

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 276**

51 Int. Cl.:

C12Q 1/68 (2008.01)
C40B 20/00 (2006.01)
C40B 20/04 (2006.01)
C40B 40/06 (2006.01)
C40B 50/00 (2006.01)
A61K 39/395 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.12.2015** **E 22168503 (5)**

97 Fecha y número de publicación de la concesión europea: **24.07.2024** **EP 4112738**

54 Título: **Análisis multigénico de muestras tumorales**

30 Prioridad:

05.12.2014 US 201462088457 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
25.11.2024

73 Titular/es:

FOUNDATION MEDICINE, INC. (100.0%)
150 Second Street
Cambridge, MA 02141, US

72 Inventor/es:

OTTO, GEOFFREY ALAN;
NAHAS, MICHELLE y
LIPSON, DORON

74 Agente/Representante:

ILLESCAS TABOADA, Manuel

Observaciones:

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

ES 2 989 276 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

Análisis multigénico de muestras tumorales

5 Campo de la invención

La invención se define por las reivindicaciones. La presente divulgación se refiere en general a procedimientos optimizados para analizar ácidos nucleicos de muestras tumorales, por ejemplo, procedimientos que tienen una selección de ácidos nucleicos, alineación de lectura y llamada de mutaciones optimizadas integradas.

10 Antecedentes de la invención

Las pruebas genómicas de células cancerosas se muestran significativamente prometedoras para lograr una mejor comprensión de los cánceres y una mejor gestión de enfoques de tratamiento más eficaces. Las pruebas genómicas implican la secuenciación del genoma de las células cancerosas de un paciente y la identificación de cualquier alteración genómica en dichas células. Por ejemplo, el documento WO 2012/092426 describió un procedimiento para analizar una muestra tumoral que incluye adquirir una colección que comprenda una pluralidad de miembros tumorales de la muestra, poner en contacto la colección con un conjunto de cebos (del inglés "*bait set*") para aislar miembros seleccionados, adquirir una lectura para un intervalo subgenómico de un miembro seleccionado, alinear dicha lectura y asignar un valor de nucleótido a partir de dicha lectura para una posición nucleotídica preseleccionada. El documento WO 2013/028817 describe moléculas de fusión de RET y usos de las mismas. El documento WO 2014/130975 describe polinucleótidos de fusión y polipéptidos de fusión asociados con el cáncer, en particular melanoma, así como sus usos como dianas terapéuticas y diagnósticas. Las alteraciones genómicas pueden incluir, por ejemplo, mutaciones, sustituciones de bases, reordenamientos, deleciones, inserciones, entre otras opciones. Comprender estas alteraciones genómicas a medida que se encuentran en el cáncer de un paciente específico también puede ayudar a desarrollar mejores tratamientos y ayudar a identificar los mejores enfoques para tratar variantes de cáncer específicas usando información de alteración genómica.

30 Sumario de la invención

La presente invención se define por las reivindicaciones. En consecuencia, la presente invención se refiere a un procedimiento para evaluar un sujeto, que comprende:

- (a) adquirir una primera colección de ácidos nucleicos que comprenda ADN genómico de una muestra del sujeto;
- (b) poner en contacto la primera colección de ácidos nucleicos con un primer conjunto de cebos para proporcionar una primera pluralidad de miembros seleccionados que comprendan un intervalo subgenómico o una porción del mismo;
- (c) amplificar cada miembro de la primera pluralidad de miembros seleccionados;
- (d) adquirir una segunda colección de ácidos nucleicos que comprenda ADNc obtenido a partir de ARN de la muestra;
- (e) poner en contacto la segunda colección de ácidos nucleicos con un segundo conjunto de cebos para proporcionar una segunda pluralidad de miembros seleccionados que comprendan un intervalo subgenómico expresado o una porción del mismo, en el que el intervalo subgenómico expresado incluya al menos un gen de anticuerpo, un gen de receptor de la superfamilia de las inmunoglobulinas (Ig) o un fragmento de los mismos;
- (f) amplificar cada miembro de la segunda pluralidad de miembros seleccionados;
- (g) adquirir la secuencia del intervalo subgenómico y el intervalo subgenómico expresado; evaluando de este modo al sujeto.

De acuerdo con un modo de realización preferente, el intervalo subgenómico expresado es un gen del receptor de linfocitos B (BCR) o un gen del receptor de linfocitos T (TCR), o un fragmento de los mismos.

De acuerdo con otro modo de realización preferente, el intervalo subgenómico es un gen de anticuerpo, un gen de receptor de la superfamilia de las inmunoglobulinas (Ig) o un fragmento de los mismos.

De acuerdo con otro modo de realización preferente, el intervalo subgenómico es un gen del receptor de linfocitos B (BCR) o un gen del receptor de linfocitos T (TCR), o un fragmento de los mismos.

De acuerdo con un modo de realización preferente, el intervalo subgenómico comprende una secuencia de un receptor de la superfamilia de las inmunoglobulinas (Ig), segmento V, un segmento D, un segmento J o una secuencia de la tabla 10.

Tabla 10
secuencia que codifica un anticuerpo;
secuencia que codifica una cadena ligera de anticuerpo;
secuencia que codifica una cadena pesada de anticuerpo;
secuencia que codifica una región variable, por ejemplo, una región variable de la cadena pesada o cadena ligera;
secuencia que codifica una región constante, por ejemplo, una región constante de la cadena pesada o cadena ligera;
secuencia diagnóstica para el tipo de cadena ligera, por ejemplo, secuencia específica kappa o lambda;
secuencia de región constante, por ejemplo, una secuencia de región constante que puede distinguir una o más clases y subclases de Ig, por ejemplo, IgG (por ejemplo, IgG1, IgG2, IgG3 o IgG4), IgM, IgD, IgA (por ejemplo, IgA1 o IgA2) o IgE;
secuencia de una CDR, por ejemplo, una CDR1, CDR2 o CDR3 de la cadena ligera o una CDR1, CDR2 o CDR3 de la cadena pesada, por ejemplo, una CDR3 de la cadena pesada;
secuencia que codifica un receptor de linfocitos T, por ejemplo, secuencia que codifica una cadena alfa, beta, delta o gamma;
secuencia que codifica una región variable de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de diversidad de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de acoplamiento de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de conmutación de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región constante de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica un segmento V (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica un segmento D (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica un segmento J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica segmentos V, D y J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica segmentos V y J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión VD de un receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión DJ de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión VJ de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;

De acuerdo con un modo de realización preferente, la muestra se ha obtenido de una neoplasia maligna o premaligna hematológica.

5

De acuerdo con un modo de realización más preferente, la neoplasia maligna hematológica es leucemia, leucemia linfocítica aguda (LLA), leucemia mielógena aguda (LMA), leucemia linfocítica crónica (LLC), leucemia mielógena crónica (LMC), tricoleucemia, leucemia monocítica aguda (LMoA), leucemia mielomonocítica crónica (LMMC), leucemia mielomonocítica juvenil (LMMJ), leucemia linfocítica granular de células grandes, linfoma, linfoma relacionado con el SIDA, linfoma cutáneo de linfocitos T, linfoma hodgkiniano, linfoma hodgkiniano clásico o linfoma hodgkiniano con predominio linfocitario nodular, micosis fungoide, linfoma no hodgkiniano, linfoma no hodgkiniano

10

de linfocitos B, linfoma de Burkitt, linfoma linfocítico de células pequeñas (LLC/LLP), linfoma difuso de linfocitos B grandes, linfoma folicular, linfoma inmunoblástico de células grandes, linfoma linfoblástico de precursores de linfocitos B, linfoma de células del manto, linfoma no hodgkiniano de linfocitos T, micosis fungoide, linfoma anaplásico de células grandes o linfoma linfoblástico de precursores de linfocitos B, linfoma primario del sistema nervioso central, síndrome de Sézary, macroglobulinemia de Waldenström, neoplasia mieloproliferativa crónica, histiocitosis de células de Langerhans, mieloma múltiple/neoplasia de células plasmáticas, síndrome mielodisplásico o neoplasia mielodisplásica/mieloproliferativa.

De acuerdo con un modo de realización preferente, la muestra se ha obtenido de una célula no cancerosa o célula no maligna.

De acuerdo con otro modo de realización preferente, la muestra se ha obtenido de un linfocito infiltrante del tumor.

De acuerdo con otro modo de realización preferente, el intervalo subgenómico se selecciona de las tablas 1-4.

De acuerdo con un modo de realización preferente, los procedimientos comprenden proporcionar la secuencia de al menos X intervalos subgenómicos, en la que X es igual a 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 o mayor.

De acuerdo con un modo de realización preferente, el intervalo subgenómico y el intervalo subgenómico expresado corresponden a la misma secuencia del mismo gen.

De acuerdo con un modo de realización preferente, el procedimiento comprende proporcionar la abundancia relativa de un acontecimiento en el intervalo subgenómico.

De acuerdo con otro modo de realización preferente, el procedimiento comprende proporcionar la abundancia relativa de un segmento V, un punto de unión VD, un punto de unión VJ o un punto de unión DJ.

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado como se entiende comúnmente por un experto en la técnica a la que pertenece la presente invención. Además, los materiales, procedimientos y ejemplos solo son ilustrativos y no se pretende que sean limitantes.

Otros rasgos característicos y ventajas de la invención serán evidentes a partir de la descripción detallada, los dibujos, y a partir de las reivindicaciones.

Breve descripción de los dibujos

Las FIG. 1A-1F muestran una representación de diagrama de flujo de una implementación de un procedimiento para el análisis multigénico de una muestra tumoral.

La FIG. 2 representa el impacto de la expectativa previa y la profundidad de lectura en la detección de mutaciones.

Descripción detallada

Se divulgan procedimientos y ensayos optimizados para secuenciar grandes cantidades de genes y productos génicos, por ejemplo, anticuerpos o receptores de la superfamilia de las inmunoglobulinas (por ejemplo, receptores de linfocitos T o receptores de linfocitos B), de muestras, por ejemplo, tejidos en estado de enfermedad o tejidos en estado de no enfermedad, de uno o más sujetos mediante la evaluación de un grupo seleccionado de genes y productos génicos. En una implementación, los procedimientos y ensayos presentados en el presente documento se usan en un formato de ensayo múltiple, por ejemplo, los ensayos incorporaron múltiples señales de una gran cantidad de acontecimientos genéticos diversos en una gran cantidad de genes. En el presente documento se divulgan procedimientos y ensayos que se basan, al menos en parte, en un grupo seleccionado de genes o productos génicos que se asocian (por ejemplo, positiva o negativamente) con un fenotipo canceroso (por ejemplo, uno o más de riesgo de cáncer, progresión del cáncer, tratamiento del cáncer o resistencia al tratamiento). Dichos genes o productos génicos preseleccionados permiten la aplicación de procedimientos de secuenciación, en particular procedimientos que se basan en la secuenciación masiva en paralelo de una gran cantidad de genes diversos, por ejemplo, de muestras tumorales o de control.

