estableció una medición cuantitativa precisa del número de equivalentes genómicos presentes en una muestra.

Métodos y resultados

Desarrollo del ensayo qPCR

[0238] La primera versión de un ensayo de equivalencia del genoma basado en la qPCR usó el cebador ACA2 (Tabla 10), pero este ensayo subestima crónicamente el número de equivalentes genómicos que están presentes en una genoteca de ADNlc (Figura 13).

Tabla 10. Cebadores de PCR usados en el desarrollo del ensayo qPCR equivalente al genoma

Nombre	Id. de sec. n.º:	Secuencia
ACA2	283	TGCAGGACCAGAGAATTCGAATACA
ACA2_FL FP	284	AATGATACGGCGACCACCGAGATCTACACGTCATGCAGGACCAG AGAATTCGAATACA
Alu_F1	285	CGGTGGCTCACGCCTGTA
Alu_R1	286	GCCTCGGCCTCCCAAAGT
Alu_F2	287	GAGGCTGAGGCAGGAGAATCG
Alu_R2	288	GTCGCCAGGCTGGAGTG

[0239] La versión mejorada del ensayo se basó en repeticiones endógenas (p. ej., repeticiones Alu) que se encuentran con una frecuencia muy alta en todo el genoma humano. Al acoplar un cebador específico de ALU con un cebador específico de adaptador, se midió de manera confiable la frecuencia con la que los adaptadores se unen a los fragmentos genómicos. Se generaron curvas estándar usando genotecas de equivalentes genómicos conocidos, y se midió el número de equivalentes genómicos en genotecas clonadas ajustándose a la curva.

[0240] Los cebadores de PCR usados para desarrollar un ensayo de qPCR basado en un adaptador Alu+ se muestran en la Tabla 10. Los cebadores de PCR para la amplificación del Alu se diseñaron a partir del consenso de una secuencia de Alu humana consensuada (Batzer & Deininger, *Nat Rev Genet*. 3(5) :370-9 (2002)) mediante el uso de PRIMER3 (Alu_F1 y Alu_R1, Id. de sec. n.º: 285 y 286, respectivamente). Los dos cebadores Alu restantes (Alu_F2 y Alu_R2, Id. de sec. n.º: 287 y 288, respectivamente) se informaron en la literatura (Marullo y col., *Genome Biology* 11:R9 (2010)).

[0241] En la Figura 14 se proporciona un esquema del diseño del ensayo. Debido a que se puede usar un único cebador de PCR para amplificar las genotecas de ADN genómico (Figura 14A), se usó un cebador que reconoce la secuencia adaptadora pero que no puede amplificar los clones genómicos. El cebador ACA2-FLFP de 58 nucleótidos (en adelante abreviado AF, Id. de sec. n.º: 284) cumple estos criterios porque su longitud induce una fuerte supresión

de la PCR en bucle troncal (figura 14B). Además, se usó un par funcional de cebadores Alu (Figura 14C). Además, se usó un par de cebadores que consistía en un cebador Alu y el cebador largo ACA2 que no amplificaba el ADN genómico (Figura 14D). Estos mismos cebadores también amplificaron los clones de la genoteca genómica (Figura 14E).

[0242] Se validaron todos los elementos necesarios para un ensayo funcional basado en ALU. Figura 15. Específicamente, el cebador largo solo era inerte, ambos conjuntos de pares de cebadores Alu reconocían el ADN genómico humano y cualquier combinación de un cebador Alu y el cebador largo ACA2 amplificaba los clones de la genoteca genómica (Figura 15A). Finalmente, la capacidad del cebador Alu más el par de cebadores ACA2 largos para discriminar entre el ADN genómico y los clones de la genoteca genómica se muestra en la Figura 15B. La combinación de los cebadores Alu_R1 y AF se usó para medir los equivalentes del genoma en las genotecas recién construidas.

[0243] En la Figura 16 se muestra una comparación directa entre los ensayos de qPCR basados en ACA2 y los basados en ALU. Se encontró una diferencia de 8 veces en los equivalentes del genoma. Además, los ensayos basados en ALU proporcionaron un rendimiento más consistente de genoteca a genoteca y una mejor alineación entre los equivalentes derivados de la qPCR y los equivalentes de etiquetas contados bioinformáticamente en las ejecuciones de secuenciación (Tabla 11).

Tabla 11. qPCR frente a etiquetas de secuenciación contadas

Muestra	qPCR basada en ALU	etiquetas contadas
Run_68 50to1	6962	3459
Run_68 1000to1	10.937	4641

Adaptadores de genotecas de alta sensibilidad para el recuento basado en secuencias de equivalentes del genoma

[0244] Como se mencionó anteriormente, la realidad de la vigilancia de enfermedades mediante el uso del ADN es que las secuencias mutantes pueden ser constituyentes poco frecuentes en un exceso, por lo demás enorme, de secuencias de ADN “normales” (es decir, de la línea germinal). Por lo tanto, se necesitan ensayos de secuenciación altamente sensibles y cuantificables. La sensibilidad del ensayo podría hacerse contando el número de secuencias únicas presentes en una genoteca de secuenciación. Sin embargo, tal recuento llevaría a una *falsa subestimación* de la sensibilidad porque los fragmentos de ADN son bastante cortos (~165 pb) y pueden conducir a lecturas idénticas que en realidad se derivaron de eventos de clonación independientes. Una solución a este problema es marcar cada clon de secuenciación independiente durante la construcción de la genoteca incluyendo, por ejemplo, un conjunto de etiquetas de ADN en los adaptadores usados para construir las genotecas.

[0245] Un conjunto de dichos adaptadores de construcción de genotecas se diseñó específicamente para medir el número de equivalentes genómicos presentes en las genotecas de ADN y, por extensión, la sensibilidad de los ensayos de secuenciación usados para monitorizar las secuencias mutantes.

[0246] La arquitectura de los adaptadores de genotecas de alta sensibilidad que se configuraron para acomodar un gran número de equivalentes de genoma en genotecas de ADN se muestra en la Figura 17. Existe una cantidad sustancial de ingeniería molecular dentro de la cadena de ligación de 45 nucleótidos, que es la cadena que se une a los fragmentos de ADN reparados en los extremos. Los adaptadores comprenden al menos cinco elementos.

[0247] El elemento 1 es un sitio de unión al cebador de PCR para el cebador de amplificación de la genoteca de un solo cebador ACA2 (tabla 12).

Tabla 12

Elemento	Función	Número de secuencias	Secuencia (5' ->3')	Id. de sec. n.º:
Elemento 1	Sitio de unión del cebador de PCR	1	TGCAGGACCAGAGAATTCGAATACA	289

[0248] El elemento 2 es un código de lectura de 5 nucleótidos. La combinación de este código con la secuencia de ADN genómico constituye la etiqueta de ADN que se usó para identificar de forma única cada lectura. Los códigos de 5 nucleótidos constan de 256 posibles secuencias únicas que se eligieron por ser cambios de 2 bases diferentes de cualquier otro código del conjunto. Esta característica permitía diferenciar las lecturas únicas y distintas de las lecturas que parecían únicas debido a un error de secuenciación en la región del código. Se eliminaron siete códigos en los que los residuos G están sobrerepresentados y que se demostró empíricamente que interferían con la función del adaptador, dejando 249 códigos aleatorios. Tabla 13.

Tabla 13

5	Elemento		Función		Número de secuencias			
	Elemento 2		Etiquetas de secuencia distintas		249			
10	Secuencias (5' - > 3')	Id. de sec. n.º:	Secuencia (5' -> 3')	Id. de sec. n.º:	Secuencia (5' -> 3')	Id. de sec. n.º:	Secuencia (5' -> 3')	Id. de sec. n.º:
	CGGGT	290	GGGTC	354	AGAGA	418	CCGGA	482
	CGGTG	291	GGTCG	355	AGCCG	419	CGACG	483
15	CGTGG	292	GGTGC	356	AGCGC	420	CGAGC	484
	GCGGT	293	GTCGG	357	AGGAA	421	CGCAG	485
	GCGTG	294	GTGCG	358	AGGCC	422	CGCGA	486
	GCTGG	295	GTGGC	359	AGGTT	423	CGGAC	487
20	GGCGT	296	TGCGG	360	AGTGT	424	CGGCA	488
	GGCTG	297	TGGCG	361	AGTTG	425	GAAAG	489
	GGGCT	298	TGGGC	362	ATGGT	426	GAAGA	490
25	TTAAA	299	AAAGG a continuación				GACCG	491
	TTACC	300	AAGAG	364	ATTGG	428	GACGC	492
	TTATT	301	AAGGA	365	CACGG	429	GAGAA	493
30	TTCAC	302	ACCGG	366	CAGCG	430	GAGCC	494
	TTCCA	303	ACGCG	367	CAGGC	431	GAGTT	495
	TTTAT	304	ACGGC	368	CCAGG	432	GATGT	496
35	TTTTA	305	AGAAG	369	CCGAG	433	GATTG	497
	GCACG	306	GTGTA	370	AACTG	434	AGTCA	498
	GCAGC	307	GTTAG	371	AAGCT	435	ATACG	499
	GCCAG	308	GTTGA	372	AAGTC	436	ATAGC	500
40	GCCGA	309	TAGGT	373	AATCG	437	ATCAG	501
	GCGAC	310	TAGTG	374	AATGC	438	ATCGA	502
	GCGCA	311	TATGG	375	ACAGT	439	ATGAC	503
45	GGAAA	312	TGAGT	376	ACATG	440	ATGCA	504
	GGACC	313	TGATG	377	ACGAT	441	CAAGT	505
	GGATT	314	TGGAT	378	ACGTA	442	CAATG	506
50	GGCAC	315	TGGTA	379	ACTAG	443	CAGAT	507
	GGCCA	316	TGTAG	380	ACTGA	444	CAGTA	508
	GGTAT	317	TGTGA	381	AGACT	445	CATAG	509
	GGTTA	318	TTAGG	382	AGATC	446	CATGA	510
55	GTAGT	319	TTGAG	383	AGCAT	447	CCCGT	511
	GTATG	320	TTGGA	384	AGCTA	448	CCCTG	512
	GTGAT	321	AACGT	385	AGTAC	449	CCGCT	513
60	CCGTC	322	CTGTT	386	GTAAC	450	TCCGC	514
	CCTCG	323	CTTGT	387	GTACA	451	TCGAA	515
	CCTGC	324	CTTTG	388	GTCAA	452	TCGCC	516
65	CGAAT	325	GAACT	389	GTCCC	453	TCGTT	517

5	CGATA	326	GAATC	390	GTCTT	454	TCTGT	518
	CGCCT	327	GACAT	391	GTTCT	455	TCTTG	519
	CGCTC	328	GACTA	392	GTTTC	456	TGAAC	520
	CGTAA	329	GATAC	393	TAACG	457	TGACA	521
	CGTCC	330	GATCA	394	TAAGC	458	TGCAA	522
10	CGTTT	331	GCAAT	395	TACAG	459	TGCCC	523
	CTAAG	332	GCATA	396	TACGA	460	TGCTT	524
	CTAGA	333	GCCCT	397	TAGAC	461	TGTCT	525
	CTCCG	334	GCCTC	398	TAGCA	462	TGTTC	526
15	CTCGC	335	GCTAA	399	TCAAG	463	TTCGT	527
	CTGAA	336	GCTCC	400	TCAGA	464	TTCTG	528
	CTGCC	337	GCTTT	401	TCCCG	465	TTGCT	529
20	TTGTC	338	ACTTC	402	CCACC	466	TATAA	530
	TTTCG	339	ATAAT	403	CCATT	467	TATCC	531
	TTTGC	340	ATATA	404	CCCAC	468	TATTT	532
25	AAAAA	341	ATCCT	405	CCCCA	469	TCACT	533
	AAACC	342	ATCTC	a continuación		70	TCATC	534
	AAATT	343	ATTAA	407	CCTTA	471	TCCAT	535
	AACAC	344	ATTCC	408	CTACT	472	TCCTA	536
30	AACCA	345	ATTTT	409	CTATC	473	TCTAC	537
	AATAT	346	CAAAC	410	CTCAT	474	TCTCA	538
	AATTA	347	CAACA	411	CTCTA	475		
35	ACAAC	348	CACAA	412	CTTAC	476		
	ACACA	349	CACCC	413	CTTCA	477		
	ACCAA	350	CACTT	414	TAAAT	478		
40	ACCCC	351	CATCT	415	TAATA	479		
	ACCTT	352	CATTC	416	TACCT	480		
	ACTCT	353	CCAAA	417	TACTC	481		

45 [0249] El elemento 3 es un código de muestra de 3 nucleótidos que difiere en al menos dos cambios de base. Este elemento se usó para identificar diferentes muestras y permitió la multiplexación de muestras dentro de una secuencia de secuenciación. Tabla 14.

Tabla 14

50	Elemento	Función	Número de secuencias	Secuencia (5' - >3')	Id. de sec. n.º:
55	Elemento 3	Etiquetas de muestra distintas; multiplexación de muestras	16	AAG	539
				CTC	540
				GGT	541
60				TCA	542
				ACT	543
				CGA	544
				GTG	545
65				TAC	546

			AGC	547
			CCG	548
5			GAA	549
			TTT	550
			ATA	551
10			CAT	552
			GCC	553
			TGG	554

[0250] El elemento 4 es una secuencia de anclaje de 12 nucleótidos con tres características importantes con respecto a la construcción de la genoteca y la secuenciación posterior. Tabla 15. Estas son: A) cada extensión de 12 bases es parte de una familia de cuatro extensiones de 12 bases que representan colectivamente cada una de las cuatro posibles bases de ADN en cada sitio dentro de la extensión. El instrumento de secuenciación de Illumina requiere esta función, la representación de bases equilibrada para calibrar las llamadas de base adecuadas en las lecturas de secuenciación. B) Cada extensión está compuesta solo por dos de las cuatro bases posibles, y estas se eligen específicamente para que sean 6 A + 6 C o 6 G + 6 T. Esta extensión formada a partir de solo dos bases reduce en gran medida la posibilidad de que la secuencia de extensión participe en la formación de una estructura secundaria que impediría la función adaptadora adecuada. C) Como cada extensión está compuesta por números iguales de A+C o G+T, cada extensión comparte aproximadamente la misma temperatura de fusión y estabilidad dúplex que cualquier otra extensión de un conjunto de cuatro.

Tabla 15.

Elemento	Función	Número de secuencias	Secuencia (5' - >3')	Id. de sec. n.º:
Elemento 4	Extensión de 12 nucleótidos que proporciona un sitio de duplexación para el oligonucleótido asociado	4	ACCCACACCAAA	555
			CAAACACAACCC	556
			GTGTGGGTTGTT	557
			TGTGTTTGGTGG	558

[0251] El elemento 5 es la secuencia de dos bases que se encuentra en el extremo 3' del elemento 4. Las dos extensiones de bases particulares se eligieron basándose en datos empíricos que muestran que estas dos secuencias de bases son sustratos eficaces para la ligación. Tabla 15. 4.

[0252] El módulo adaptador se hibrida con un oligonucleótido asociado. Tabla 16. La hibridación tiene lugar entre la secuencia dentro del elemento 4 y el oligonucleótido asociado. El adaptador bicatenario se ligó al ADNlc con extremos reparados.

Tabla 16.

Elemento	Función	Número de secuencias	Secuencia (5' - >3')	Id. de sec. n.º:
Elemento 4	Extensión de 12 nucleótidos que proporciona un sitio de duplexación para el oligonucleótido asociado	4	TTTGGTGTGGGT	559
			GGGTTGTGTTTG	560
			AACAACCCACAC	561
			CCACCAAACACA	562

[0253] Para convertir un conjunto de 256 cadenas de ligamiento agrupadas y sintetizadas de forma independiente (cada una de las cuales comparte un código de muestra común y por lo tanto constituye un único conjunto de adaptadores de muestra) en dúplex adecuados para la ligación, la cadena de ligación de 45 nucleótidos se combinó con la cadena asociada complementaria de 12 nucleótidos apropiada, se calentó a 95 °C, se enfrió a 65 °C durante 5 minutos y después se enfrió a temperatura ambiente. Este dúplex formó un sustrato de ligación de extremo romo como se muestra en la Figura 17B. Tras la ligación y la purificación del ADN, una etapa mediada por la ADN polimerasa que se produce antes de la amplificación por PCR desplazó la cadena asociada y copió la cadena de ligación para formar un adaptador bicatenario que era adecuado para la amplificación exponencial mediante PCR de un solo cebador.

[0254] A continuación, se realizó el análisis cuantitativo de los equivalentes del genoma derivados de los datos de secuenciación dirigida. Cada lectura única se consideró un evento de ligación único y la suma de las lecturas únicas se consideró equivalente al número de equivalentes del genoma analizados.

[0255] Se realizó un cálculo aproximado “rápido” y “regla general” para determinar el número de equivalentes genómicos que podían analizarse. Cada clon de ADNlc tenía aproximadamente 150 pares de bases, de los cuales se requerían 50 pares de bases para unirse a las sondas de captura. Esto dejó aproximadamente 100 posibles sitios de inicio de secuencia dentro de cualquier clon de ADNlc capturado. El acoplamiento de 249 códigos aleatorios a cada uno de los 100 posibles sitios de inicio creó un repertorio total de -249.000 posibles clones únicos. A medida que el número de clones únicos se acerca al número total de posibles combinaciones de secuencias, la probabilidad dicta que las mismas combinaciones de código y sitio de inicio se creen mediante eventos independientes y que estos eventos independientes se agrupen de manera inapropiada en familias individuales. El resultado neto será una subestimación de los equivalentes genómicos analizados, y las lecturas de mutantes raros pueden descartarse como errores de secuenciación porque se superponen con las lecturas silvestres que contienen los mismos identificadores. Para evitar esto, se hicieron esfuerzos usando el ensayo qPCR para restringir las entradas genómicas a una décima parte o menos del número de posibles clones únicos. Por ejemplo, un único adaptador tiene 24.900 clones posibles y, por lo tanto, tiene una capacidad fiable para proporcionar análisis precisos para genotecas que constan de 2500 o menos equivalentes de genoma.

[0256] El procedimiento que se describe se proporciona como un ejemplo y los métodos contemplados en la presente descripción no pretenden limitarse a este ejemplo. En algunos casos, el número de equivalentes genómicos a analizar puede superar con creces el límite de 2500 ilustrado en el párrafo anterior. Para ampliar la profundidad de los equivalentes del genoma, hay dos soluciones disponibles para este problema. La primera solución es usar más de un juego de adaptadores por muestra. Al combinar adaptadores, es posible ampliar el número total de clones posibles y, por lo tanto, ampliar los cómodos límites de la entrada genómica. Como ejemplo no limitativo, la combinación de cuatro conjuntos de adaptadores usados para una muestra ampliaría el análisis a $24.900 \times 4 = 99.600$ secuencias posibles y -10.000 equivalentes genómicos razonablemente analizados. La segunda solución es expandir el código en el elemento 2 de la Figura 17A a 6, 7 o más bases. El número de códigos posibles que difieren en al menos 2 bases de cualquier otro código se escala como $4^{(n-1)}$, donde n es el número de bases dentro del elemento 2. Por lo tanto, en el ejemplo no limitativo presentado en la presente descripción, $n = 5$ y $4^{(5-1)} = 256$; por lo tanto, la inclusión de bases adicionales amplía el repertorio disponible en un factor de cuatro por cada base adicional.

Conclusiones

[0257] Los resultados de este ejemplo mostraron que dos métodos independientes para la determinación de los equivalentes del genoma tienen utilidad en el flujo de trabajo de procesamiento de muestras. El primer método, qPCR, se implementó durante la fase de construcción de la genoteca del análisis de ADNlc y se usó como paso de control de calidad para garantizar que se transfiera un número adecuado de equivalentes del genoma mediante la amplificación de la genoteca, la captura de secuencias dirigidas y la secuenciación del ADN. El otro método usa el recuento explícito de lecturas únicas como una medida más directa del número real de equivalentes del genoma que se consideraron desde el punto de vista de la informática.

Ejemplo 5

Análisis genético cuantitativo

Propósito

[0258] El propósito de este ejemplo era aplicar el análisis genético cuantitativo a genomas de cáncer mezclados con ADN normal y al ADNlc no caracterizado aislado del plasma de pacientes con cáncer.

Antecedentes

[0259] En los cánceres humanos prevalecen tres tipos de eventos genómicos. Se trata de mutaciones somáticas que alteran la función del gen afectado y de sus productos proteicos expresados; reordenamientos genómicos que crean fusiones génicas quiméricas y, por lo tanto, expresan proteínas de fusión con propiedades biológicas novedosas; y cambios en el número de copias de los genes que conducen a la pérdida de los genes y a la subexpresión de los productos génicos o, por el contrario, a la amplificación de los genes y a la sobrerrepresentación de los productos génicos correspondientes. En el ADN circulante de un paciente con cáncer, estos loci aberrantes, muchos de los cuales tienen una importancia crítica para guiar la atención del paciente, se mezclan con el ADN normal de la línea germinal del paciente.

Resumen

[0260] En los ejemplos anteriores, se ha descrito una tecnología que se configuró para el análisis del ADN circulante libre de células (ADNlc), con el objetivo de vigilar el cáncer. Sin embargo, la tecnología es ampliamente aplicable a cualquier paradigma analítico, de diagnóstico y de monitorización, incluidas, entre otras, las enfermedades genéticas; pruebas fetales; trastornos mendelianos; detección de patógenos; y la monitorización de trasplantes de órganos en los que el ADN circulante es un posible analito. En este ejemplo, las características técnicas destacadas en los ejemplos anteriores se aplican al análisis de muestras de cáncer mezcladas. En la primera fase de esta validación, las líneas celulares derivadas del cáncer se mezclaron con ADN humano normal en diluciones definidas y se realizó un análisis genético cuantitativo. En la segunda fase de este estudio, se aisló el ADNlc no caracterizado del plasma de pacientes con cáncer y, posteriormente, se examinó mediante un análisis genético cuantitativo.

Métodos

Mezclas de ADN genómico de líneas celulares con ADN humano normal

[0261] Se usaron las siguientes muestras de ADN:

- NA06994 - ADN genómico humano normal (depósito de Coriell);
- NCI-H2228: línea celular de cáncer de pulmón de células no pequeñas (ATCC), que alberga una mutación en la fusión de los genes TP53 (Q331*) y EML4-ALK (punto de quiebre desconocido); y
- NCI-H69: línea celular de cáncer de pulmón de células pequeñas (ATCC), alberga la amplificación del gen MYCN (~100 copias).

[0262] **Preparación de la genoteca:** El ADN genómico aislado de las líneas celulares (las tres anteriores) es un material de alto peso molecular que es diferente al pequeño tamaño del ADNlc. Para imitar el ADNlc en estos experimentos de validación, los ADN genómicos se fragmentaron primero en el entorno de “150 pb” usando un sonicador acústico Covaris. La sonicación generalmente produce un frotis amplio y el ADN se procesó posteriormente mediante una selección de perlas “bilateral”. Se añadió una solución diluida de perlas de purificación de ADN a la muestra y se desecharon los fragmentos de mayor masa molecular que se adhieren a las perlas (el tamaño del ADN purificado es proporcional a la cantidad de perlas añadidas). Se añade una alícuota adicional de perlas al sobrenadante restante y, en esta segunda ronda, se purifica el ADN que se adhiere a las perlas (en una concentración global más alta de regulador de unión). Esta purificación “bilateral” produce una distribución de tamaños estrecha que es un indicador razonable del ADNlc (Figura 18).

[0263] El ADN genómico fragmentado se reparó al final, se cuantificó y se mezcló en las diversas proporciones que se muestran en la Tabla 17 y se describen en la sección de resultados a continuación.

Tabla 17. Muestras y mezclas usadas para los estudios de validación

Muestra, mezcla	equivalentes del genoma (qPCR)
H2228 puro	2248
NA06994:H2228 4:1	2616
NA06994:H2228 10:1	2600
NA06994:H2228 20:1	2968
NA06994:H2228 50:1	5000
NA06994:H2228 1000:1	10000
Muestra, mezcla	equivalentes del genoma (qPCR)
H69 puro	2472
NA06994:H69 4:1	2768
NA06994:H69 10:1	3088
NA06994:H69 20:1	2944
NA06994:H69 40:1	1616
NA06994:H69 100:1	1920
NA06994:H69 200:1	2920
NA06994:H69 500:1	17520

[0264] Las genotecas de ADNlc pueden tener entradas de ADN limitadas. La cantidad de ADNlc obtenida por ml de plasma del paciente es muy variable, pero los límites inferiores (p. ej., el ejemplo 3) son generalmente de ~10 ng/ml, lo que equivale a 3300 genomas humanos. Para evitar cantidades limitadas de ADNlc, los experimentos de mezcla se modelaron para reflejar los límites inferiores de ADNlc que se recolectaron rutinariamente de los pacientes. Esta restricción se aplicó a todos los aditivos excepto a las mezclas más extremas. En estas últimas mezclas, se crearon genotecas para imitar las entradas de 4 ml (NA06994:H2228 1000:1) u 8 ml (NA06994:H69 500:1) de ADNlc de pacientes de bajo rendimiento. Las muestras mezcladas se ligaron después a los conjuntos de adaptadores descritos en el Ejemplo 4. La medición de los equivalentes del genoma en cada genoteca purificada usando qPCR (Ejemplo 4) también se muestra en la Tabla 17. Las genotecas se amplificaron, cuantificaron y se agruparon las masas equivalentes de cada genoteca (500 ng de cada una). La muestra agrupada se hibridó con las sondas de captura de 40 mer de alta densidad de prueba de concepto enumeradas en la Tabla 6 del Ejemplo 2. Los complejos resultantes se capturaron en perlas recubiertas con estreptavidina, se lavaron, se procesaron, se amplificaron, se purificaron y se seleccionaron por tamaño como se describe en los ejemplos anteriores. La genoteca resultante se analizó usando un kit de secuenciación Miseq Illumina 150 bp-V3 en el instrumento Illumina MiSeq.

[0265] Para el análisis bioinformático, se usó una variante somática poco frecuente para detectar mutaciones, un alineador de lectura dividida para detectar genes de fusión y un análisis interno que cuantifica y ajusta estadísticamente las etiquetas para denominar variación del número de copias (CNV).

[0266] La detección de una mutación puntual mixta en el gen TP53 se muestra en la Figura 19. La frecuencia “esperada” se desvía de la proporción de mezcla porque se sabe que el TP53 es hemicígote en la línea celular NCI-H2228. El software automatizado pudo llamar al alelo mutante en la mezcla 50:1. Se requirió una curación manual para llamar al evento mutante al 1000:1. Con respecto a la especificidad, el filtrado de etiquetas descrito en el ejemplo 1 se aplicó al análisis y no se detectó ninguna otra llamada de mutación en TP53 después de aplicar este filtro de etiquetas.

[0267] Se sabe que la línea celular NCI-H2228 alberga un gen de fusión entre EML4 y ALK; la línea celular sirve como control positivo tanto en los ensayos de hibridación fluorescente *in situ* como en la detección de transcripciones de genes de fusión mediante RT-PCR. No hay informes publicados sobre la ubicación exacta de la unión de fusión de genes. Mediante el uso de una cobertura de sonda densa de la región del intrón 19 de ALK, el análisis de secuencia reveló la ubicación precisa y la secuencia de la unión formada cuando los dos genes se fusionaron (Figura 20). La frecuencia de las lecturas normales frente a las lecturas de unión en la línea celular NCI-H2228 (378 frente a 249, respectivamente) sugiere que el gen de fusión es heterocigoto con una copia normal de ALK.