Los procedimientos divulgados en el presente documento proporcionan un análisis clonal de intervalos de sujeto, por ejemplo, en uno o ambos de los niveles de ADN y ARN. Por ejemplo, los procedimientos divulgados en el presente documento permiten proporcionar la abundancia relativa de un acontecimiento en un intervalo de sujeto, por ejemplo, la abundancia relativa de un segmento V o punto de unión VD, VJ o DJ particular, por ejemplo, en una inmunoglobulina, receptor de linfocitos B o receptor de linfocitos T. Otros procedimientos permiten la evaluación de los niveles de diversidad, por ejemplo, de la hipermutación somática, en un intervalo de sujeto. Otros procedimientos permiten el análisis de reordenamientos cromosómicos grandes, por ejemplo, de todo el brazo.

Otros procedimientos permiten la evaluación, por ejemplo, en uno o ambos de los niveles de ADN o ARN, de los intervalos de sujeto implicados en neoplasias malignas (o premalignas) hematológicas.

En primer lugar, se definen determinados términos. Se definen términos adicionales a lo largo de la memoria descriptiva.

Como se usa en el presente documento, los artículos "un" y "una" se refieren a uno o a más de uno (por ejemplo, a al menos uno) del objeto gramatical del artículo.

"Aproximadamente" significará, en general, un grado de error aceptable para la cantidad medida dada la naturaleza o precisión de las mediciones. Los grados ejemplares de error están dentro de un 20 por ciento (%), típicamente dentro de un 10 %, y más típicamente dentro de un 5 % de un valor dado o intervalo de valores.

"Adquirir" o "adquisición", como se usan los términos en el presente documento, se refiere a obtener la posesión de una entidad física, o de un valor, por ejemplo, un valor numérico, al "adquirir directamente" o "adquirir indirectamente" la entidad física o valor. "Adquirir directamente" significa realizar un procedimiento (por ejemplo, realizar un procedimiento sintético o analítico) para obtener la entidad física o valor. "Adquirir indirectamente" se refiere a recibir la entidad física o el valor de otra parte o fuente (por ejemplo, un laboratorio externo que adquirió directamente la entidad física o valor). Adquirir directamente una entidad física incluye realizar un procedimiento que incluya un cambio físico en una sustancia física, por ejemplo, un material de partida. Los cambios ejemplares incluyen preparar una entidad física a partir de dos o más materiales de partida, cortar o fragmentar una sustancia, separar o purificar una sustancia, combinar dos o más entidades separadas en una mezcla, realizar una reacción química que incluya romper o formar un enlace covalente o no covalente. Adquirir directamente un valor incluye realizar un procedimiento que incluya un cambio físico en una muestra u otra sustancia, por ejemplo, realizar un procedimiento analítico que incluya un cambio físico en una sustancia, por ejemplo, una muestra, analito o reactivo (a veces denominado en el presente documento "análisis físico"), realizar un procedimiento analítico, por ejemplo, un procedimiento que incluya uno o más de los siguientes: separar o purificar una sustancia, por ejemplo, un analito, o un fragmento u otro derivado del mismo, de otra sustancia; combinar un analito, o fragmento u otro derivado del mismo, con otra sustancia, por ejemplo, un tampón, disolvente o reactivo; o cambiar la estructura de un analito, o un fragmento u otro derivado del mismo, por ejemplo, rompiendo o formando un enlace covalente o no covalente entre un primer y un segundo átomo del analito; o cambiando la estructura de un reactivo, o un fragmento u otro derivado del mismo, por ejemplo, rompiendo o formando un enlace covalente o no covalente entre un primer y un segundo átomo del reactivo.

"Adquirir una secuencia" o "adquirir una lectura", como se usa el término en el presente documento, se refiere a obtener la posesión de una secuencia de nucleótidos o una secuencia de aminoácidos al "adquirir directamente" o "adquirir indirectamente" la secuencia o lectura. "Adquirir directamente" una secuencia o lectura significa realizar un procedimiento (por ejemplo, realizar un procedimiento sintético o analítico) para obtener la secuencia, tal como realizar un procedimiento de secuenciación (por ejemplo, un procedimiento de secuenciación de nueva generación (NGS)). "Adquirir indirectamente" una secuencia o lectura se refiere a recibir información o conocimiento de, o recibir, la secuencia de otra parte o fuente (por ejemplo, un laboratorio externo que adquirió directamente la secuencia). La secuencia o lectura adquirida no necesita ser una secuencia completa, por ejemplo, la secuenciación de al menos un nucleótido o la obtención de información o conocimiento que identifique una o más de las alteraciones divulgadas en el presente documento como que están presente en un sujeto constituye adquirir una secuencia.

Adquirir directamente una secuencia o lectura incluye realizar un procedimiento que incluya un cambio físico en una sustancia física, por ejemplo, un material de partida, tal como una muestra tisular o muestra celular, por ejemplo, una biopsia o una muestra de ácido nucleico aislado (por ejemplo, ADN o ARN). Los cambios ejemplares incluyen preparar una entidad física a partir de dos o más materiales de partida, cortar o fragmentar una sustancia, tal como un fragmento de ADN genómico; separar o purificar una sustancia (por ejemplo, aislar una muestra de ácido nucleico de un tejido); combinar dos o más entidades separadas en una mezcla, realizar una reacción química que incluya romper o formar un enlace covalente o no covalente. Adquirir directamente un valor incluye realizar un procedimiento que incluya un cambio físico en una muestra u otra sustancia como se describe anteriormente.

"Adquirir una muestra", como se usa el término en el presente documento, se refiere a obtener la posesión de una muestra, por ejemplo, una muestra tisular o muestra de ácido nucleico, al "adquirir directamente" o "adquirir indirectamente" la muestra. "Adquirir directamente una muestra" significa realizar un procedimiento (por ejemplo, realizar un procedimiento físico, tal como una cirugía o extracción) para obtener la muestra. "Adquirir indirectamente una muestra" se refiere a recibir la muestra de otra parte o fuente (por ejemplo, un laboratorio externo que adquirió directamente la muestra). Adquirir directamente una muestra incluye realizar un procedimiento que incluya un cambio físico en una sustancia física, por ejemplo, un material de partida, tal como un tejido, por ejemplo, un tejido en un paciente humano o un tejido que se aisló previamente de un paciente. Los cambios ejemplares incluyen preparar una entidad física a partir de un material de partida, disecar o raspar un tejido; separar o purificar una sustancia (por ejemplo, una muestra tisular o una muestra de ácido nucleico); combinar dos o más

entidades separadas en una mezcla; realizar una reacción química que incluya romper o formar un enlace covalente o no covalente. Adquirir directamente una muestra incluye realizar un procedimiento que incluya un cambio físico en una muestra u otra sustancia, por ejemplo, como se describe anteriormente.

"Selector de alineación", como se usa en el presente documento, se refiere a un parámetro que permite o dirige la selección de un procedimiento de alineación, por ejemplo, un algoritmo o parámetro de alineación, que puede optimizar la secuenciación de un intervalo subgenómico preseleccionado. Un selector de alineación puede ser específico para, o seleccionarse como, una función, por ejemplo, de uno o más de los siguientes:

1. El contexto de secuencia, por ejemplo, contexto de secuencia de un intervalo subgenómico (por ejemplo, la posición nucleotídica preseleccionada que se va a evaluar) que se asocia con una propensión a la alineación incorrecta de las lecturas para dicho intervalo subgenómico. Por ejemplo, la aparición de un elemento de secuencia en o cerca del intervalo subgenómico que se va a evaluar que se repite en otra parte del genoma puede causar alineación incorrecta y, de este modo, reducir el rendimiento. El rendimiento se puede potenciar seleccionando un algoritmo o un parámetro de algoritmo que minimice la alineación incorrecta. En este caso, el valor para el selector de alineación puede ser una función del contexto de secuencia, por ejemplo, la presencia o ausencia de una secuencia de longitud preseleccionada que se repite al menos un número preseleccionado de veces en el genoma (o en la porción del genoma que se está analizando).

2. El tipo de tumor que se está analizando. Por ejemplo, un tipo de tumor específico se puede caracterizar por una tasa incrementada de deleciones. Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo que sea más sensible a los indels. En este caso, el valor para el selector de alineación puede ser una función del tipo de tumor, por ejemplo, un identificador para el tipo de tumor. En una implementación, el valor es la identidad del tipo de tumor, por ejemplo, una neoplasia maligna (o premaligna) hematológica.

3. El gen, o tipo de gen, que se está analizando, por ejemplo, un gen, o tipo de gen, se puede analizar. Los oncogenes, a modo de ejemplo, se caracterizan a menudo por sustituciones o indels sin cambio de pauta de lectura. Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo que sea en particular sensible a estas variantes y específico frente a otras. Los supresores tumorales se caracterizan a menudo por indels con desplazamiento del marco de lectura. Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo que sea en particular sensible a estas variantes. Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo emparejado con el intervalo subgenómico. En este caso, el valor para el selector de alineación puede ser una función del gen o tipo de gen, por ejemplo, un identificador para el gen o tipo de gen. En una implementación, el valor es la identidad del gen.

4. El sitio (por ejemplo, la posición del nucleótido) que se está analizando. En este caso, el valor para el selector de alineación puede ser una función del sitio o el tipo de sitio, por ejemplo, un identificador para el sitio o tipo de sitio. En una implementación, el valor es la identidad del sitio. Por ejemplo, si el gen que contiene el sitio es altamente homólogo con otro gen, los algoritmos de alineación de lectura corta normal/rápida (por ejemplo, BWA) pueden tener dificultades para distinguir entre los dos genes, requiriendo potencialmente procedimientos de alineación más intensivos (Smith-Waterman) o incluso ensamblaje (ARACHNE). De forma similar, si la secuencia génica contiene regiones de baja complejidad (por ejemplo, AAAAAA), pueden ser necesarios procedimientos de alineación más intensivos.

5. La variante, o tipo de variante, asociada con el intervalo subgenómico que se está evaluando. Por ejemplo, una sustitución, inserción, delección, translocación u otro reordenamiento. Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo que sea más sensible al tipo de variante específico. En este caso, el valor para el selector de alineación puede ser una función del tipo de variante, por ejemplo, un identificador para el tipo de variante. En una implementación, el valor es la identidad del tipo de variante, por ejemplo, una sustitución.

6. El tipo de muestra, una muestra FFIP u otra muestra fijada. El tipo/calidad de la muestra puede afectar a la tasa de error (observación espuria de la secuencia sin referencia). Por tanto, se puede potenciar el rendimiento seleccionando un algoritmo o parámetro de algoritmo que modelice con exactitud la tasa de error real en la muestra. En este caso, el valor para el selector de alineación puede ser una función del tipo de muestra, por ejemplo, un identificador para el tipo de muestra. En una implementación, el valor es la identidad del tipo de muestra, por ejemplo, una muestra fijada.

"Alteración" o "estructura alterada", como se usa en el presente documento, de un gen o producto génico (por ejemplo, un gen o producto génico marcador) se refiere a la presencia de una mutación o mutaciones dentro del gen o producto génico, por ejemplo, una mutación que afecta a la cantidad o actividad del gen o producto génico en comparación con el gen normal o natural. La alteración puede ser en cantidad, estructura y/o actividad en un tejido canceroso o célula cancerosa, en comparación con su cantidad, estructura y/o actividad en un tejido o célula normal o sano (por ejemplo, un control), y se asocia con un estado de enfermedad, tal como cáncer. Por ejemplo, una alteración que se asocia con el cáncer, o predictiva de la reactividad a tratamientos antineoplásico, puede

tener una secuencia de nucleótidos (por ejemplo, una mutación), secuencia de aminoácidos, translocación cromosómica, inversión intracromosómica, número de copias, nivel de expresión, nivel de proteína, actividad de proteína o estado de metilación alterados en un tejido canceroso o célula cancerosa, en comparación con un tejido o célula normal y sana. Las mutaciones ejemplares incluyen, pero no se limitan a, mutaciones puntuales (por ejemplo, silenciosas, de sentido alterado (missense) o finalizadoras (nonsense)), deleciones, inserciones, inversiones, mutaciones de enlace, duplicaciones, translocaciones, reordenamientos intercromosómicos e intracromosómicos. Las mutaciones pueden estar presentes en la región codificante o no codificante del gen. En determinadas implementaciones, la alteración o alteraciones se detectan como un reordenamiento, por ejemplo, un reordenamiento genómico que comprende uno o más intrones o fragmentos de los mismos los mismos (por ejemplo, uno o más reordenamientos en la 5'-UTR y/o 3'-UTR). En determinadas implementaciones, las alteraciones se asocian (o no asocian) con un fenotipo, por ejemplo, un fenotipo canceroso (por ejemplo, uno o más de riesgo de cáncer, progresión del cáncer, tratamiento del cáncer o resistencia al tratamiento del cáncer). En una implementación, la alteración se asocia con uno o más de: un factor de riesgo genético de cáncer, un predictor de respuesta positiva al tratamiento, un predictor de respuesta negativa al tratamiento, un factor pronóstico positivo, un factor pronóstico negativo o un factor diagnóstico.

"Perfil clonal", como se usa ese término en el presente documento, se refiere a la aparición, identidad, variabilidad, distribución, expresión (la aparición o el nivel de copias transcritas de una firma subgenómica) o abundancia, por ejemplo, la abundancia relativa, de una o más secuencias, por ejemplo, un alelo o firma, de un intervalo de sujeto (o de una célula que comprende el mismo). En una implementación, el perfil clonal es un valor para la abundancia relativa para una secuencia, alelo o firma, para un intervalo de sujeto (o de una célula que comprende el mismo) cuando una pluralidad de secuencias, alelos o firmas para ese intervalo de sujeto están presentes en una muestra. Por ejemplo, en una implementación, un perfil clonal comprende un valor para la abundancia relativa de una o más de una pluralidad de combinaciones VDJ o VJ para un intervalo de sujeto. En una implementación, un perfil clonal comprende un valor para la abundancia relativa de un segmento V seleccionado para un intervalo de sujeto. En una implementación, un perfil clonal comprende un valor para la diversidad, por ejemplo, como surge de la hipermutación somática, dentro de las secuencias de un intervalo de sujeto. En una implementación, un perfil clonal comprende un valor para la aparición o el nivel de expresión de una secuencia, alelo o firma, por ejemplo, como se evidencia por la aparición o nivel de un intervalo subgenómico expresado que comprende la secuencia, alelo o firma.

"Intervalo subgenómico expresado", como se usa ese término en el presente documento, se refiere a la secuencia transcrita de un intervalo subgenómico. En una implementación, la secuencia del intervalo subgenómico expresado diferirá del intervalo subgenómico a partir del cual se transcribe, por ejemplo, ya que es posible que alguna secuencia no se transcriba.

"Firma", como se usa ese término en el presente documento, se refiere a una secuencia de un intervalo de sujeto. Una firma puede ser diagnóstica de la aparición de una de una pluralidad de posibilidades en un intervalo de sujeto, por ejemplo, una firma puede ser diagnóstica de: la aparición de un segmento V seleccionado en un gen de la región variable de la cadena pesada o ligera reordenado; la aparición de un punto de unión VJ seleccionado, por ejemplo, la aparición de un segmento V seleccionado y un segmento J seleccionado en un gen de la región variable de la cadena pesada reordenado. En una implementación, una firma comprende una pluralidad de secuencias de ácido nucleico específicas. Por tanto, una firma no se limita a una secuencia de ácido nucleico específica, sino que es lo suficientemente única como para poder distinguir entre un primer grupo de secuencias o posibilidades en un intervalo de sujeto y un segundo grupo de posibilidades en un intervalo de sujeto, por ejemplo, puede distinguir entre un primer segmento V y un segundo segmento V, permitiendo, por ejemplo, la evaluación del uso de diversos segmentos V. El término firma comprende el término firma específica, que es una secuencia de ácido nucleico específica. En una implementación, la firma es indicativa de, o es el producto de, un acontecimiento específico, por ejemplo, un acontecimiento de reordenamiento.

"Intervalo subgenómico", como se usa ese término en el presente documento, se refiere a una porción de la secuencia genómica. En una implementación, un intervalo subgenómico puede ser una única posición nucleotídica, por ejemplo, variantes de una posición nucleotídica que se asocian (positiva o negativamente) con un fenotipo tumoral. En una implementación, un intervalo subgenómico comprende más de una posición nucleotídica. Dichas implementaciones incluyen secuencias de al menos 2, 5, 10, 50, 100, 150 o 250 posiciones nucleotídicas de longitud. Los intervalos subgenómicos pueden comprender un gen completo, o una porción preseleccionada del mismo, por ejemplo, la región codificante (o porciones de la misma), un intrón (o porción del mismo) o exón (o porción del mismo) preseleccionado. Un intervalo subgenómico puede comprender la totalidad o una parte de un fragmento de un ácido nucleico natural, por ejemplo, genómico. Por ejemplo, un intervalo subgenómico puede corresponder a un fragmento de ADN genómico que se somete a una reacción de secuenciación. En implementaciones, un intervalo subgenómico es una secuencia continua de una fuente genómica. En implementaciones, un intervalo subgenómico incluye secuencias que no son contiguas en el genoma, por ejemplo, puede incluir puntos de unión formados que se encuentran en los puntos de unión exón-exón en el ADNc.

En una implementación, un intervalo subgenómico corresponde a una secuencia reordenada, por ejemplo, una secuencia en un linfocito B o T que surge como resultado del acoplamiento de un segmento V a un segmento D,

un segmento D a un segmento J, un segmento V a un segmento J o un segmento J a un segmento J a un segmento de clase.

En una implementación, no hay diversidad en un intervalo subgenómico.

- 5 En una implementación, hay diversidad en un intervalo subgenómico, por ejemplo, el intervalo subgenómico está representado por más de una secuencia, por ejemplo, el intervalo subgenómico que cubre una secuencia VD puede estar representado por más de una firma.

10 En una implementación, un intervalo subgenómico comprende o consiste en: una única posición nucleotídica; una región intragénica o una región intergénica; un exón o un intrón, o un fragmento de los mismos, típicamente una secuencia de exón o un fragmento de la misma; una región codificante o una región no codificante, por ejemplo, un promotor, un potenciador, una región 5' no traducida (5'-UTR) o una región 3' no traducida (3'-UTR) o un fragmento de la misma; un ADNc o un fragmento del mismo; un polimorfismo mononucleotídico (SNP); una mutación somática, una mutación de estirpe germinal o ambas; una alteración, por ejemplo, una mutación puntual o una única; una mutación de delección (por ejemplo, una delección sin cambio de pauta de lectura, una delección intragénica, una delección génica completa); una mutación de inserción (por ejemplo, una inserción intragénica); una mutación de inversión (por ejemplo, una inversión intracromosómica); una mutación de enlace; una mutación de inserción enlazada; una mutación de duplicación invertida; una duplicación en tándem (por ejemplo, una duplicación en tándem intracromosómica); una translocación (por ejemplo, una translocación cromosómica, una translocación no recíproca); un reordenamiento (por ejemplo, un reordenamiento genómico (por ejemplo, un reordenamiento de uno o más intrones, o un fragmento de los mismos; un intrón reordenado puede incluir una 5'-UTR y/o 3'-UTR); un cambio en el número de copias del gen; un cambio en la expresión génica; un cambio en los niveles de ARN o una combinación de los mismos. El "número de copias de un gen" se refiere al número de secuencias de ADN en una célula que codifica un producto génico particular. En general, para un gen dado, un mamífero tiene dos copias de cada gen. El número de copias se puede incrementar, por ejemplo, mediante amplificación o duplicación génica, o reducir mediante delección.

25

"Intervalo de sujeto", como se usa ese término en el presente documento, se refiere a un intervalo subgenómico o un intervalo subgenómico expresado. En una implementación, un intervalo subgenómico y un intervalo subgenómico expresado se corresponden, lo que significa que el intervalo subgenómico expresado comprende una secuencia expresada del intervalo subgenómico correspondiente. En una implementación, un intervalo subgenómico y un intervalo subgenómico expresado no se corresponden, lo que significa que el intervalo subgenómico expresado no comprende la secuencia expresada del intervalo subgenómico no correspondiente, sino que se corresponde con un intervalo subgenómico diferente. En una implementación, un intervalo subgenómico y un intervalo subgenómico expresado se corresponden parcialmente, lo que significa que el intervalo subgenómico expresado comprende la secuencia expresada del intervalo subgenómico correspondiente y la secuencia expresada de un intervalo subgenómico correspondiente diferente.

30

35

Como se usa en el presente documento, el término "colección" se refiere a un conjunto de miembros. En una implementación, la colección incluye un conjunto de miembros de ácido nucleico, por ejemplo, un conjunto de fragmentos genómicos, subgenómicos completos, ADNc, fragmentos de ADNc, ARN, por ejemplo, ARNm, fragmentos de ARN o una combinación de los mismos. En una implementación, una porción o todos los miembros de la colección comprenden una secuencia de adaptador. La secuencia del adaptador se puede localizar en uno o ambos extremos. La secuencia de adaptador puede ser útil, por ejemplo, para un procedimiento de secuenciación (por ejemplo, un procedimiento NGS), para amplificación, para retrotranscripción o para clonación en un vector.