[0268] La detección de las lecturas de unión como una función de la mezcla se muestra en la Figura 21. Al igual que con la detección de mutaciones puntuales, los valores esperados se ajustaron para reflejar el hecho de que el alelo mutante se encuentra en una copia por genoma diploide. No se detectaron lecturas de fusión en los 1000: 1 muestra mezclada.

[0269] La Figura 22 muestra los resultados de la determinación de la CNV para el gen MYCN como una función de la mezcla. La línea celular NCI-H69 alberga un gen MYCN altamente amplificado. El MYCN se encuentra normalmente como un gen de una sola copia, dos por genoma diploide y, por lo tanto, el resultado esperado para mezclas cada vez más diluidas es que la CNV calculada por la etiqueta debería acercarse asintóticamente a 2 copias (la asíntota está resaltada en la figura). El experimento de validación mostrado en la presente descripción indicó que el sistema de ensayo descrito en esta invención es muy sensible a los genes altamente amplificados.

Descubrimiento de variantes en el ADNlc de pacientes con cáncer

[0270] La validación más rigurosa de la tecnología contemplada en la presente descripción es aplicarla a muestras de ADNlc en las que se desconoce el espectro de mutaciones. Se realizó un análisis secuenciando muestras coincidentes de ADNlc, tumor y tejido adyacente normal (NAT) de dos pacientes con cáncer de ovario. Además, se analizaron dos muestras de ADNlc de pacientes con cáncer colorrectal (CRC) y dos muestras de ADNlc de voluntarios sanos. En ningún caso se detectaron mutaciones, fusiones o anomalías de la CNV en las muestras de voluntarios sanos.

[0271] Las genotecas de ADNlc de los cuatro pacientes con cáncer se examinaron inicialmente usando las sondas dirigidas descritas en la Tabla 6 del Ejemplo 2. Estas sondas se configuraron principalmente para detectar mutaciones puntuales en el gen TP53, fusiones de genes con ALK y amplificaciones del MYCN. Los resultados de este ensayo se muestran en la Figura 23. Se encontró una mutación puntual que se produjo en la misma posición de base en el ADNlc, el tumor, y la NAT de una paciente ovárica. No se encontraron mutaciones en el TP53 en el otro conjunto de muestras coincidentes de pacientes con cáncer de ovario. También se detectaron mutaciones puntuales en las dos genotecas de ADNlc del CRC para las que no estaban disponibles los tejidos coincidentes. Todas estas mutaciones puntuales se han identificado previamente en tumores y se sabe que todas son factores causales de la tumorigénesis. La detección de secuencias mutantes en la genoteca de ADNlc CRC406 del 0,9 % estuvo dentro del alcance de la sensibilidad del ensayo. La sensibilidad se define por la presencia de múltiples familias de lecturas etiquetadas, todas

las cuales poseen la secuencia mutante. Estos datos destacan la utilidad clínica del sistema contemplado en la presente descripción.

[0272] Para explorar más la detección de cambios relacionados con el cáncer en las genotecas de ADNlc y los tejidos asociados, las mismas genotecas se hibridaron con un conjunto de 679 sondas que se dirigen a un total de 20 genes diferentes relacionados con el cáncer (Tabla 18). En este conjunto de sondas, se seleccionaron como objetivo todas las regiones codificantes de 14 genes, mientras que los loci seleccionados se dirigieron a los 6 genes restantes.

Tabla 18. Objetivos de los genes del cáncer

Genes: región codificante diana	Genes: seleccione las regiones diana
BRCA1	ALK
BRCA2	DPYD
BRAF	EPHX1
CDH1	MYC
ERBB2	TNFRSF14
JAK2	ALDH4A1
NF2	
PIK3CA	
RB1	
CDKN2A	
KRAS	
MYCN	
PTEN	
TP53	

[0273] Como se muestra en la Figura 24, la muestra de OVA1, que carecía de cualquier cambio detectable en TP53, portaba una mutación en KRAS que se encontró tanto en el ADNlc como en el tumor correspondiente. Esta observación destaca una característica importante del sistema de ensayo descrito en la presente descripción. Las genotecas creadas a partir de ADNlc se pueden interrogar con cientos (como en este ejemplo) e incluso miles de sondas de direccionamiento. La secuenciación de las genotecas diana resultantes reveló mutaciones somáticas que residen dentro del tumor y no dentro de la línea germinal del individuo afectado. Estos marcadores somáticos asociados a tumores también se pueden usar para cuantificar la cantidad de ADN circulante que se desprende del tumor (a diferencia del ADNlc que tiene una secuencia de línea germinal). Por lo tanto, el descubrimiento de mutaciones, independientemente de su importancia biológica, también proporciona una estimación del contenido tumoral en el ADNlc mezclado.

[0274] Muchas terapias dirigidas tienen más éxito en presencia de genes normales (p. ej., los inhibidores del EGFR solo funcionan en presencia del KRAS silvestre). La evaluación cuantitativa de los niveles de ADN tumoral circulante adquiere especial importancia en estos casos en los que no se encuentran mutaciones en los genes. En otras palabras, la presencia demostrable de ADN tumoral circulante junto con un resultado de secuenciación silvestre en un gen diana particular sugiere que el gen diana es normal dentro del tumor, y tales resultados pueden tener una influencia significativa a la hora de guiar la elección de las terapias. Tal es el caso de la muestra OVA1 resaltada en la Figura 24. La presencia de la mutación KRAS en la genoteca de ADNlc sugirió que el tumor de este paciente alberga uno o más genes TP53 silvestres.

[0275] Otro ejemplo de descubrimiento de genes aberrantes se muestra en la Figura 25. El sistema de análisis genético cuantitativo dirigido reveló la presencia de una amplificación significativa en el gen ERBB2, también denominado HER-2/neu. Si bien este tipo de amplificación se ha publicitado mucho en el cáncer de mama, en ocasiones también se identifica en los carcinomas colorrectales.

Conclusiones

[0276] Los experimentos de validación con el ADN de líneas celulares revelaron los umbrales de detección de tres tipos de variación genética que son fundamentales para impulsar el crecimiento neoplásico en los cánceres. La caracterización del ADNlc derivado de pacientes con cáncer reveló cambios genéticos asociados a tumores que estaban muy por encima de los umbrales establecidos por los experimentos de reconstrucción en las cuatro muestras

analizadas. Estos datos indicaron que el sistema de análisis cuantitativo contemplado en la presente descripción puede tener una utilidad clínica significativa, especialmente en entornos en los que las biopsias líquidas son más apropiadas.

REIVINDICACIONES

1. Un método para el análisis genético dirigido del ADN libre de células (ADNlc) que comprende:
 - (a) tratar el ADNlc con una o más enzimas reparadoras terminales para generar ADNlc reparado terminalmente y ligar un módulo adaptador a cada extremo del ADNlc reparado terminalmente para generar una genoteca de ADNlc;
 - (b) amplificar la genoteca de ADNlc para generar clones de la genoteca de ADNlc y determinar el número de equivalentes genómicos en la genoteca de clones de ADNlc; y
 - (c) realizar un análisis genético cuantitativo de un locus genético diana en los clones de la genoteca de ADNlc que comprenden:
 - (i) hibridar uno o más módulos de sonda de captura con un locus genético diana en la genoteca de clones de ADNlc para formar uno o más complejos de clones de ADNlc del módulo de sonda de captura, en donde el locus genético diana es una región de un gen, en donde los módulos de sonda de captura comprenden sondas de captura que tienen de 25 a 45 nucleótidos de longitud e hibridan de manera comparable, igual de bien, o mejor que las sondas de captura de 60 nucleótidos de longitud;
 - (ii) aislar el uno o más complejos aislados de módulo de sonda de captura y clon de ADNlc de (i);
 - (iii) procesar enzimáticamente el uno o más complejos de clones de ADNlc del módulo de sonda de captura aislados de (ii), en donde el procesamiento enzimático comprende realizar una extensión 5'-3' del ADN polimerasa de la sonda de captura usando el clon de ADNlc aislado en el complejo como plantilla;
 - (iv) realizar PCR en el complejo procesado enzimáticamente de (iii) para generar moléculas de ácido nucleico híbridas amplificadas, en donde las moléculas de ácido nucleico híbridas amplificadas comprenden una secuencia diana capaz de hibridarse con la sonda de captura y el complemento de la secuencia de cola del módulo de la sonda de captura; y
 - (v) secuenciar las moléculas de ácido nucleico híbridas amplificadas de (iv) para generar una pluralidad de lecturas de secuenciación y realizar análisis bioinformáticos en la pluralidad de lecturas de secuenciación.
2. El método de la reivindicación 1, en donde el módulo adaptador comprende un sitio de unión al cebador para la amplificación de la genoteca de ADNlc.
3. El método de la reivindicación 1 o la reivindicación 2, en donde el módulo adaptador comprende uno o más códigos de lectura únicos.
4. El método de una cualquiera de las reivindicaciones 1 a 3, en donde el módulo adaptador comprende una o más secuencias para la secuenciación del ADN.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en donde el módulo adaptador comprende uno o más códigos de muestra para la multiplexación de muestras.
6. El método de la reivindicación 1, en donde una pluralidad de módulos de sondas de captura se hibridan con el locus genético diana en la genoteca de clones de ADNlc.
7. El método de la reivindicación 1, en donde el análisis genético dirigido se realiza en una pluralidad de loci genéticos diana.
8. El método de una cualquiera de las reivindicaciones 1 a 7, en donde las sondas de captura tienen una longitud de 25 nucleótidos, 26 nucleótidos, 27 nucleótidos, 28 nucleótidos, 29 nucleótidos, 30 nucleótidos, 31 nucleótidos, 32 nucleótidos, 33 nucleótidos, 34 nucleótidos, 35 nucleótidos, 36 nucleótidos, 37 nucleótidos, 38 nucleótidos, 39 nucleótidos, 40 nucleótidos, 41 nucleótidos, 42 nucleótidos, 43 nucleótidos, 44 nucleótidos o 45 nucleótidos.
9. El método de una cualquiera de las reivindicaciones 1 a 8, en donde las sondas de captura tienen una longitud de 35 a 45, 36 a 44, 37 a 43, o 38 a 42, o 39 a 41 nucleótidos.
10. El método de una cualquiera de las reivindicaciones 1 a 8, en donde las sondas de captura tienen 40 nucleótidos.
11. El método de la reivindicación 1, en donde el análisis bioinformático se usa para detectar una o más lesiones genéticas dentro del locus genético diana.

12. El método de la reivindicación 11, en donde una o más lesiones genéticas comprenden una transición o transversión de nucleótidos, una inserción o eliminación de nucleótidos, un reordenamiento genómico, o un cambio en el número de copias.
13. El método de una cualquiera de las reivindicaciones 1 a 12, en donde el ADNlc se aísla de una muestra biológica de un sujeto seleccionado del grupo que consiste en líquido amniótico, sangre, plasma, suero, semen, líquido linfático, líquido cefalorraquídeo, líquido ocular, orina, saliva, heces, mucosa y sudor.
14. El método de una cualquiera de las reivindicaciones 1 a 13, en donde el análisis genético cuantitativo se usa para identificar o detectar una o más lesiones genéticas que causan o están asociadas a una enfermedad genética.
15. El método de la reivindicación 14, en donde la enfermedad genética es cáncer.
16. El método de una cualquiera de las reivindicaciones 1 a 15, en donde al menos un módulo de sonda de captura se hibrida corriente abajo de una región diana dentro del locus genético diana y al menos un módulo de sonda de captura se hibrida corriente arriba de la región diana.
17. El método de una cualquiera de las reivindicaciones 1 a 16, en donde una pluralidad de sondas de captura se hibrida con el locus genético diana, en donde cada una de la pluralidad de sondas de captura se hibrida con el locus genético diana dentro de aproximadamente 200 bp de cualquier otra sonda de captura.