40

45

La colección puede comprender un conjunto de miembros, por ejemplo, un miembro diana (por ejemplo, un miembro tumoral, un miembro de referencia, un miembro PGx o una combinación de los mismos). Los miembros de la colección pueden proceder de un único individuo. En implementaciones, una colección puede comprender miembros de más de un sujeto (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 o más sujetos), por ejemplo, se pueden combinar dos o más colecciones de diferentes sujetos a partir de una colección que comprende miembros de más de un sujeto. En una implementación, el sujeto es un ser humano que tiene, o está en riesgo de tener, un cáncer o tumor.

50

"Captura de colección" se refiere a un subconjunto de una colección, por ejemplo, un subconjunto enriquecido para intervalos subgenómicos preseleccionados, por ejemplo, producto capturado por hibridación con cebos preseleccionados.

55

"Miembro" o "miembro de la colección" u otro término similar, como se usa en el presente documento, se refiere a una molécula de ácido nucleico, por ejemplo, un ADN, ARN o una combinación de los mismos, que es el miembro de una colección. Típicamente, un miembro es una molécula de ADN, por ejemplo, ADN genómico o ADNc. Un miembro puede estar fragmentado, por ejemplo, ADN genómico cortado o preparado enzimáticamente. Los miembros comprenden una secuencia de un sujeto y también pueden comprender una secuencia no derivada del sujeto, por ejemplo, una secuencia de adaptadores, una secuencia de cebador u otras secuencias que permitan la identificación, por ejemplo, secuencias de "código de barras".

60

65

"Cebo", como se usa en el presente documento, es un tipo de reactivo de captura de híbridos. Un cebo puede ser una molécula de ácido nucleico, por ejemplo, una molécula de ADN o ARN, que se puede hibridar (por ejemplo, ser complementaria a) y, de este modo, permitir la captura de un ácido nucleico diana. En una implementación, un cebo es una molécula de ARN (por ejemplo, una molécula de ARN natural o modificada); una molécula de ADN (por ejemplo, una molécula de ADN natural o modificada) o una combinación de las mismas. En otras implementaciones, un cebo incluye una entidad de unión, por ejemplo, una marca de afinidad, que permite la captura y separación, por ejemplo, uniéndose a una entidad de unión, de un híbrido formado por un cebo y un ácido nucleico hibridado al cebo. En una implementación, un cebo es adecuado para la hibridación en fase de solución. En una implementación, un cebo es una molécula de ácido nucleico bicíclico (BNA).

"Conjunto de cebos", como se usa en el presente documento, se refiere a una o una pluralidad de moléculas de cebo.

"Entidad de unión" significa cualquier molécula a la que se pueden fijar directa o indirectamente marcas moleculares que se pueden unir específicamente a un analito. La entidad de unión puede ser una marca de afinidad en cada secuencia de cebo. En determinadas implementaciones, la entidad de unión permite la separación de los híbridos cebo/miembro de una mezcla de hibridación uniéndose a un compañero de unión, tal como una molécula de avidina, o un anticuerpo que se une al hapteno o a un fragmento de unión a antígeno del mismo. Las entidades de unión ejemplares incluyen, pero no se limitan a, una molécula de biotina, un hapteno, un anticuerpo, un fragmento de unión a anticuerpo, un péptido y una proteína.

"Complementario" se refiere a la complementariedad de secuencias entre regiones de dos hebras de ácido nucleico o entre dos regiones de la misma hebra de ácido nucleico. Es conocido que un residuo de adenina de una primera región de ácido nucleico puede formar enlaces de hidrógeno específicos ("emparejamiento de bases") con un residuo de una segunda región de ácido nucleico que es antiparalela a la primera región si el residuo es timina o uracilo. De forma similar, es conocido que un residuo de citosina de una primera hebra de ácido nucleico se puede someter a emparejamiento de bases con un residuo de una segunda hebra de ácido nucleico que es antiparalela a la primera hebra si el residuo es guanina. Una primera región de un ácido nucleico es complementaria a una segunda región del mismo ácido nucleico o uno diferente si, cuando las dos regiones están dispuestas en una forma antiparalela, al menos un residuo nucleotídico de la primera región se puede someter a emparejamiento de bases con un residuo de la segunda región. En determinadas implementaciones, la primera región comprende una primera porción y la segunda región comprende una segunda porción, con lo que, cuando las primera y segunda porciones están dispuestas en una forma antiparalela, al menos aproximadamente un 50 %, al menos aproximadamente un 75 %, al menos aproximadamente un 90 % o al menos aproximadamente un 95 % de los residuos nucleotídicos de la primera porción se pueden someter a emparejamiento de bases con residuos nucleotídicos en la segunda porción. En otras implementaciones, todos los residuos nucleotídicos de la primera porción se pueden someter a emparejamiento de bases con residuos nucleotídicos en la segunda porción.

El término "cáncer" o "tumor" se usa de manera intercambiable en el presente documento. Estos términos se refieren a la presencia de células que poseen características típicas de las células que provocan cáncer, tales como proliferación no controlada, inmortalidad, potencial metastásico, rápida tasa de crecimiento y proliferación, y determinados rasgos característicos morfológicos característicos. Las células cancerosas a menudo están en la forma de un tumor, pero dichas células pueden existir solas dentro de un animal, o pueden ser una célula cancerosa no oncogénica, tal como una célula de leucemia. Estos términos incluyen un tumor sólido, un tumor de tejidos blandos o una lesión metastásica. Como se usa en el presente documento, el término "cáncer" incluye cánceres premalignos, así como malignos.

"Probable que" o "probabilidad incrementada", como se usa en el presente documento, se refiere a una probabilidad incrementada de que se produzca un asunto, objeto, cosa o persona. Por tanto, en un ejemplo, un sujeto que es probable que responda al tratamiento tiene una probabilidad incrementada de responder al tratamiento en relación con un sujeto o grupo de sujetos de referencia.

"Improbable que" se refiere a una probabilidad disminuida de que se produzca un acontecimiento, asunto, objeto, cosa o persona con respecto a una referencia. Por tanto, un sujeto que es improbable que responda al tratamiento tiene una probabilidad disminuida de responder al tratamiento en relación con un sujeto o grupo de sujetos de referencia.

"Miembro de control" se refiere a un miembro que tiene una secuencia de una célula no tumoral.

"Selector de secuencia de alineación de indel", como se usa en el presente documento, se refiere a un parámetro que permite o dirige la selección de una secuencia con la que se va a alinear una lectura en el caso de un indel preseleccionado. El uso de dicha secuencia puede optimizar la secuenciación de un intervalo subgenómico preseleccionado que comprende un indel. El valor para un selector de secuencia de alineación de indel es una función de un indel preseleccionado, por ejemplo, un identificador para el indel. En una implementación, el valor es la identidad del indel.

"Secuenciación de nueva generación o NGS o secuenciación NG", como se usa en el presente documento, se refiere a cualquier procedimiento de secuenciación que determine la secuencia de nucleótidos de moléculas de ácido nucleico individuales (por ejemplo, en la secuenciación de molécula única) o bien sustitutos expandidos clonalmente para moléculas de ácido nucleico individuales en un modo de alto rendimiento (por ejemplo, se secuencian simultáneamente más de 10^3 , 10^4 , 10^5 moléculas). En una implementación, la abundancia relativa de las especies de ácido nucleico en la colección se puede estimar contando el número relativo de apariciones de sus secuencias afines en los datos generados por el experimento de secuenciación. Son conocidos procedimientos de secuenciación de nueva generación en la técnica y se describen, por ejemplo, en Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46. La secuenciación de nueva generación puede detectar una variante presente en menos de un 5 % de los ácidos nucleicos en una muestra.

"Valor de nucleótido", como se hace referencia en el presente documento, representa la identidad del nucleótido o nucleótidos que ocupan o están asignados a una posición nucleotídica preseleccionada. Los valores de nucleótido típicos incluyen: faltante (por ejemplo, deleccionado); adicional (por ejemplo, una inserción de uno o más nucleótidos, la identidad de los cuales puede o no estar incluida) o presente (ocupado); A; T; C; o G. Otros valores pueden ser, por ejemplo, no Y, en el que Y es A, T, G o C; A o X, en el que X es uno o dos de T, G o C; T o X, en el que X es uno o dos de A, G o C; G o X, en el que X es uno o dos de T, A o C; C o X, en el que X es uno o dos de T, G o A; un nucleótido pirimidina; o un nucleótido purina. Un valor de nucleótido puede ser una frecuencia para 1 o más, por ejemplo, 2, 3 o 4 bases (u otro valor descrito en el presente documento, por ejemplo, faltante o adicional) en una posición nucleotídica. Por ejemplo, un valor de nucleótido puede comprender una frecuencia para A, y una frecuencia para G, en una posición nucleotídica.

El término "o" se usa en el presente documento para significar, y se usa de manera intercambiable con, el término "y/o", a menos que el contexto lo indique claramente de otro modo. El uso del término "y/o" en algunos lugares del presente documento no significa que los usos del término "o" no sean intercambiables con el término "y/o", a menos que el contexto lo indique claramente de otro modo.

"Control primario" se refiere a un tejido no tumoral distinto del tejido contiguo normal (TCN) en una muestra tumoral. La sangre es un control primario típico.

"Selector de secuencia de alineación de reordenamiento", como se usa en el presente documento, se refiere a un parámetro que permite o dirige la selección de una secuencia con la que se va a alinear una lectura en el caso de un reordenamiento preseleccionado. El uso de dicha secuencia puede optimizar la secuenciación de un intervalo subgenómico preseleccionado que comprende un reordenamiento. El valor para un selector de secuencia de alineación de reordenamiento es una función de un indel preseleccionado, por ejemplo, un identificador para el reordenamiento. En una implementación, el valor es la identidad del reordenamiento. Un "selector de secuencia de alineación de indel" (también definido en otra parte del presente documento) es un ejemplo de un selector de secuencia de alineación de reordenamiento.

"Muestra", "muestra tisular", "muestra del paciente" o "muestra celular o tisular del paciente" se refieren cada una a un conjunto de células similares obtenidas de un tejido, o de células circundantes, de un sujeto o paciente. La fuente de la muestra tisular puede ser tejido sólido como de un órgano recién obtenido, congelado y/o conservado, muestra tisular, biopsia o aspirado; sangre o cualquier componente de la sangre; líquidos corporales, tales como líquido cefalorraquídeo, líquido amniótico, líquido peritoneal o líquido intersticial; o células de cualquier momento de la gestación o desarrollo del sujeto. La muestra tisular puede contener compuestos que no estén naturalmente entremezclados con el tejido en la naturaleza, tales como conservantes, anticoagulantes, tampones, fijadores, nutrientes, antibióticos o similares. En una implementación, la muestra se conserva como una muestra congelada o como una preparación tisular fijada con formaldehído o paraformaldehído e incluida en parafina (FFIP). Por ejemplo, la muestra se puede incluir en una matriz, por ejemplo, un bloque FFIP o una muestra congelada. En otra implementación, la muestra es una muestra de sangre. Aún en otra implementación, la muestra es una muestra de aspirado de médula ósea.

En una implementación, la muestra es una célula asociada con un tumor, por ejemplo, una célula tumoral o un linfocito infiltrante del tumor (TIL). En una implementación, la muestra es una muestra tumoral, por ejemplo, incluye una o más células premalignas o malignas. En una implementación, la muestra se adquiere de una neoplasia maligna (o premaligna) hematológica, por ejemplo, una neoplasia maligna (o premaligna) hematológica descrita en el presente documento. En determinadas implementaciones, la muestra, por ejemplo, la muestra tumoral, se adquiere de un tumor sólido, un tumor de tejidos blandos o una lesión metastásica. En otras implementaciones, la muestra, por ejemplo, la muestra tumoral, incluye tejido o células de un margen quirúrgico. En otra implementación, la muestra, por ejemplo, muestra tumoral, incluye una o más células tumorales circulantes (CTC) (por ejemplo, una CTC adquirida de una muestra de sangre). En una implementación, la muestra es una célula no asociada con un tumor, por ejemplo, una célula no tumoral o un linfocito de sangre periférica.

"Sensibilidad", como se usa en el presente documento, es una medida de la capacidad de un procedimiento para detectar una variante de secuencia preseleccionada en una población heterogénea de secuencias. Un procedimiento tiene una sensibilidad de S % para variantes de F % si, dada una muestra en la que la variante de

secuencia preseleccionada está presente como al menos F % de las secuencias en la muestra, el procedimiento puede detectar la secuencia preseleccionada con una confianza preseleccionada de C %, un S % de las veces. A modo de ejemplo, un procedimiento tiene una sensibilidad de un 90 % para variantes del 5 % si, dada una muestra en la que la secuencia variante preseleccionada está presente como al menos un 5 % de las secuencias en la muestra, el procedimiento puede detectar la secuencia preseleccionada con una confianza preseleccionada de 99 %, 9 de 10 veces ($F = 5\%$; $C = 99\%$; $S = 90\%$). Las sensibilidades ejemplares incluyen las de $S = 90\%$, 95% , 99% para variantes de secuencia con $F = 1\%$, 5% , 10% , 20% , 50% , 100% a niveles de confianza de $C = 90\%$, 95% , 99% y $99,9\%$.

"Especificidad", como se usa en el presente documento, es una medida de la capacidad de un procedimiento para distinguir una variante de secuencia preseleccionada que ocurre verdaderamente respecto de artefactos de secuenciación u otras secuencias estrechamente relacionadas. Es la capacidad de evitar detecciones de positivos falsos. Las detecciones de positivos falsos pueden surgir de errores introducidos en la secuencia de interés durante la preparación de la muestra, error de secuenciación o secuenciación inadvertida de secuencias estrechamente relacionadas como pseudogenes o miembros de una familia de genes. Un procedimiento tiene una especificidad de X % si, cuando se aplica a un conjunto de muestras de N_{Total} secuencias, en las que $X_{\text{Verdaderas}}$ secuencias son verdaderamente variantes y $X_{\text{No verdaderas}}$ no son verdaderamente variantes, el procedimiento selecciona al menos X % de las no verdaderamente variantes como no variantes. Por ejemplo, un procedimiento tiene una especificidad de un 90 % si, cuando se aplica a un conjunto de muestras de 1000 secuencias, en el que 500 secuencias son verdaderamente variantes y 500 no son verdaderamente variantes, el procedimiento selecciona un 90 % de las 500 secuencias no verdaderamente variantes como no variantes. Las especificidades ejemplares incluyen un 90, 95, 98 y 99 %.

Una "muestra de ácido nucleico tumoral", como se usa en el presente documento, se refiere a moléculas de ácido nucleico de una muestra tumoral o cancerosa. Típicamente, es ADN, por ejemplo, ADN genómico, o ADNc derivado de ARN, de una muestra tumoral o cancerosa. En determinadas implementaciones, la muestra de ácido nucleico tumoral se purifica o aísla (por ejemplo, se retira de su estado natural).

Una "muestra de ácido nucleico" de "control" o de "referencia", como se usa en el presente documento, se refiere a moléculas de ácido nucleico de una muestra de control o de referencia. Típicamente, es ADN, por ejemplo, ADN genómico, o ADNc derivado de ARN, que no contiene la alteración o variación en el gen o producto génico. En determinadas implementaciones, la muestra de ácido nucleico de referencia o de control es una secuencia natural o una no mutada. En determinadas implementaciones, la muestra de ácido nucleico de referencia se purifica o aísla (por ejemplo, se retira de su estado natural). En otras implementaciones, la muestra de ácido nucleico de referencia es de una muestra no tumoral, por ejemplo, un control de sangre, un tumor contiguo normal (TCN) o cualquier otra muestra no cancerosa del mismo sujeto o de uno diferente.

"Secuenciar" una molécula de ácido nucleico requiere determinar la identidad de al menos 1 nucleótido en la molécula (por ejemplo, una molécula de ADN, una molécula de ARN o una molécula de ADNc derivada de una molécula de ARN). En implementaciones, se determina la identidad de menos de la totalidad de los nucleótidos en una molécula. En otras implementaciones, se determina la identidad de la mayoría o la totalidad de los nucleótidos en la molécula.

"Valor umbral", como se usa en el presente documento, es un valor que es una función del número de lecturas que se requiere que estén presentes para asignar un valor de nucleótido a un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado). Por ejemplo, es una función del número de lecturas que tienen un valor de nucleótido específico, por ejemplo, A, en una posición nucleotídica, requerido para asignar ese valor de nucleótido a esa posición nucleotídica en el intervalo subgenómico. El valor umbral se puede expresar, por ejemplo, como (o como una función de) un número de lecturas, por ejemplo, un número entero, o como una proporción de lecturas que tienen el valor preseleccionado. A modo de ejemplo, si el valor umbral es X, y están presentes X+1 lecturas que tienen el valor de nucleótido de "A", entonces el valor de "A" se asigna a la posición preseleccionada en el intervalo de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado). El valor umbral también se puede expresar como una función de una expectativa de mutación o variante, frecuencia de mutación o de previo bayesiano. En una implementación, una frecuencia de mutación preseleccionada requeriría un número o proporción preseleccionado de lecturas que tengan un valor de nucleótido, por ejemplo, A o G, en una posición preseleccionada, para considerar a ese valor el valor de nucleótido. En implementaciones, el valor umbral puede ser una función de la expectativa de mutación, por ejemplo, la frecuencia de mutación y el tipo de tumor. Por ejemplo, una variante preseleccionada en una posición nucleotídica preseleccionada podría tener un primer valor umbral si el paciente tiene un primer tipo de tumor y un segundo valor umbral si el paciente tiene un segundo tipo de tumor.

Como se usa en el presente documento, "miembro diana" se refiere a una molécula de ácido nucleico que se desea aislar de la colección de ácidos nucleicos. En una implementación, los miembros diana pueden ser un miembro tumoral, un miembro de referencia, un miembro de control o un miembro PGx como se describe en el presente documento.

"Miembro tumoral" u otro término similar (por ejemplo, un "miembro asociado a tumor o cáncer"), como se usa en el presente documento, se refiere a un miembro que tiene una secuencia de una célula tumoral. En una implementación, el miembro tumoral incluye un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) que tiene una secuencia (por ejemplo, una secuencia de nucleótidos) que tiene una alteración (por ejemplo, una mutación) asociada con un fenotipo canceroso. En otras implementaciones, el miembro tumoral incluye un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) que tiene una secuencia natural (por ejemplo, una secuencia de nucleótidos natural). Por ejemplo, un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) de un alelo natural heterocigótico u homocigótico presente en una célula cancerosa. Un miembro tumoral puede incluir un miembro de referencia o un miembro PGx.

"Miembro de referencia" u otro término similar (por ejemplo, un "miembro de control"), como se usa en el presente documento, se refiere a un miembro que comprende un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) que tiene una secuencia (por ejemplo, una secuencia de nucleótidos) que no se asocia con el fenotipo canceroso. En una implementación, el miembro de referencia incluye una secuencia de nucleótidos natural o no mutada de un gen o producto génico que, cuando está mutada, se asocia con el fenotipo canceroso. El miembro de referencia puede estar presente en una célula cancerosa o no cancerosa.

"Miembro PGx" u otro término similar, como se usa en el presente documento, se refiere a un miembro que comprende un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) que se asocia con el perfil farmacogenético o farmacogenómico de un gen. En una implementación, el miembro PGx incluye un SNP (por ejemplo, un SNP como se describe en el presente documento). En otras implementaciones, el miembro PGx incluye un intervalo de sujeto (por ejemplo, un intervalo subgenómico o un intervalo subgenómico expresado) de acuerdo con las **tablas 1-4**.

"Variante", como se usa en el presente documento, se refiere a una estructura que puede estar presente en un intervalo subgenómico que puede tener más de una estructura, por ejemplo, un alelo en un locus polimórfico.

Los encabezados, por ejemplo, (a), (b), (i), etc., se presentan meramente para facilitar la lectura de la memoria descriptiva y las reivindicaciones. El uso de encabezados en la memoria descriptiva o reivindicaciones no requiere que las etapas o elementos se realicen en orden alfabético o numérico o el orden en que se presentan.

Selección de gen o productos génicos

Los genes o productos génicos seleccionados (también denominados en el presente documento "genes o productos génicos diana") pueden incluir intervalos de sujeto (por ejemplo, intervalos subgenómicos, intervalos subgenómicos expresados o ambos) que comprenden regiones intragénicas o regiones intergénicas. Por ejemplo, los intervalos de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado) pueden incluir un exón o un intrón, o un fragmento de los mismos, típicamente una secuencia de exón o un fragmento de la misma. El intervalo de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado) puede incluir una región codificante o una región no codificante, por ejemplo, un promotor, un potenciador, una región 5' no traducida (5'-UTR) o una región 3' no traducida (3'-UTR), o un fragmento de los mismos. En otras implementaciones, el intervalo de sujeto incluye un ADNc o un fragmento del mismo. En otras implementaciones, el intervalo de sujeto incluye un SNP, por ejemplo, como se describe en el presente documento.