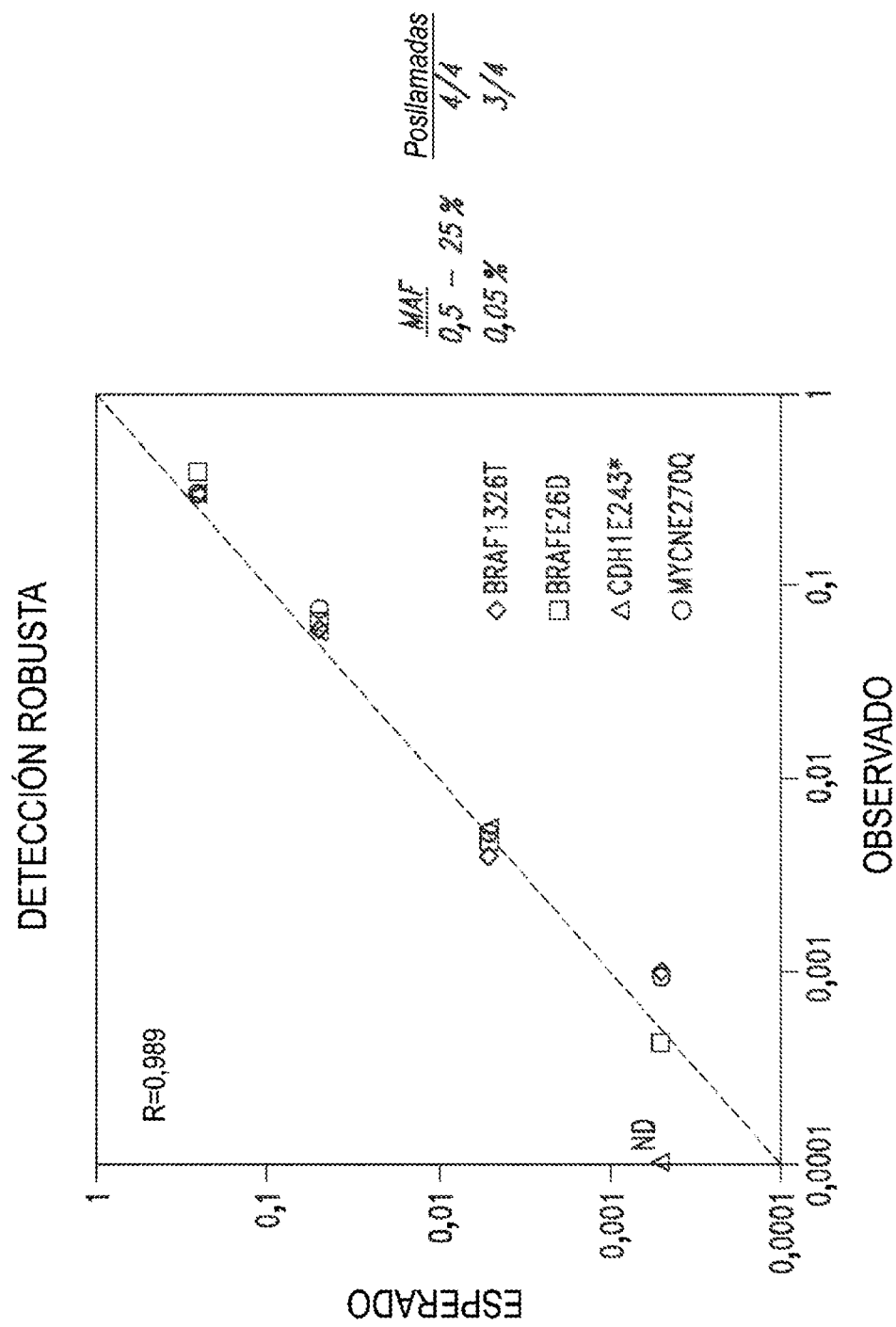


Figura 1

BRAF 1326T SNV

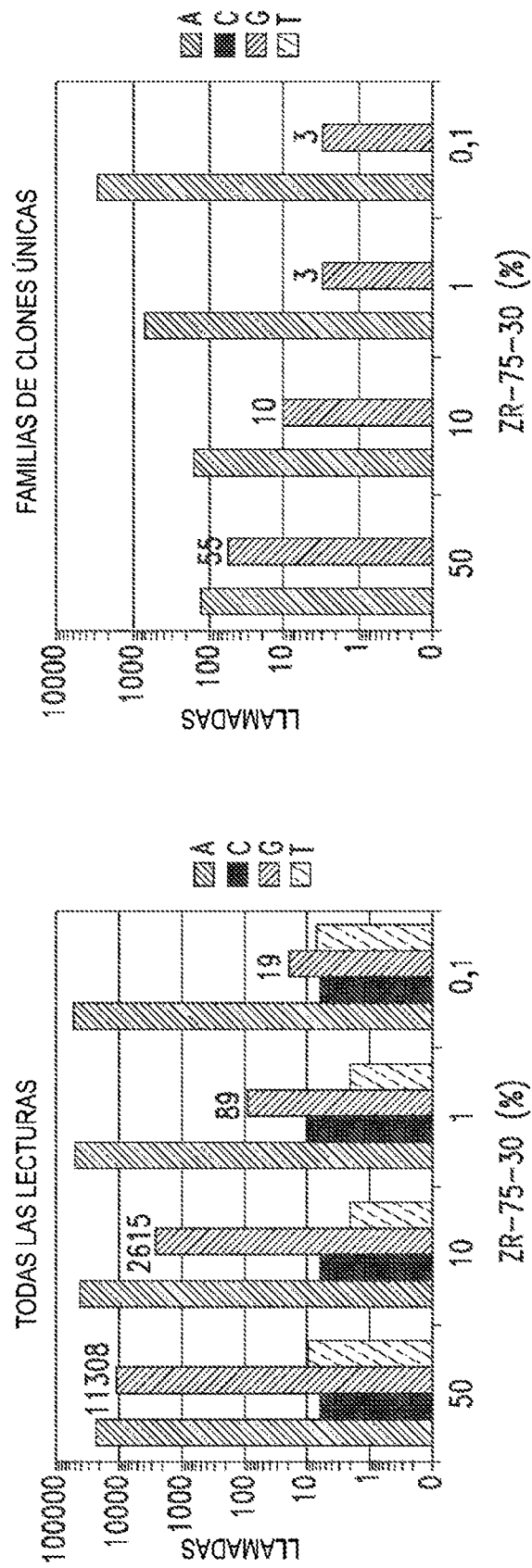


Figura 2

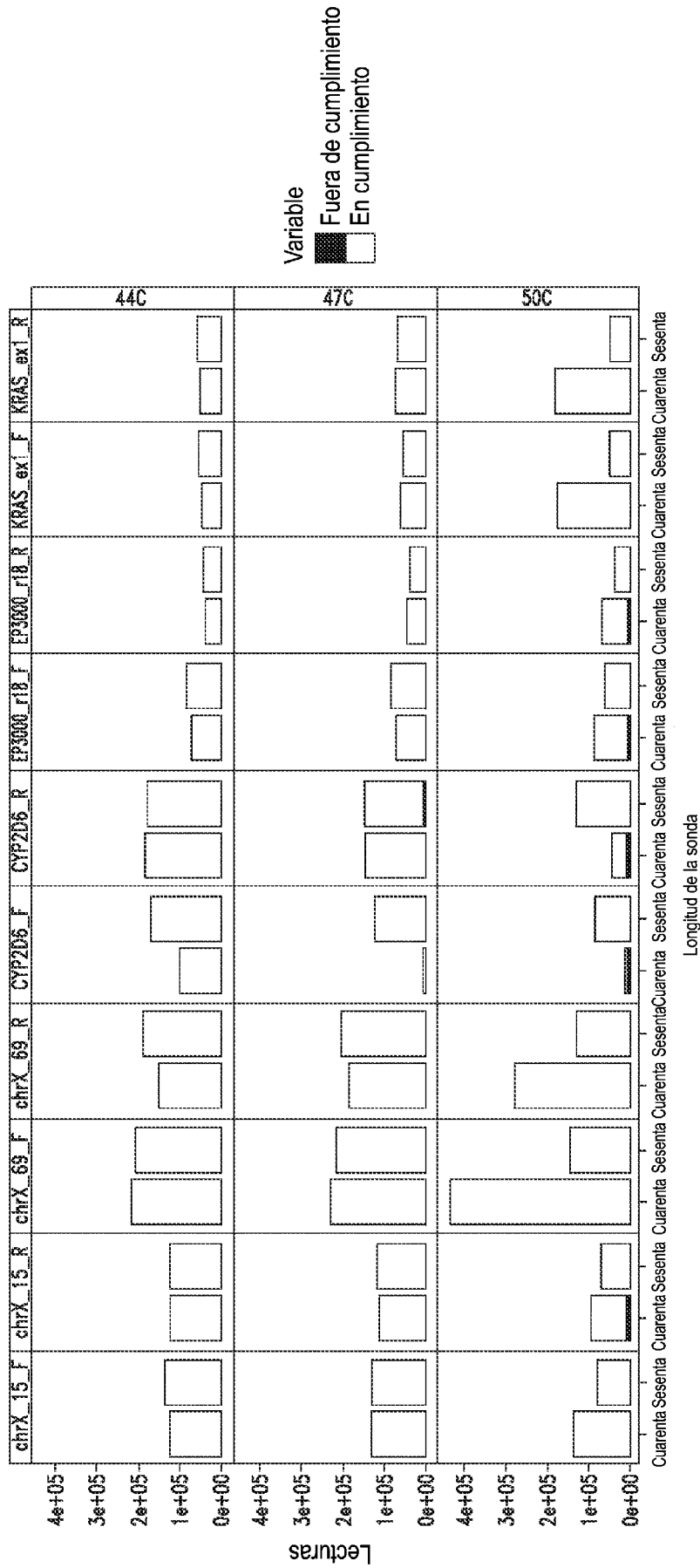


Figura 3

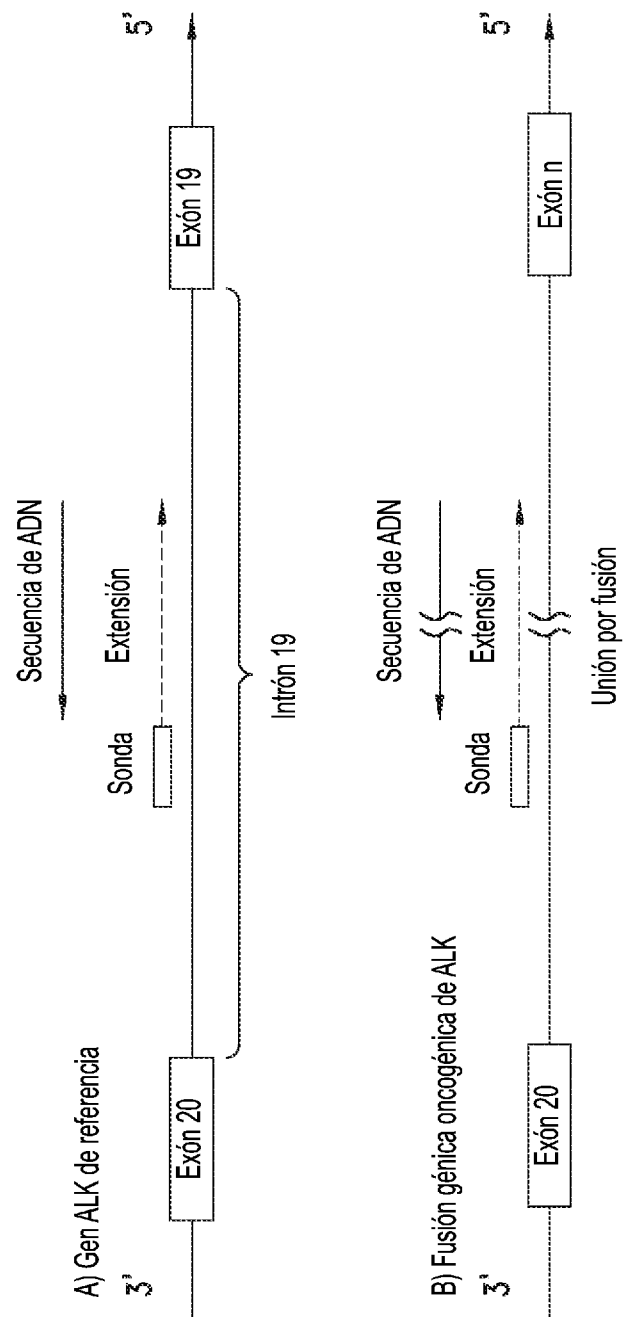


Figura 4

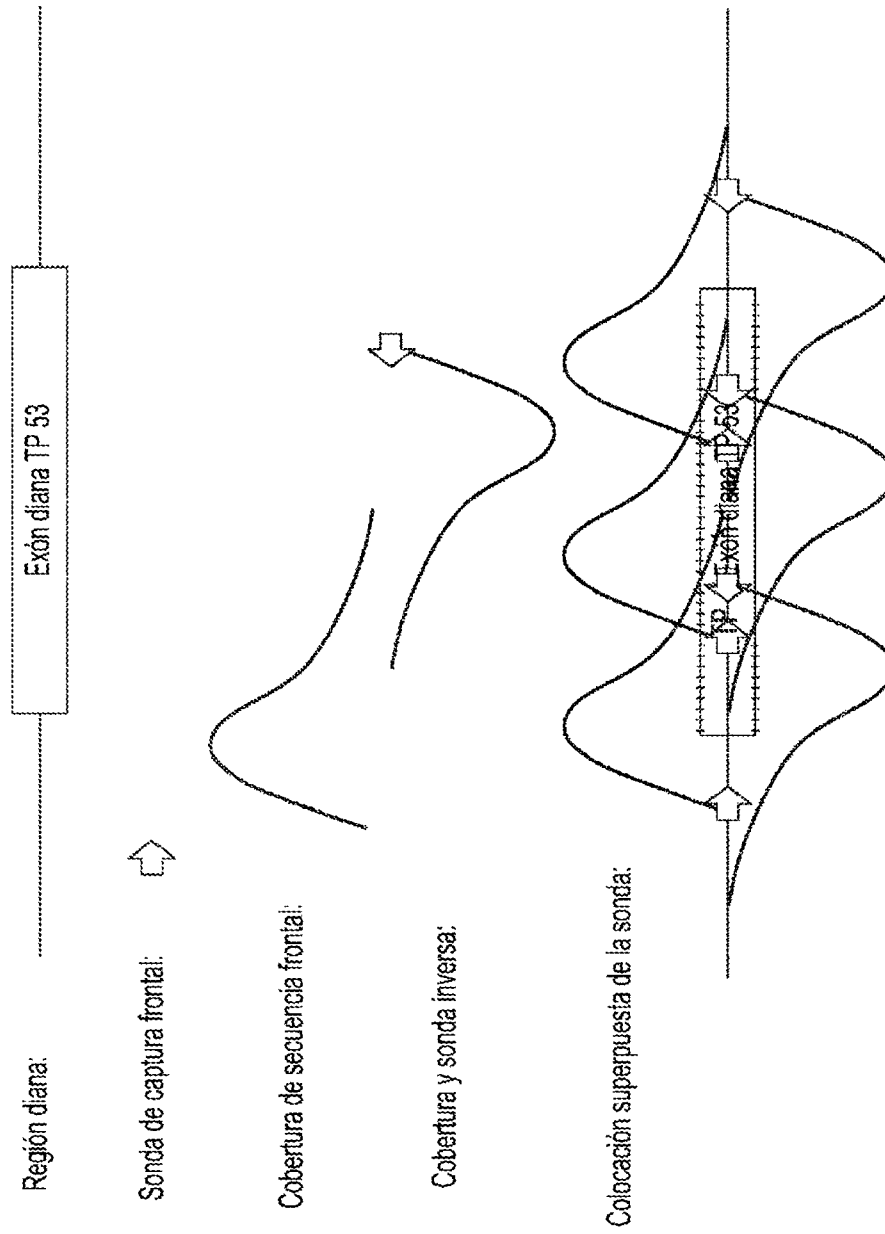


Figura 5

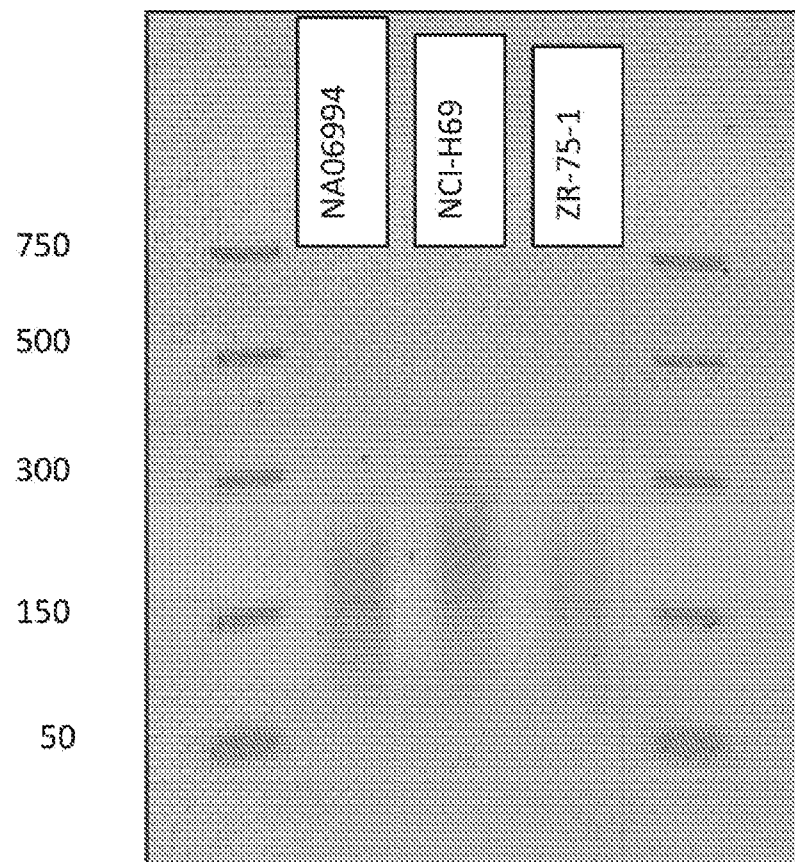
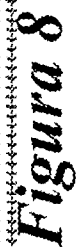