En otras implementaciones, los intervalos de sujeto (por ejemplo, intervalos subgenómicos, intervalos subgenómicos expresados o ambos) incluyen sustancialmente todos los exones en un genoma, por ejemplo, uno o más de los intervalos de sujeto (por ejemplo, intervalos subgenómicos, intervalos subgenómicos expresados o ambos) como se describe en el presente documento (por ejemplo, exones de genes o productos génicos seleccionados de interés (por ejemplo, genes o productos génicos asociados con un fenotipo canceroso como se describe en el presente documento)). En una implementación, el intervalo de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado) incluye una mutación somática, una mutación de estirpe germinal o ambas. En una implementación, el intervalo de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado) incluye una alteración, por ejemplo, una mutación puntual o una única, una mutación de delección (por ejemplo, una delección sin cambio de pauta de lectura, una delección intragénica, una delección génica completa), una mutación de inserción (por ejemplo, inserción intragénica), una mutación de inversión (por ejemplo, una inversión intracromosómica), una mutación de enlace, una mutación de inserción enlazada, una mutación de duplicación invertida, una duplicación en tándem (por ejemplo, una duplicación en tándem intracromosómica), una translocación (por ejemplo, una translocación cromosómica, una translocación no recíproca), un reordenamiento, un cambio en el número de copias del gen o una combinación de los mismos. En determinadas implementaciones, el intervalo de sujeto (por ejemplo, intervalo subgenómico o intervalo subgenómico expresado) constituye menos de un 5, 1, 0,5, 0,1 %, 0,01 %, 0,001 % de la región codificante del genoma de las células tumorales en una muestra. En otras implementaciones, los intervalos de sujeto (por ejemplo, intervalos subgenómicos, intervalos subgenómicos expresados o ambos) no están involucrados en una enfermedad, por ejemplo, no se asocian con un fenotipo canceroso como se describe en el presente documento.

En una implementación, el gen o producto génico diana es un biomarcador. Como se usa en el presente documento, un "biomarcador" o "marcador" es un gen, ARNm o proteína que se puede alterar, en el que dicha alteración se asocia con el cáncer. La alteración puede ser en cantidad, estructura y/o actividad en un tejido canceroso o célula cancerosa, en comparación con su cantidad, estructura y/o actividad en un tejido o célula normal o sano (por ejemplo, un control), y se asocia con un estado de enfermedad, tal como cáncer. Por ejemplo, un marcador asociado con el cáncer, o predictivo de la reactividad a tratamientos antineoplásico, puede tener una secuencia de nucleótidos, secuencia de aminoácidos, translocación cromosómica, inversión intracromosómica, número de copias, nivel de expresión, nivel de proteína, actividad de proteína o estado de metilación alterados en un tejido canceroso o célula cancerosa, en comparación con un tejido o célula normal y sana. Además, un "marcador" incluye una molécula cuya estructura está alterada, por ejemplo, mutada (contiene una mutación), por ejemplo, difiere de la secuencia natural a nivel de nucleótidos o aminoácidos, por ejemplo, por sustitución, delección o inserción, cuando está presente en un tejido o célula asociado con un estado de enfermedad, tal como cáncer.

En una implementación, el gen o producto génico diana incluye un polimorfismo mononucleotídico (SNP). En otra implementación, el gen o producto génico tiene una delección pequeña, por ejemplo, una delección intragénica pequeña (por ejemplo, una delección sin cambio de pauta de lectura o con desplazamiento del marco de lectura). Aún en otra implementación, la secuencia diana resulta de la delección de un gen completo. Todavía en otra implementación, la secuencia diana tiene una inserción pequeña, por ejemplo, una inserción intragénica pequeña. En una implementación, la secuencia diana resulta de una inversión, por ejemplo, una inversión intracromosómica. En otra implementación, la secuencia diana resulta de una translocación intercromosómica. Aún en otra implementación, la secuencia diana tiene una duplicación en tándem. En una implementación, la secuencia diana tiene un rasgo característico indeseable (por ejemplo, alto contenido de GC o elemento de repetición). En otra implementación, la secuencia diana tiene una porción de la secuencia de nucleótidos a la que no es posible dirigirse con éxito, por ejemplo, debido a su naturaleza repetitiva. En una implementación, la secuencia diana resulta de un corte y empalme alternativo. En otra implementación, la secuencia diana se elige de un gen o producto génico, o un fragmento del mismo, de acuerdo con las **tablas 1-4**.

En una implementación, el gen o producto génico diana, o un fragmento del mismo, es un gen o producto génico de anticuerpo, un gen o producto génico de receptor de la superfamilia de las inmunoglobulinas (por ejemplo, receptor de linfocitos B (BCR) o receptor de linfocitos T (TCR)), o un fragmento de los mismos.

Las moléculas de anticuerpo humano (y los receptores de linfocitos B) se componen de cadenas pesadas y ligeras con regiones constantes (C) y variables (V) que se codifican por genes en al menos los siguientes tres locus.

1. Locus pesado de inmunoglobulina (IGH@) en el cromosoma 14, que contiene segmentos génicos para la cadena pesada de inmunoglobulina;
2. Locus kappa (κ) de inmunoglobulina (IGK@) en el cromosoma 2, que contiene segmentos génicos para la cadena ligera de inmunoglobulina;
3. Locus lambda (λ) de inmunoglobulina (IGL@) en el cromosoma 22, que contiene segmentos génicos para la cadena ligera de inmunoglobulina.

Cada gen de cadena pesada y cadena ligera contiene múltiples copias de tres tipos diferentes de segmentos génicos para las regiones variables de las proteínas de anticuerpo. Por ejemplo, la región de la cadena pesada de inmunoglobulina puede contener una de cinco clases diferentes γ , δ , α , μ y ϵ , 44 segmentos génicos variables (V), 27 segmentos génicos de diversidad (D) y 6 segmentos génicos de acoplamiento (J). Las cadenas ligeras también pueden poseer numerosos segmentos génicos V y J, pero no tienen segmentos génicos D. La cadena ligera lambda tiene 7 regiones C posibles y la cadena ligera kappa tiene 1.