Figura 6





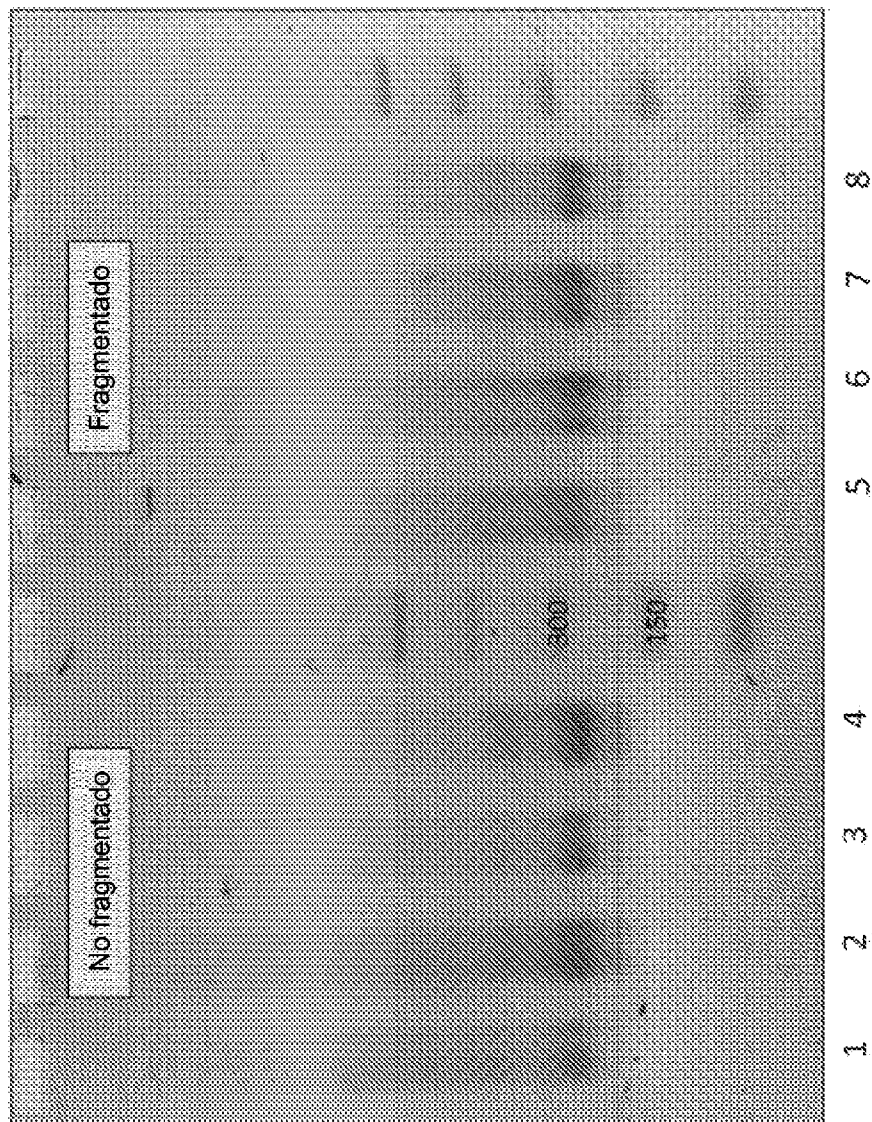


Figura 9A

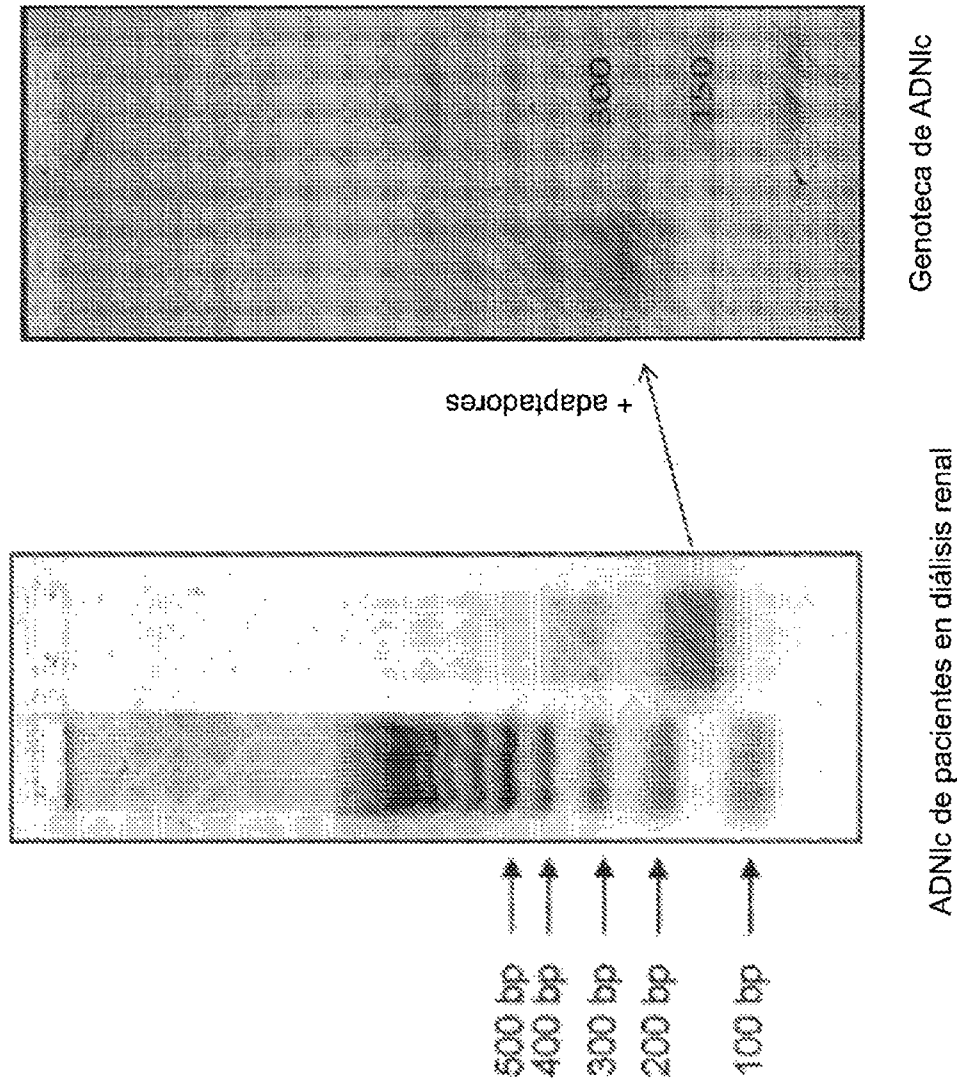


Figura 9B

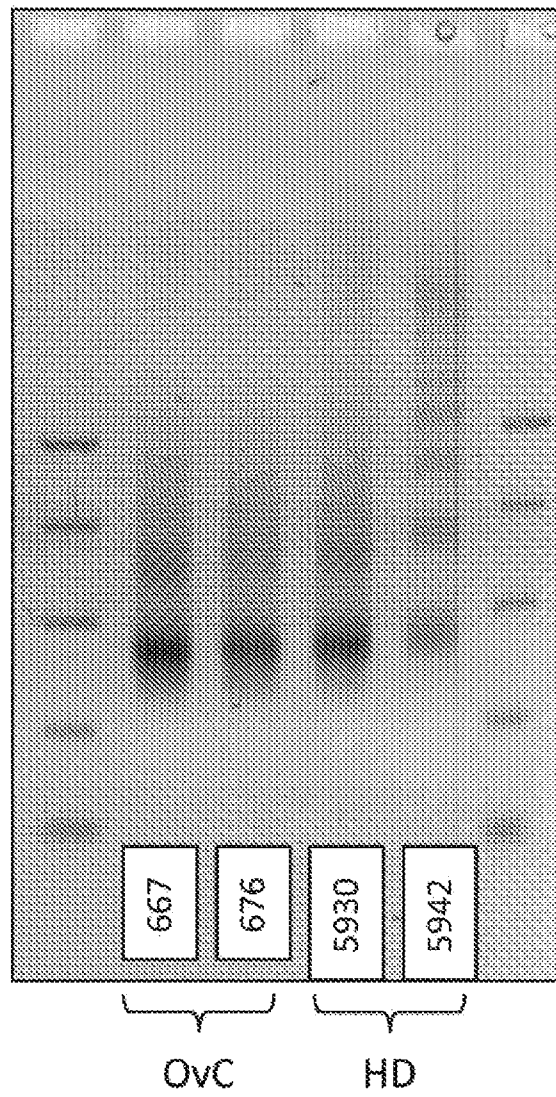


Figura 9C

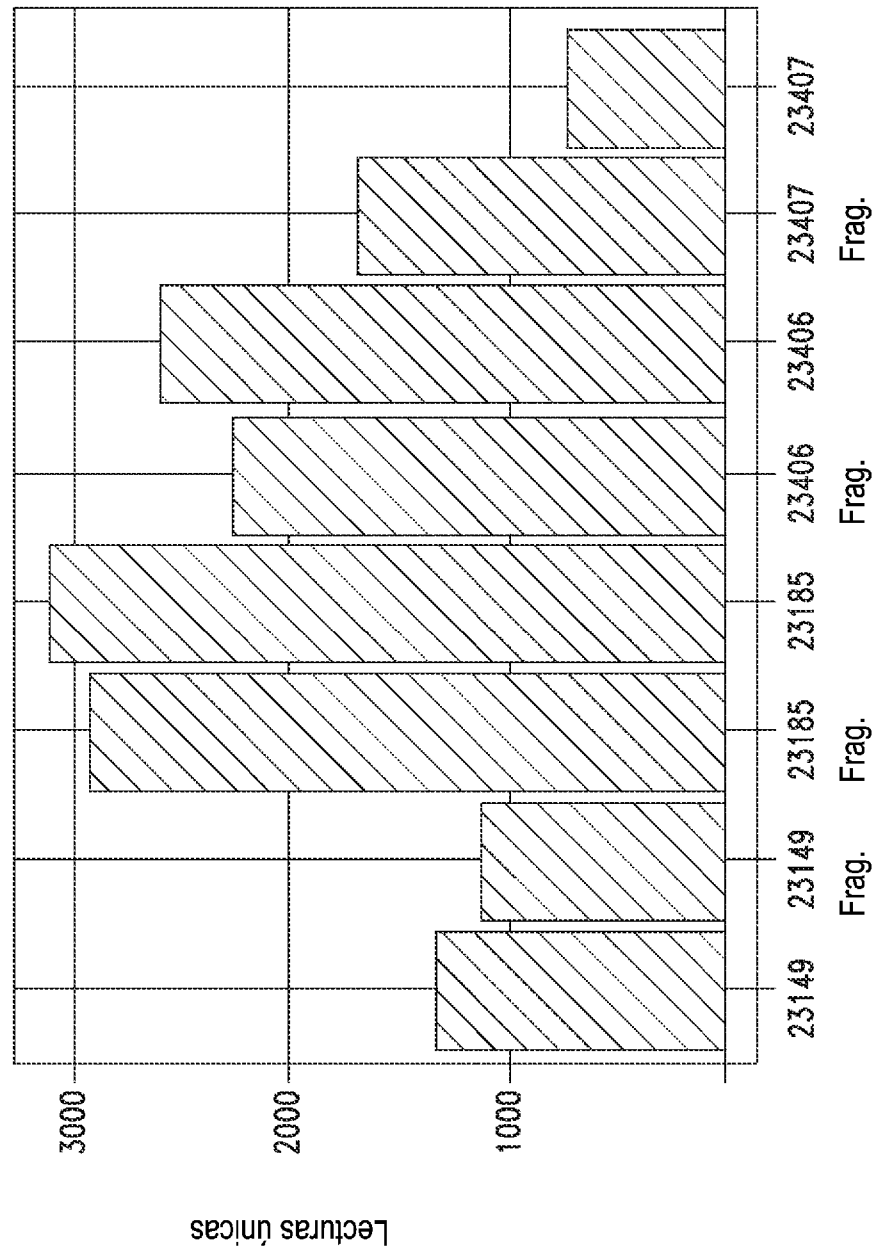


Figura 10

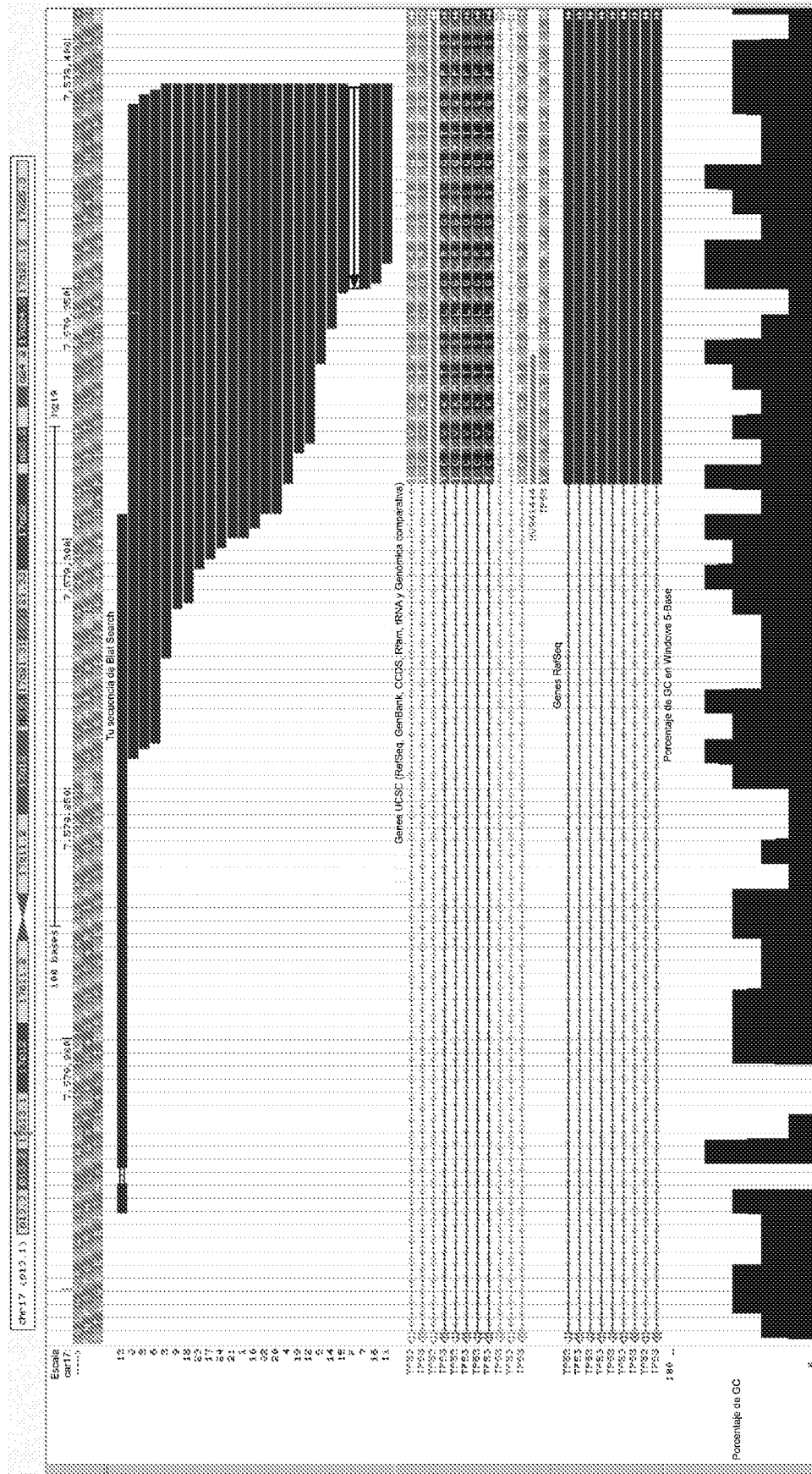