El locus pesado de inmunoglobulina (IGH@) es una región en el cromosoma humano 14 que contiene genes para las cadenas pesadas de anticuerpos (o inmunoglobulinas) humanos. Por ejemplo, el locus IGH incluye los genes IGHV (variable), IGHD (diversidad), IGHJ (acoplamiento) e IGHC (constante). Los genes ejemplares que codifican las cadenas pesadas de inmunoglobulina incluyen, pero no se limitan a IGHV1-2, IGHV1-3, IGHV1-8, IGHV1-12, IGHV1-14, IGHV1-17, IGHV1-18, IGHV1-24, IGHV1-45, IGHV1-46, IGHV1-58, IGHV1-67, IGHV1-68, IGHV1-69, IGHV1-38-4, IGHV1-69-2, IGHV2-5, IGHV2-10, IGHV2-26, IGHV2-70, IGHV3-6, IGHV3-7, IGHV3-9, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-16, IGHV3-19, IGHV3-20, IGHV3-21, IGHV3-22, IGHV3-23, IGHV3-25, IGHV3-29, IGHV3-30, IGHV3-30-2, IGHV3-30-3, IGHV3-30-5, IGHV3-32, IGHV3-33, IGHV3-33-2, IGHV3-35, IGHV3-36, IGHV3-37, IGHV3-38, IGHV3-41, IGHV3-42, IGHV3-43, IGHV3-47, IGHV3-48, IGHV3-49, IGHV3-50, IGHV3-52, IGHV3-53, IGHV3-54, IGHV3-57, IGHV3-60, IGHV3-62, IGHV3-63, IGHV3-64, IGHV3-65, IGHV3-66, IGHV3-71, IGHV3-72, IGHV3-73, IGHV3-74, IGHV3-75, IGHV3-76, IGHV3-79, IGHV3-38-3, IGHV3-69-1, IGHV4-4, IGHV4-28, IGHV4-30-1, IGHV4-30-2, IGHV4-30-4, IGHV4-31, IGHV4-34, IGHV4-39, IGHV4-55, IGHV4-59, IGHV4-61, IGHV4-80, IGHV4-38-2, IGHV5-51, IGHV5-78, IGHV5-10-1, IGHV6-1, IGHV7-4-1, IGHV7-27, IGHV7-34-1, IGHV7-40, IGHV7-56, IGHV7-81, IGHVD-1-1, IGHV11-15-1, IGHVD-20-1, IGHV11-22-1, IGHV11-26-2, IGHV11-28-1, IGHV11-30-1, IGHV11-31-1, IGHV11-33-1, IGHV11-40-1, IGHV11-43-1, IGHV11-44-2, IGHV11-46-1, IGHVD-49-1, IGHV11-51-2, IGHV11-53-1, IGHV11-60-1, IGHV11-62-1, IGHV11-65-1, IGHV11-67-1, IGHV11-74-1, IGHV11-78-1, IGHV11-81-1, IGHV11-82-1, IGHV11-83-1, IGHV11-84-1, IGHV11-85-1, IGHV11-86-1, IGHV11-87-1, IGHV11-88-1, IGHV11-89-1, IGHV11-90-1, IGHV11-91-1, IGHV11-92-1, IGHV11-93-1, IGHV11-94-1, IGHV11-95-1, IGHV11-96-1, IGHV11-97-1, IGHV11-98-1, IGHV11-99-1, IGHV11-100-1, IGHV11-101-1, IGHV11-102-1, IGHV11-103-1, IGHV11-104-1, IGHV11-105-1, IGHV11-106-1, IGHV11-107-1, IGHV11-108-1, IGHV11-109-1, IGHV11-110-1, IGHV11-111-1, IGHV11-112-1, IGHV11-113-1, IGHV11-114-1, IGHV11-115-1, IGHV11-116-1, IGHV11-117-1, IGHV11-118-1, IGHV11-119-1, IGHV11-120-1, IGHV11-121-1, IGHV11-122-1, IGHV11-123-1, IGHV11-124-1, IGHV11-125-1, IGHV11-126-1, IGHV11-127-1, IGHV11-128-1, IGHV11-129-1, IGHV11-130-1, IGHV11-131-1, IGHV11-132-1, IGHV11-133-1, IGHV11-134-1, IGHV11-135-1, IGHV11-136-1, IGHV11-137-1, IGHV11-138-1, IGHV11-139-1, IGHV11-140-1, IGHV11-141-1, IGHV11-142-1, IGHV11-143-1, IGHV11-144-1, IGHV11-145-1, IGHV11-146-1, IGHV11-147-1, IGHV11-148-1, IGHV11-149-1, IGHV11-150-1, IGHV11-151-1, IGHV11-152-1, IGHV11-153-1, IGHV11-154-1, IGHV11-155-1, IGHV11-156-1, IGHV11-157-1, IGHV11-158-1, IGHV11-159-1, IGHV11-160-1, IGHV11-161-1, IGHV11-162-1, IGHV11-163-1, IGHV11-164-1, IGHV11-165-1, IGHV11-166-1, IGHV11-167-1, IGHV11-168-1, IGHV11-169-1, IGHV11-170-1, IGHV11-171-1, IGHV11-172-1, IGHV11-173-1, IGHV11-174-1, IGHV11-175-1, IGHV11-176-1, IGHV11-177-1, IGHV11-178-1, IGHV11-179-1, IGHV11-180-1, IGHV11-181-1, IGHV11-182-1, IGHV11-183-1, IGHV11-184-1, IGHV11-185-1, IGHV11-186-1, IGHV11-187-1, IGHV11-188-1, IGHV11-189-1, IGHV11-190-1, IGHV11-191-1, IGHV11-192-1, IGHV11-193-1, IGHV11-194-1, IGHV11-195-1, IGHV11-196-1, IGHV11-197-1, IGHV11-198-1, IGHV11-199-1, IGHV11-200-1, IGHV11-201-1, IGHV11-202-1, IGHV11-203-1, IGHV11-204-1, IGHV11-205-1, IGHV11-206-1, IGHV11-207-1, IGHV11-208-1, IGHV11-209-1, IGHV11-210-1, IGHV11-211-1, IGHV11-212-1, IGHV11-213-1, IGHV11-214-1, IGHV11-215-1, IGHV11-216-1, IGHV11-217-1, IGHV11-218-1, IGHV11-219-1, IGHV11-220-1, IGHV11-221-1, IGHV11-222-1, IGHV11-223-1, IGHV11-224-1, IGHV11-225-1, IGHV11-226-1, IGHV11-227-1, IGHV11-228-1, IGHV11-229-1, IGHV11-230-1, IGHV11-231-1, IGHV11-232-1, IGHV11-233-1, IGHV11-234-1, IGHV11-235-1, IGHV11-236-1, IGHV11-237-1, IGHV11-238-1, IGHV11-239-1, IGHV11-240-1, IGHV11-241-1, IGHV11-242-1, IGHV11-243-1, IGHV11-244-1, IGHV11-245-1, IGHV11-246-1, IGHV11-247-1, IGHV11-248-1, IGHV11-249-1, IGHV11-250-1, IGHV11-251-1, IGHV11-252-1, IGHV11-253-1, IGHV11-254-1, IGHV11-255-1, IGHV11-256-1, IGHV11-257-1, IGHV11-258-1, IGHV11-259-1, IGHV11-260-1, IGHV11-261-1, IGHV11-262-1, IGHV11-263-1, IGHV11-264-1, IGHV11-265-1, IGHV11-266-1, IGHV11-267-1, IGHV11-268-1, IGHV11-269-1, IGHV11-270-1, IGHV11-271-1, IGHV11-272-1, IGHV11-273-1, IGHV11-274-1, IGHV11-275-1, IGHV11-276-1, IGHV11-277-1, IGHV11-278-1, IGHV11-279-1, IGHV11-280-1, IGHV11-281-1, IGHV11-282-1, IGHV11-283-1, IGHV11-284-1, IGHV11-285-1, IGHV11-286-1, IGHV11-287-1, IGHV11-288-1, IGHV11-289-1, IGHV11-290-1, IGHV11-291-1, IGHV11-292-1, IGHV11-293-1, IGHV11-294-1, IGHV11-295-1, IGHV11-296-1, IGHV11-297-1, IGHV11-298-1, IGHV11-299-1, IGHV11-300-1, IGHV11-301-1, IGHV11-302-1, IGHV11-303-1, IGHV11-304-1, IGHV11-305-1, IGHV11-306-1, IGHV11-307-1, IGHV11-308-1, IGHV11-309-1, IGHV11-310-1, IGHV11-311-1, IGHV11-312-1, IGHV11-313-1, IGHV11-314-1, IGHV11-315-1, IGHV11-316-1, IGHV11-317-1, IGHV11-318-1, IGHV11-319-1, IGHV11-320-1, IGHV11-321-1, IGHV11-322-1, IGHV11-323-1, IGHV11-324-1, IGHV11-325-1, IGHV11-326-1, IGHV11-327-1, IGHV11-328-1, IGHV11-329-1, IGHV11-330-1, IGHV11-331-1, IGHV11-332-1, IGHV11-333-1, IGHV11-334-1, IGHV11-335-1, IGHV11-336-1, IGHV11-337-1, IGHV11-338-1, IGHV11-339-1, IGHV11-340-1, IGHV11-341-1, IGHV11-342-1, IGHV11-343-1, IGHV11-344-1, IGHV11-345-1, IGHV11-346-1, IGHV11-347-1, IGHV11-348-1, IGHV11-349-1, IGHV11-350-1, IGHV11-351-1, IGHV11-352-1, IGHV11-353-1, IGHV11-354-1, IGHV11-355-1, IGHV11-356-1, IGHV11-357-1, IGHV11-358-1, IGHV11-359-1, IGHV11-360-1, IGHV11-361-1, IGHV11-362-1, IGHV11-363-1, IGHV11-364-1, IGHV11-365-1, IGHV11-366-1, IGHV11-367-1, IGHV11-368-1, IGHV11-369-1, IGHV11-370-1, IGHV11-371-1, IGHV11-372-1, IGHV11-373-1, IGHV11-374-1, IGHV11-375-1, IGHV11-376-1, IGHV11-377-1, IGHV11-378-1, IGHV11-379-1, IGHV11-380-1, IGHV11-381-1, IGHV11-382-1, IGHV11-383-1, IGHV11-384-1, IGHV11-385-1, IGHV11-386-1, IGHV11-387-1, IGHV11-388-1, IGHV11-389-1, IGHV11-390-1, IGHV11-391-1, IGHV11-392-1, IGHV11-393-1, IGHV11-394-1, IGHV11-395-1, IGHV11-396-1, IGHV11-397-1, IGHV11-398-1, IGHV11-399-1, IGHV11-400-1, IGHV11-401-1, IGHV11-402-1, IGHV11-403-1, IGHV11-404-1, IGHV11-405-1, IGHV11-406-1, IGHV11-407-1, IGHV11-408-1, IGHV11-409-1, IGHV11-410-1, IGHV11-411-1, IGHV11-412-1, IGHV11-413-1, IGHV11-414-1, IGHV11-415-1, IGHV11-416-1, IGHV11-417-1, IGHV11-418-1, IGHV11-419-1, IGHV11-420-1, IGHV11-421-1, IGHV11-422-1, IGHV11-423-1, IGHV11-424-1, IGHV11-425-1, IGHV11-426-1, IGHV11-427-1, IGHV11-428-1, IGHV11-429-1, IGHV11-430-1, IGHV11-431-1, IGHV11-432-1, IGHV11-433-1, IGHV11-434-1, IGHV11-435-1, IGHV11-436-1, IGHV11-437-1, IGHV11-438-1, IGHV11-439-1, IGHV11-440-1, IGHV11-441-1, IGHV11-442-1, IGHV11-443-1, IGHV11-444-1, IGHV11-445-1, IGHV11-446-1, IGHV11-447-1, IGHV11-448-1, IGHV11-449-1, IGHV11-450-1, IGHV11-451-1, IGHV11-452-1, IGHV11-453-1, IGHV11-454-1, IGHV11-455-1, IGHV11-456-1, IGHV11-457-1, IGHV11-458-1, IGHV11-459-1, IGHV11-460-1, IGHV11-461-1, IGHV11-462-1, IGHV11-463-1, IGHV11-464-1, IGHV11-465-1, IGHV11-466-1, IGHV11-467-1, IGHV11-468-1, IGHV11-469-1, IGHV11-470-1, IGHV11-471-1, IGHV11-472-1, IGHV11-473-1, IGHV11-474-1, IGHV11-475-1, IGHV11-476-1, IGHV11-477-1, IGHV11-478-1, IGHV11-479-1, IGHV11-480-1, IGHV11-481-1, IGHV11-482-1, IGHV11-483-1, IGHV11-484-1, IGHV11-485-1, IGHV11-486-1, IGHV11-487-1, IGHV11-488-1, IGHV11-489-1, IGHV11-490-1, IGHV11-491-1, IGHV11-492-1, IGHV11-493-1, IGHV11-494-1, IGHV11-495-1, IGHV11-496-1, IGHV11-497-1, IGHV11-498-1, IGHV11-499-1, IGHV11-500-1, IGHV11-501-1, IGHV11-502-1, IGHV11-503-1, IGHV11-504-1, IGHV11-505-1, IGHV11-506-1, IGHV11-507-1, IGHV11-508-1, IGHV11-509-1, IGHV11-510-1, IGHV11-511-1, IGHV11-512-1, IGHV11-513-1, IGHV11-514-1, IGHV11-515-1, IGHV11-516-1, IGHV11-517-1, IGHV11-518-1, IGHV11-519-1, IGHV11-520-1, IGHV11-521-1, IGHV11-522-1, IGHV11-523-1, IGHV11-524-1, IGHV11-525-1, IGHV11-526-1, IGHV11-527-1, IGHV11-528-1, IGHV11-529-1, IGHV11-530-1, IGHV11-531-1, IGHV11-532-1, IGHV11-533-1, IGHV11-534-1, IGHV11-535-1, IGHV11-536-1, IGHV11-537-1, IGHV11-538-1, IGHV11-539-1, IGHV11-540-1, IGHV11-541-1, IGHV11-542-1, IGHV11-543-1, IGHV11-544-1, IGHV11-545-1, IGHV11-546-1, IGHV11-547-1, IGHV11-548-1, IGHV11-549-1, IGHV11-550-1, IGHV11-551-1, IGHV11-552-1, IGHV11-553-1, IGHV11-554-1, IGHV11-555-1, IGHV11-556-1, IGHV11-557-1, IGHV11-558-1, IGHV11-559-1, IGHV11-560-1, IGHV11-561-1, IGHV11-562-1, IGHV11-563-1, IGHV11-564-1, IGHV11-565-1, IGHV11-566-1, IGHV11-567-1, IGHV11-568-1, IGHV11-569-1, IGHV11-570-1, IGHV11-571-1, IGHV11-572-1, IGHV11-573-1, IGHV11-574-1, IGHV11-575-1, IGHV11-576-1, IGHV11-577-1, IGHV11-578-1, IGHV11-579-1, IGHV11-580-1, IGHV11-581-1, IGHV11-582-1, IGHV11-583-1, IGHV11-584-1, IGHV11-585-1, IGHV11-586-1, IGHV11-587-1, IGHV11-588-1, IGHV11-589-1, IGHV11-590-1, IGHV11-591-1, IGHV11-592-1, IGHV11-593-1, IGHV11-594-1, IGHV11-595-1, IGHV11-596-1, IGHV11-597-1, IGHV11-598-1, IGHV11-599-1, IGHV11-600-1, IGHV11-601-1, IGHV11-602-1, IGHV11-603-1, IGHV11-604-1, IGHV11-605-1, IGHV11-606-1, IGHV11-607-1, IGHV11-608-1, IGHV11-609-1, IGHV11-610-1, IGHV11-611-1, IGHV11-612-1, IGHV11-613-1, IGHV11-614-1, IGHV11-615-1, IGHV11-616-1, IGHV11-617-1, IGHV11-618-1, IGHV11-619-1, IGHV11-620-1, IGHV11-621-1, IGHV11-622-1, IGHV11-623-1, IGHV11-624-1, IGHV11-625-1, IGHV11-626-1, IGHV11-627-1, IGHV11-628-1, IGHV11-629-1, IGHV11-630-1, IGHV11-631-1, IGHV11-632-1, IGHV11-633-1, IGHV11-634-1, IGHV11-635-1, IGHV11-636-1, IGHV11-637-1, IGHV11-638-1, IGHV11-639-1, IGHV11-640-1, IGHV11-641-1, IGHV11-642-1, IGHV11-643-1, IGHV11-644-1, IGHV11-645-1, IGHV11-646-1, IGHV11-647-1, IGHV11-648-1, IGHV11-649-1, IGHV11-650-1, IGHV11-651-1, IGHV11-652-1, IGHV11-653-1, IGHV11-654-1, IGHV11-655-1, IGHV11-656-1, IGHV11-657-1, IGHV11-658-1, IGHV11-659-1, IGHV11-660-1, IGHV11-661-1, IGHV11-662-1, IGHV11-663-1, IGHV11-664-1, IGHV11-665-1, IGHV11-666-1, IGHV11-667-1, IGHV11-668-1, IGHV11-669-1, IGHV11-670-1, IGHV11-671-1, IGHV11-672-1, IGHV11-673-1, IGHV11-674-1, IGHV11-675-1, IGHV11-676-1, IGHV11-677-1, IGHV11-678-1, IGHV11-679-1, IGHV11-680-1, IGHV11-681-1, IGHV11-682-1, IGHV11-683-1, IGHV11-684-1, IGHV11-685-1, IGHV11-686-1, IGHV11-687-1, IGHV11-688-1, IGHV11-689-1, IGHV11-690-1, IGHV11-691-1, IGHV11-692-1, IGHV11-693-1, IGHV11-694-1, IGHV11-695-1, IGHV11-696-1, IGHV11-697-1, IGHV11-698-1, IGHV11-699-1, IGHV11-700-1, IGHV11-701-1, IGHV11-702-1, IGHV11-703-1, IGHV11-704-1, IGHV11-705-1, IGHV11-706-1, IGHV11-707-1, IGHV11-708-1, IGHV11-709-1, IGHV11-710-1, IGHV11-711-1, IGHV11-712-1, IGHV11-713-1, IGHV11-714-1, IGHV11-715-1, IGHV11-716-1, IGHV11-717-1, IGHV11-718-1, IGHV11-719-1, IGHV11-720-1, IGHV11-721-1, IGHV11-722-1, IGHV11-723-1, IGHV11-724-1, IGHV11-725-1, IGHV11-726-1, IGHV11-727-1, IGHV11-728-1, IGHV11-729-1, IGHV11-730-1, IGHV11-731-1, IGHV11-732-1, IGHV11-733-1, IGHV11-734-1, IGHV11-735-1, IGHV11-736-1, IGHV11-737-1, IGHV11-738-1, IGHV11-739-1, IGHV11-740-1, IGHV11-741-1, IGHV11-742-1, IGHV11-743-1, IGHV11-744-1, IGHV11-745-1, IGHV11-746-1, IGHV11-747-1, IGHV11-748-1, IGHV11-749-1, IGHV11-750-1, IGHV11-751-1, IGHV11-752-1, IGHV11-753-1, IGHV11-754-1, IGHV11-755-1, IGHV11-756-1, IGHV11-757-1, IGHV11-758-1, IGHV11-759-1, IGHV11-760-1, IGHV11-761-1, IGHV11-762-1, IGHV11-763-1, IGHV11-764-1, IGHV11-765-1, IGHV11-766-1, IGHV11-767-1, IGHV11-768-1, IGHV11-769-1, IGHV11-770-1, IGHV11-771-1, IGHV11-772-1, IGHV11-773-1, IGHV11-774-1, IGHV11-775-1, IGHV11-776-1, IGHV11-777-1, IGHV11-778-1, IGHV11-779-1, IGHV11-780-1, IGHV11-781-1, IGHV11-782-1, IGHV11-783-1, IGHV11-784-1, IGHV11-785-1, IGHV11-786-1, IGHV11-787-1, IGHV11-788-1, IGHV11-789-1, IGHV11-790-1, IGHV11-791-1, IGHV11-792-1, IGHV11-793-1, IGHV11-794-1, IGHV11-795-1, IGHV11-796-1, IGHV11-797-1, IGHV11-798-1, IGHV11-799-1, IGHV11-800-1, IGHV11-801-1, IGHV11-802-1, IGHV11-803-1, IGHV11-804-1, IGHV11-805-1, IGHV11-806-1, IGHV11-807-1, IGHV11-808-1, IGHV11-809-1, IGHV11-810-1, IGHV11-811-1, IGHV11-812-1, IGHV11-813-1, IGHV11-814-1, IGHV11-815-1, IGHV11-816-1, IGHV11-817-1, IGHV11-818-1, IGHV11-819-1, IGHV11-820-1, IGHV11-821-1, IGHV11-822-1, IGHV11-823-1, IGHV11-824-1, IGHV11-825-1, IGHV11-826-1, IGHV11-827-1, IGHV11-828-1, IGHV11-829-1, IGHV11-830-1, IGHV11-831-1, IGHV11-832-1, IGHV11-833-1, IGHV11-834-1, IGHV11-835-1, IGHV11-836-1, IGHV11-837-1, IGHV11-838-1, IGHV11-839-1, IGHV11-840-1, IGHV11-841-1, IGHV11-842-1, IGHV11-843-1, IGHV11-844-1, IGHV11-845-1, IGHV11-846-1, IGHV11-847-1, IGHV11-848-1, IGHV11-849-1, IGHV11-850-1, IGHV11-851-1, IGHV11-852-1, IGHV11-853-1, IGHV11-854-1, IGHV11-855-1, IGHV11-856-1, IGHV11-857-1, IGHV11-858-1, IGHV11-859-1, IGHV11-860-1, IGHV11-861-1, IGHV11-862-1, IGHV11-863-1, IGHV11-864-1, IGHV11-865-1, IGHV11-866-1, IGHV11-867-1, IGHV11-868-1, IGHV11-869-1, IGHV11-870-1, IGHV11-871-1, IGHV11-872-1, IGHV11-873-1, IGHV11-874-1, IGHV11-875-1, IGHV11-876-1, IGHV11-877-1, IGHV11-878-1, IGHV11-879-1, IGHV11-880-1, IGHV11-881-1, IGHV11-882-1, IGHV11-883-1, IGHV11-884-1, IGHV11-885-1, IGHV11-886-1, IGHV11-887-1, IGHV11-888-1, IGHV11-889-1, IGHV11-890-1, IGHV11-891-1, IGHV11-892-1, IGHV11-893-1, IGHV11-894-1, IGHV11-895-1, IGHV11-896-1, IGHV11-897-1, IGHV11-898-1, IGHV11-899-1, IGHV11-900-1, IGHV11-901-1, IGHV11-902-1, IGHV11-903-1, IGHV11-90