Figura 11

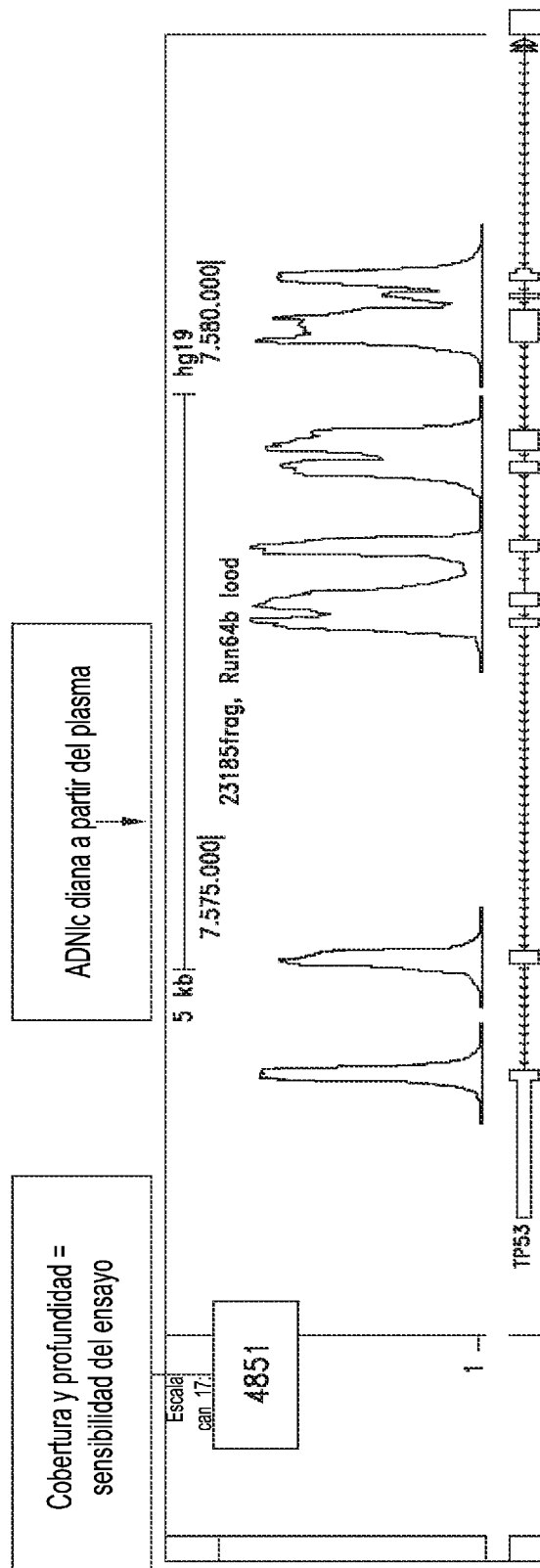


Figura 12

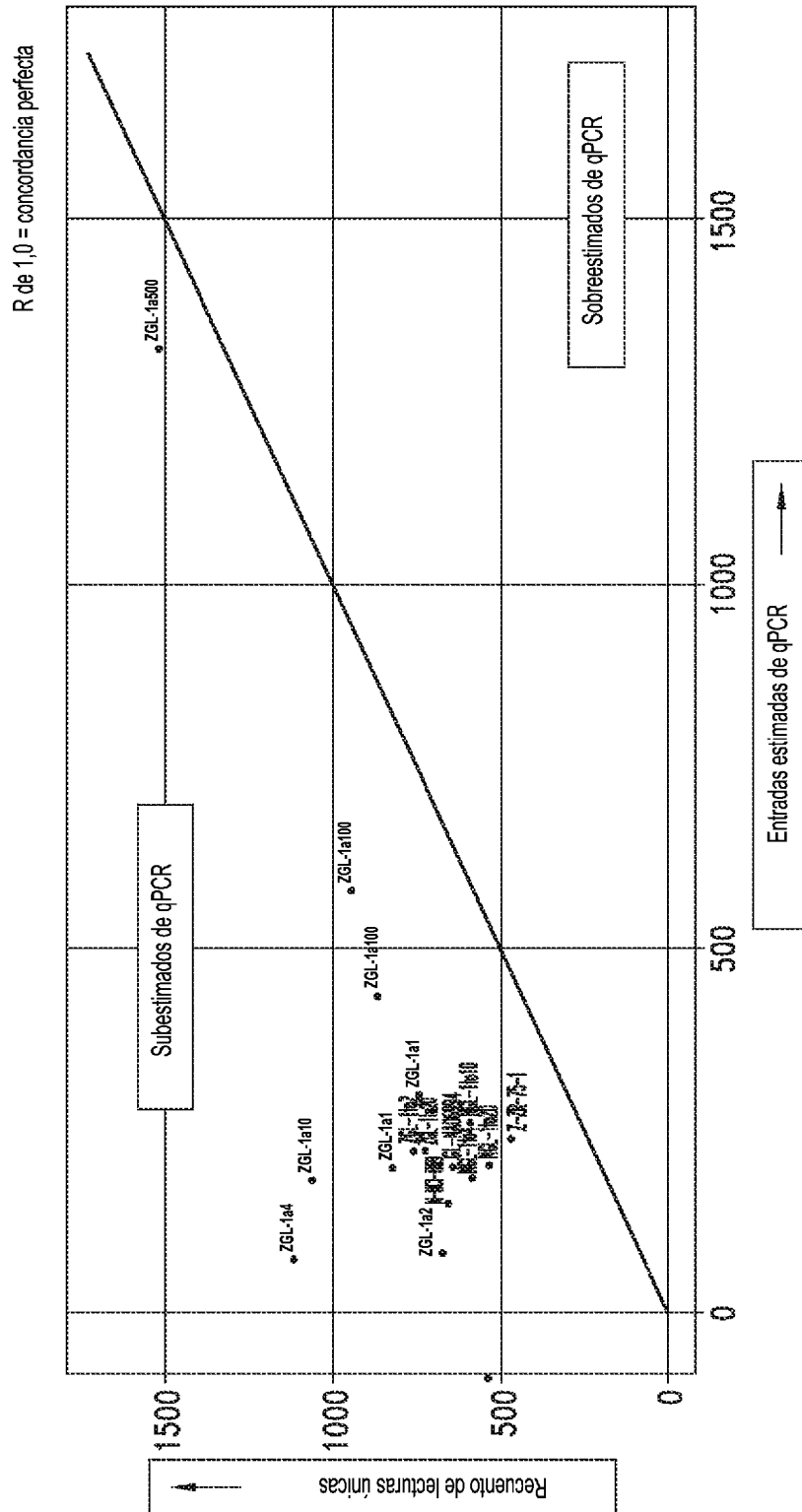


Figura 13

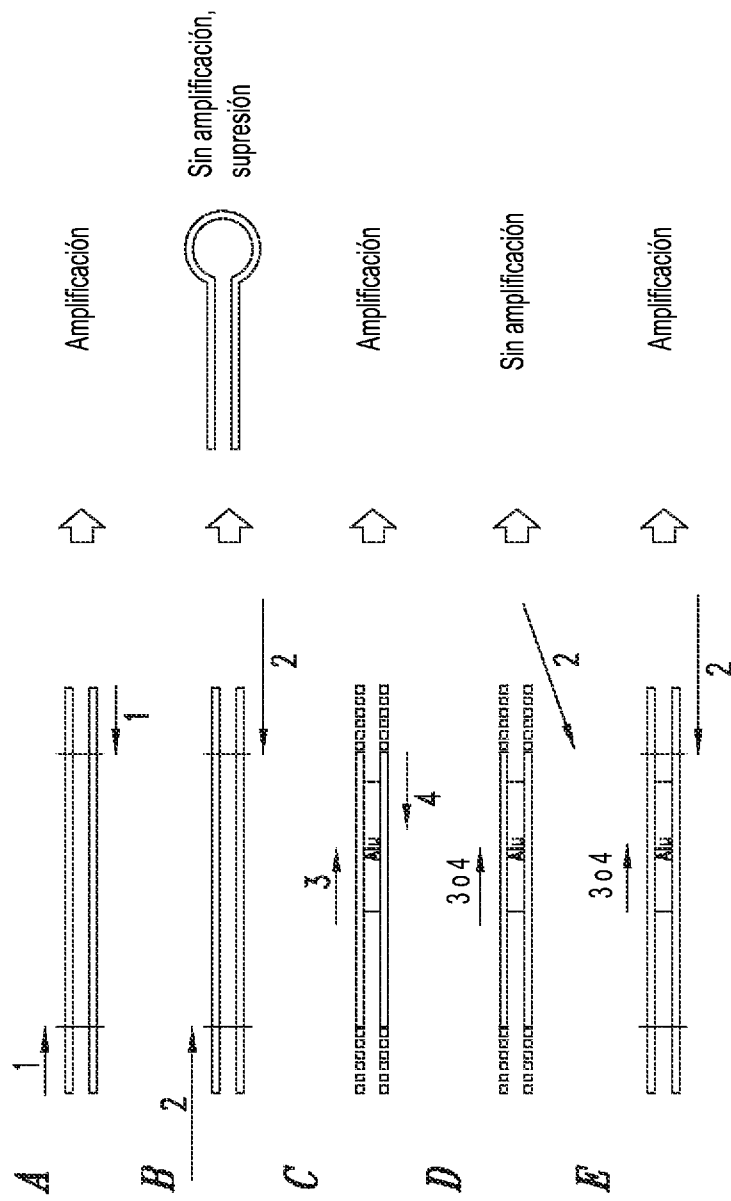
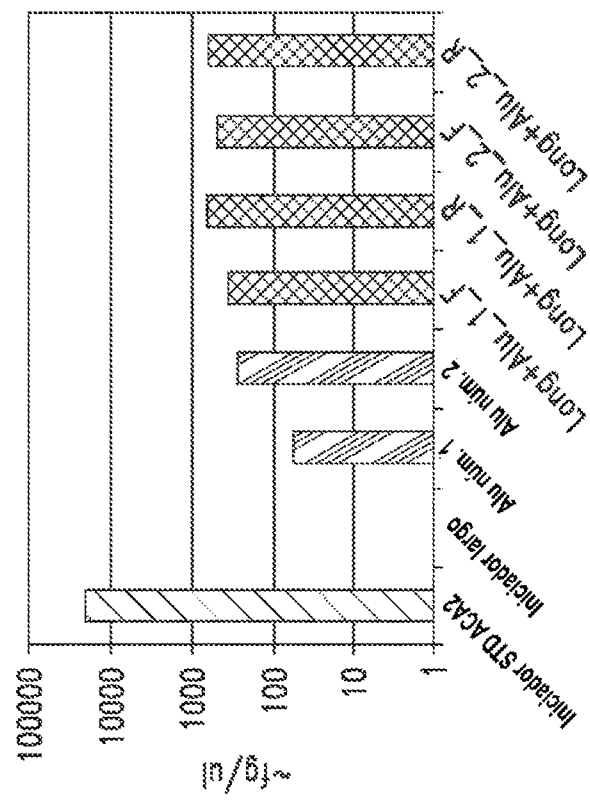
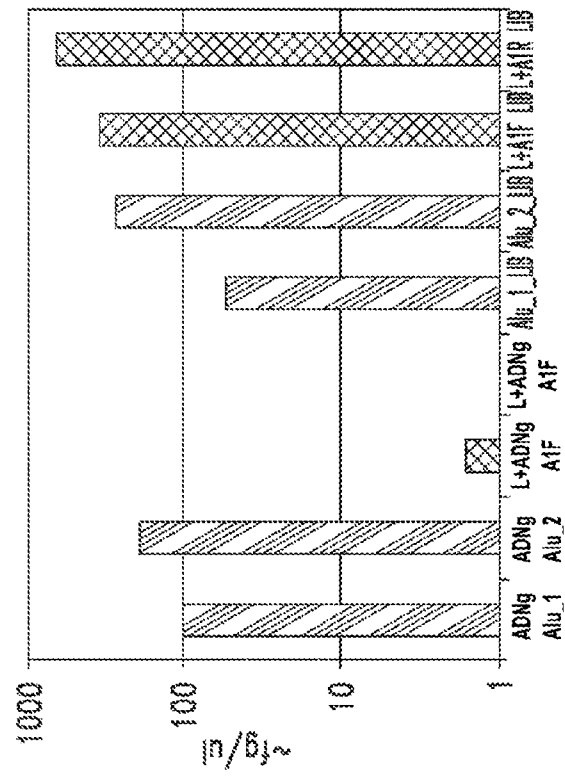


Figura 14



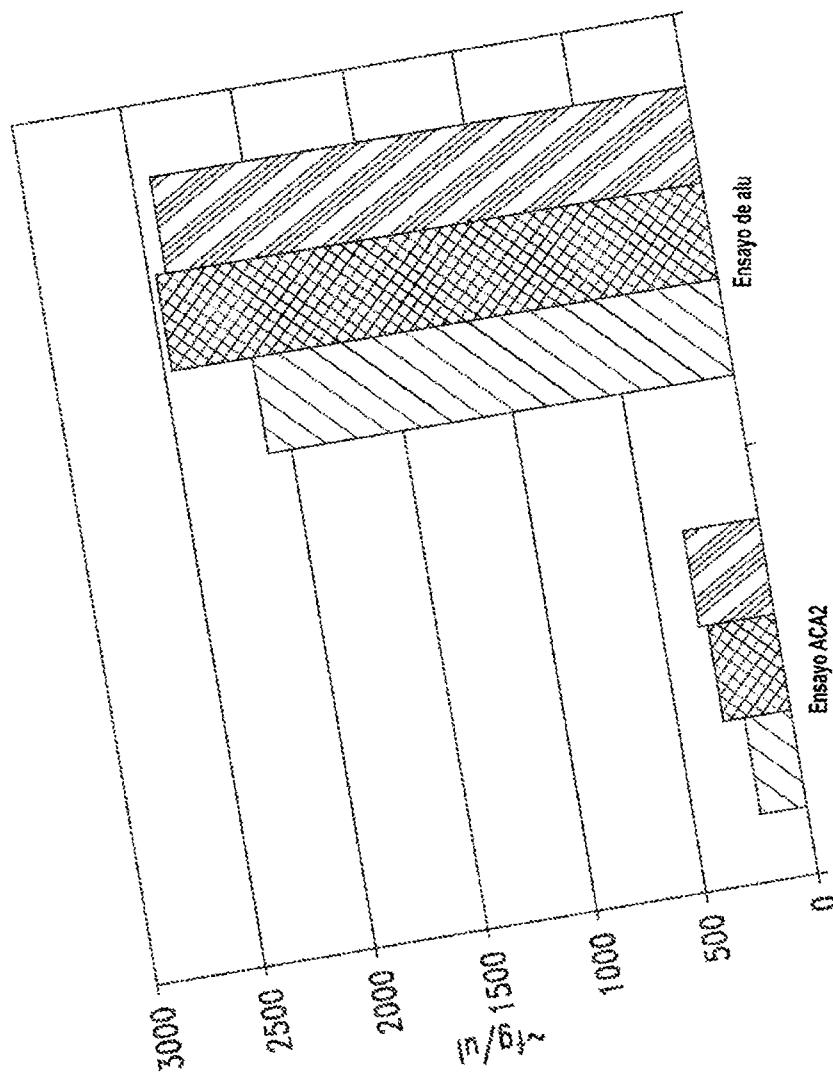


Figura 16

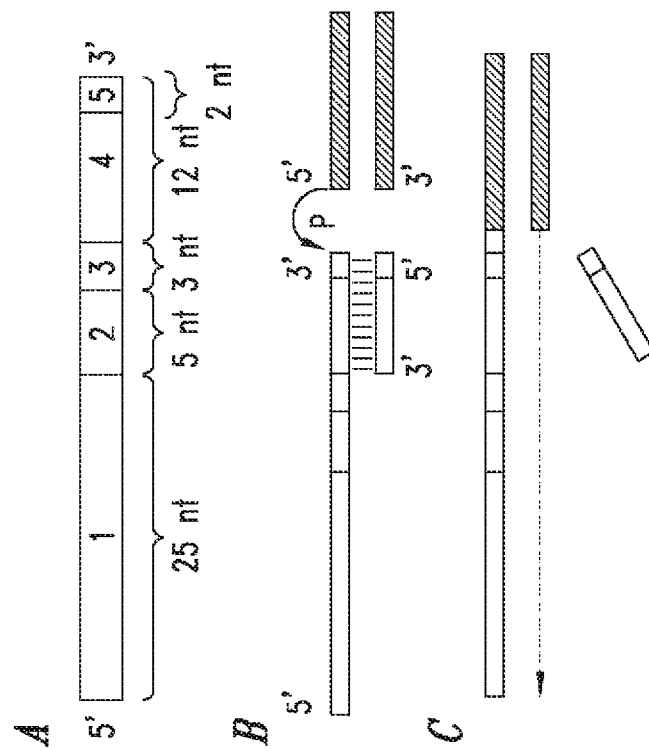


Figura 17

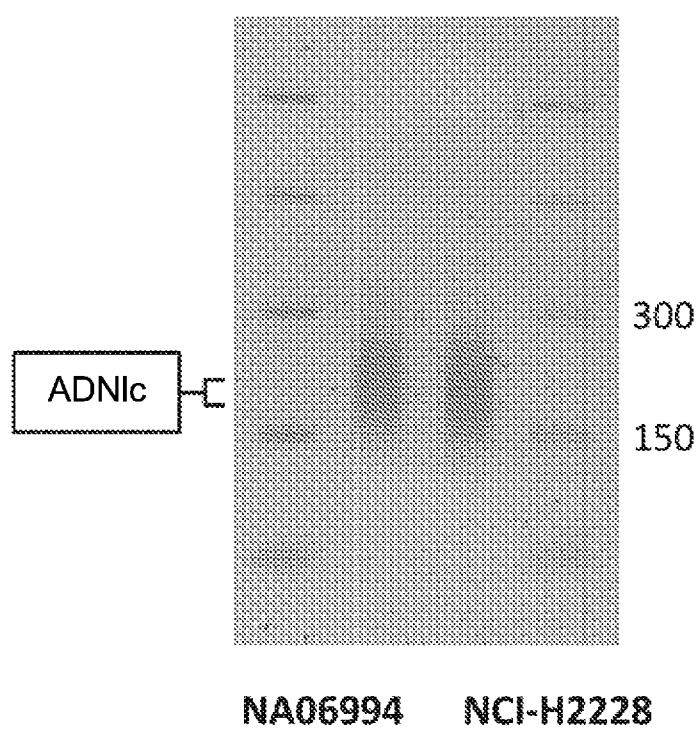


Figura 18

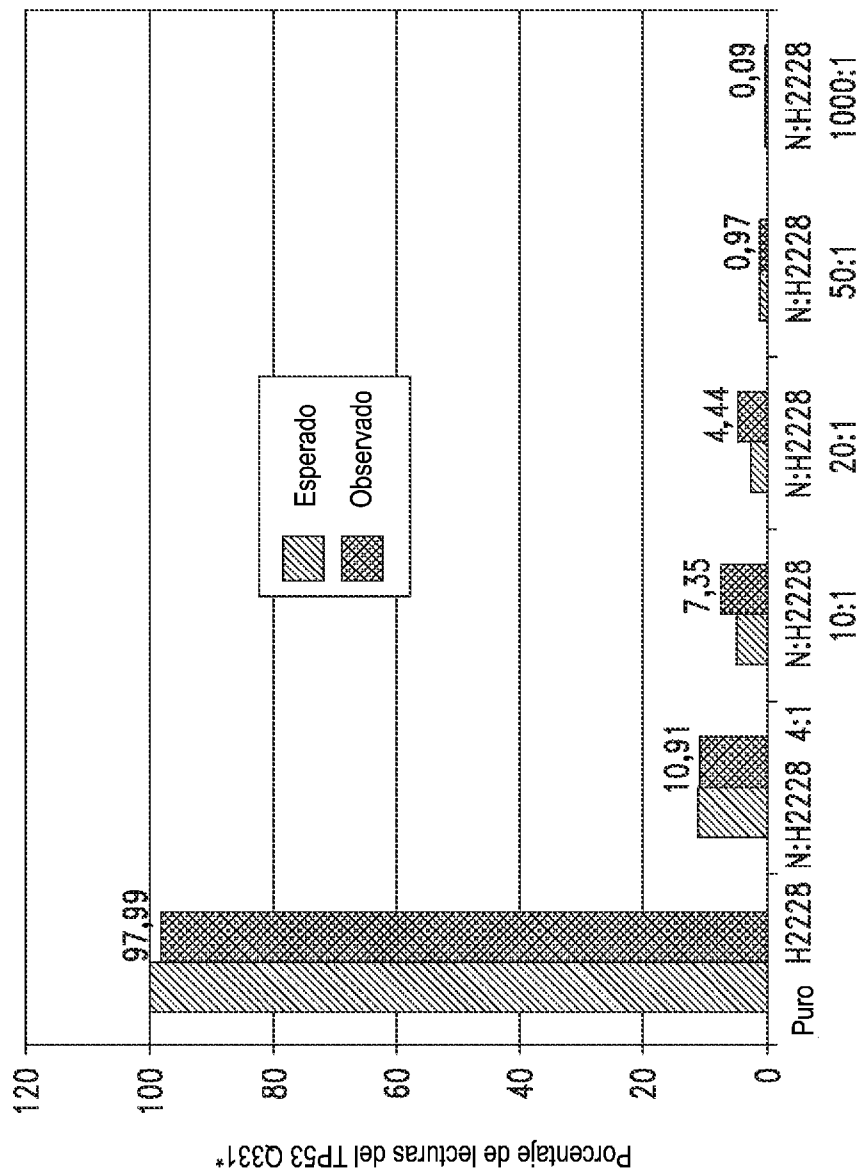


Figura 19

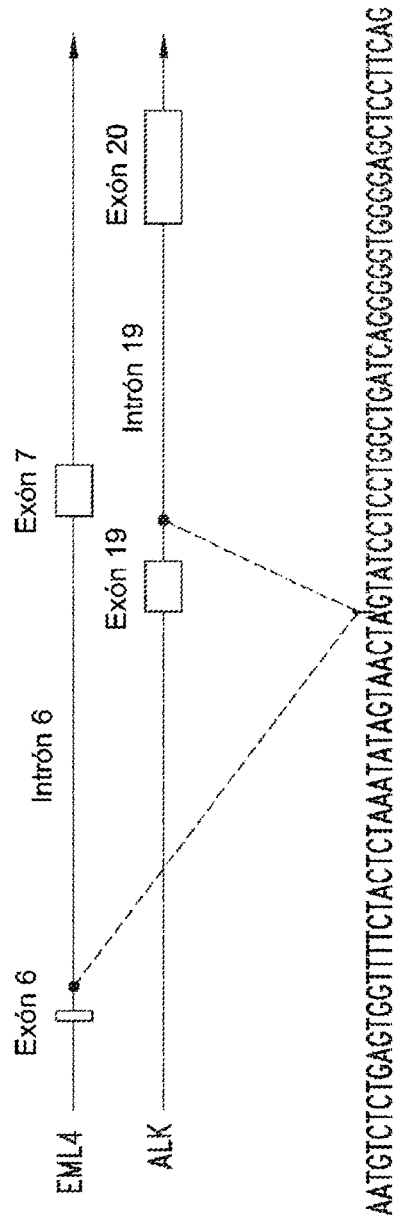


Figura 20

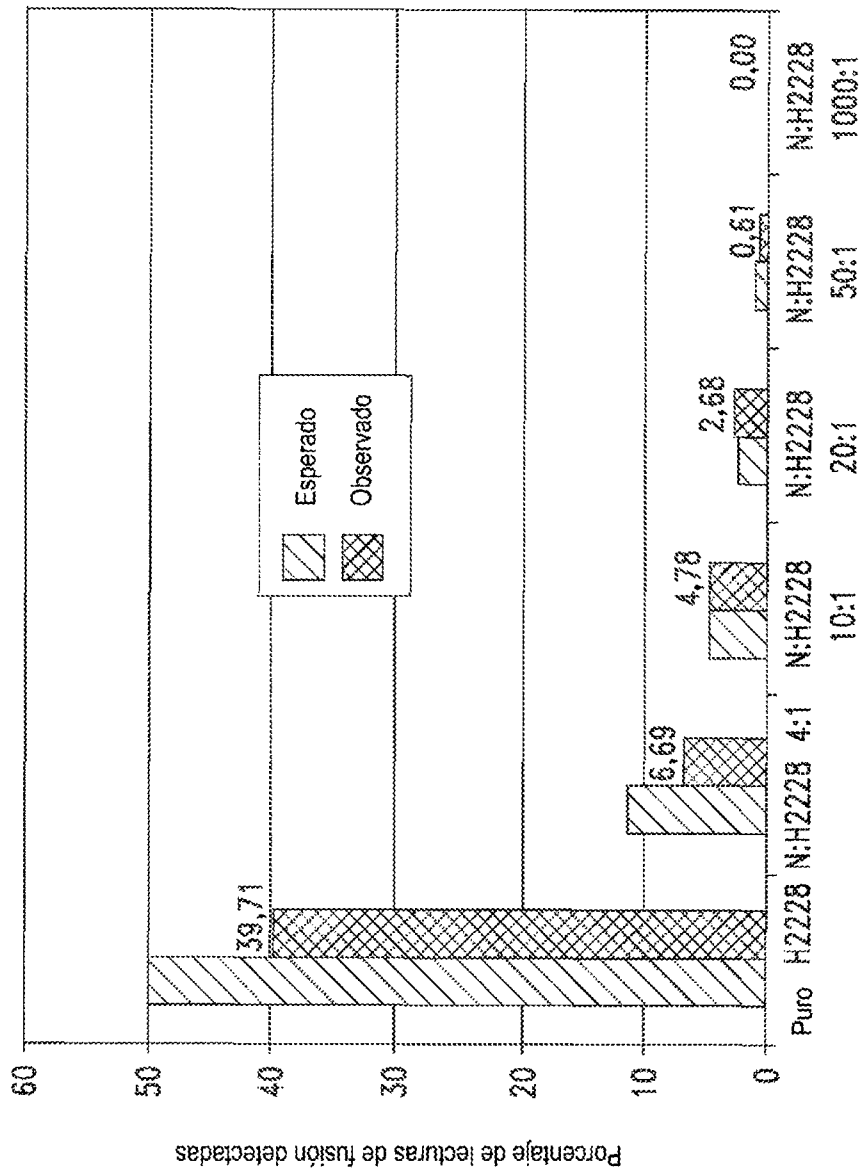


Figura 21

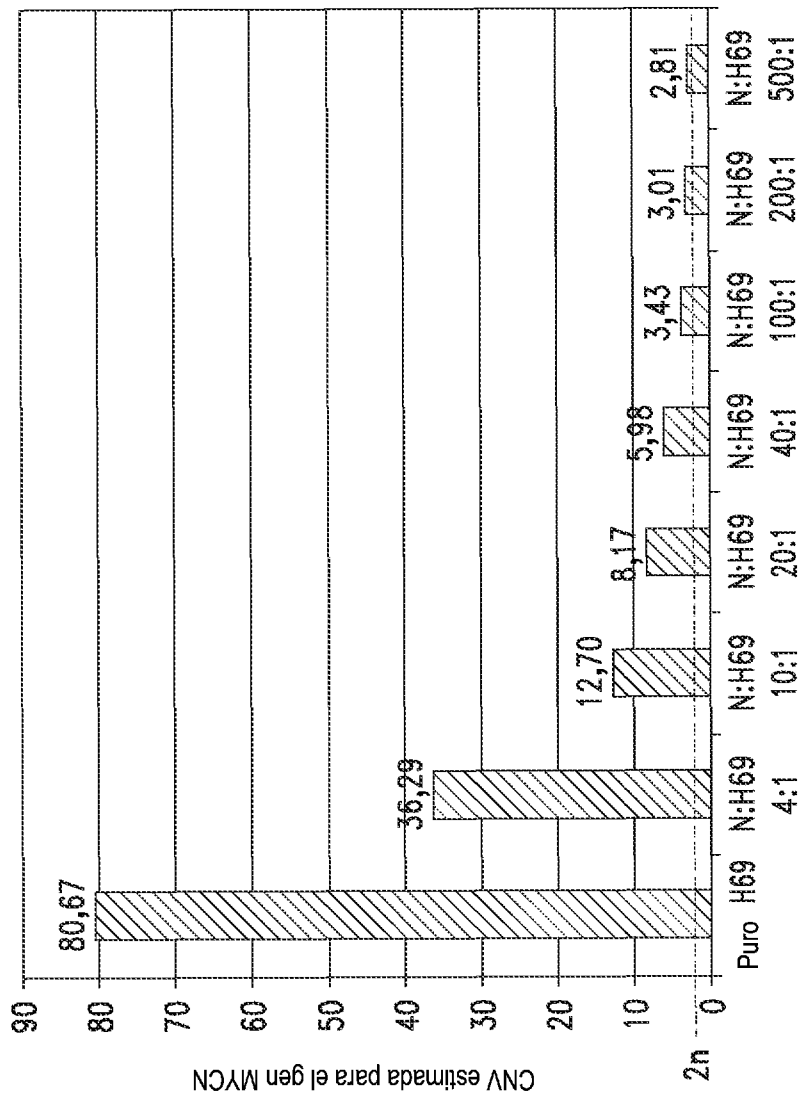


Figura 22

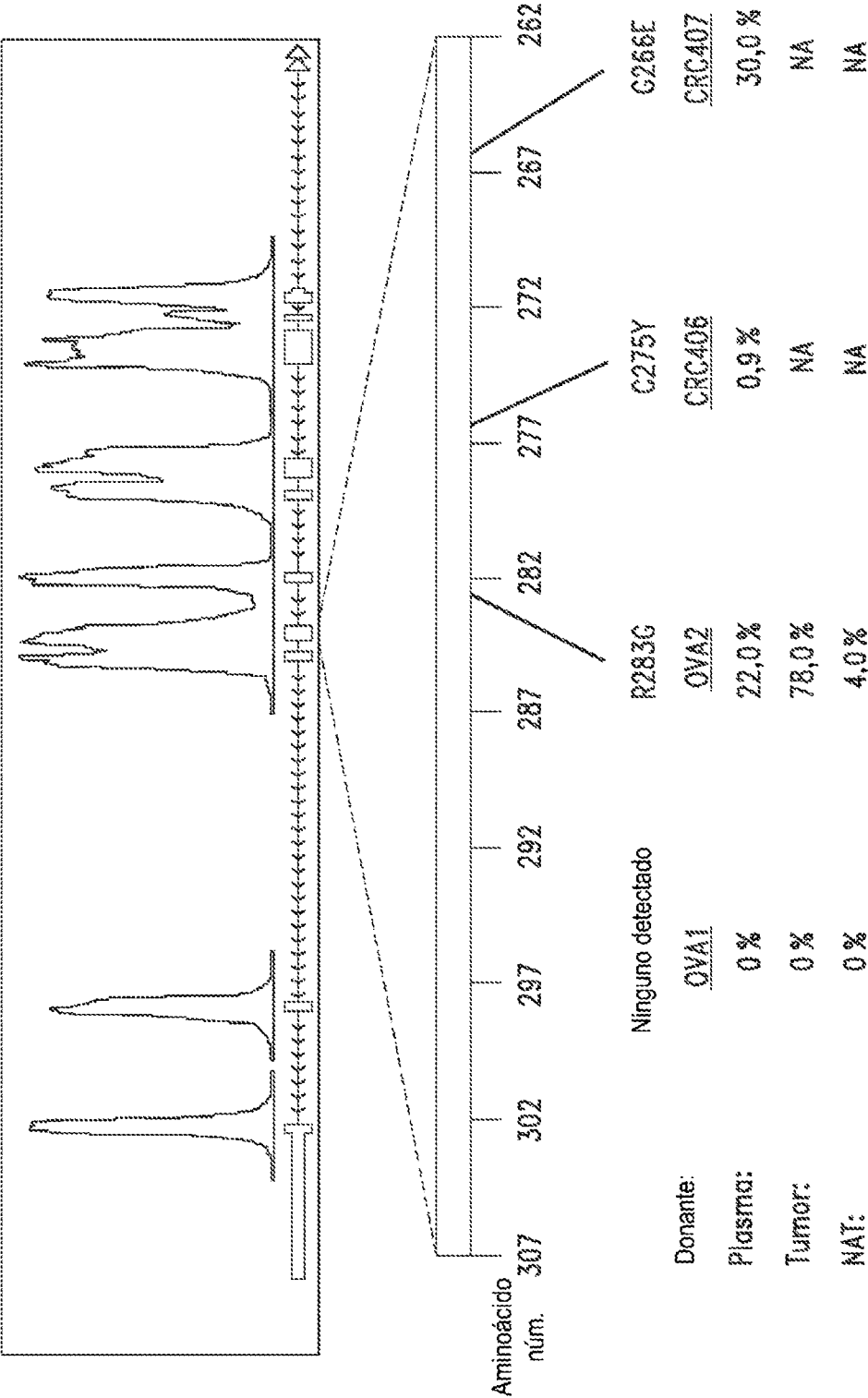


Figura 23

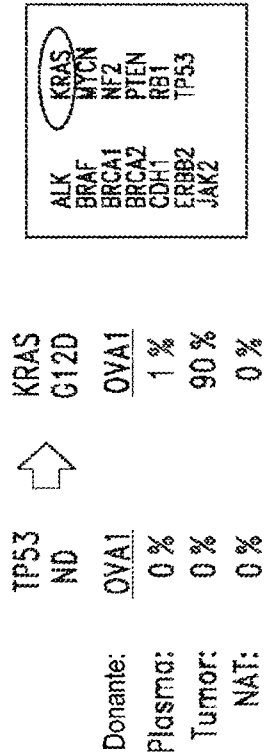


Figura 24

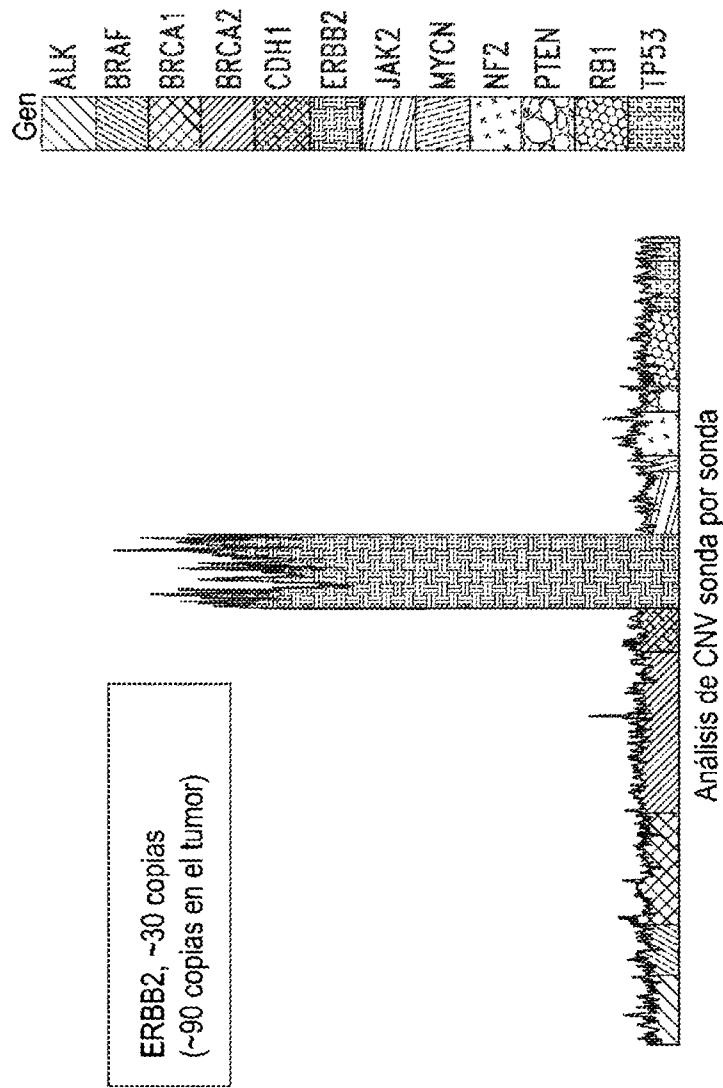


Figura 25