5-1, IGHV11-5-2, IGHV11-11-1, IGHV11-13-1, IGHVD1-16-1, IGHV11-22-2, IGHV11-25-1, IGHV11-26-1, IGHV11-38-1, IGHV11-44, IGHV11-47-1, IGHV11-51-1, IGHVD1-67-2, IGHV11-67-3, IGHV11-67-4, IGHV11-76-1, IGHV11-82, IGHV11-44-1, IGHD1-1, IGHD1-7, IGHD1-14, IGHD1-20, IGHD1-26, IGHD2-2, IGHD2-8, IGHD2-15, IGHD2-21, IGHD3-3, IGHD3-9, IGHD3-10, IGHD3-16, IGHD3-22, IGHD4-4, IGHD4-11, IGHD4-17, IGHD4-23, IGHD5-5, IGHD5-12, IGHD5-18, IGHD5-24, IGHD6-6, IGHD6-13, IGHD6-19, IGHD6-25, IGHD7-27, IGHJ1, IGHJ1P, IGHJ2, IGHJ2P, IGHJ3, IGHJ3P, IGHJ4, IGHJ5, IGHJ6, IGHA1, IGHA2, IGHG1, IGHG2, IGHG3, IGHG4, IGHGP, IGHD, IGHE, IGHEP1, IGHM e IGHV1-69D.

El locus kappa de inmunoglobulina (IGK@) es una región en el cromosoma 2 humano que contiene genes para las cadenas ligeras kappa (κ) de anticuerpos (o inmunoglobulinas). Por ejemplo, el locus IGK incluye los genes IGKV (variable), IGKJ (acoplamiento) e IGKC (constante). Los genes ejemplares que codifican las cadenas ligeras kappa de inmunoglobulina incluyen, pero no se limitan a, IGKV1-5, IGKV1-6, IGKV1-8, IGKV1-9, IGKV1-12, IGKV1-13, IGKV1-16, IGKV1-17, IGKV1-22, IGKV1-27, IGKV1-32, IGKV1-33, IGKV1-35, IGKV1-37, IGKV1-39, IGKV1D-8, IGKV1D-12, IGKV1D-13, IGKV1D-16, IGKV1D-17, IGKV1D-22, IGKV1D-27, IGKV1D-32, IGKV1D-33, IGKV1D-35, IGKV1D-37, IGKV1D-39, IGKV1D-42, IGKV1D-43, IGKV2-4, IGKV2-10, IGKV2-14, IGKV2-18, IGKV2-19, IGKV2-23, IGKV2-24, IGKV2-26, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2-36, IGKV2-38, IGKV2-40, IGKV2D-10, IGKV2D-14, IGKV2D-18, IGKV2D-19, IGKV2D-23, IGKV2D-24, IGKV2D-26, IGKV2D-28, IGKV2D-29, IGKV2D-30, IGKV2D-36, IGKV2D-38, IGKV2D-40, IGKV3-7, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3-25, IGKV3-31, IGKV3-34, IGKV3D-7, IGKV3D-11, IGKV3D-15, IGKV3D-20, IGKV3D-25, IGKV3D-31, IGKV3D-34, IGKV4-1, IGKV5-2, IGKV6-21, IGKV6D-21, IGKV6D-41, IGKV7-3, IGKJ1, IGKJ2, IGKJ3, IGKJ4, IGKJ5 e IGKC.

El locus lambda de inmunoglobulina (IGL@) es una región en el cromosoma 22 humano que contiene genes para las cadenas ligeras lambda de anticuerpos (o inmunoglobulinas). Por ejemplo, el locus IGL incluye los genes IGLV (variable), IGLJ (acoplamiento) e IGLC (constante). Los genes ejemplares que codifican las cadenas ligeras lambda de inmunoglobulina incluyen, pero no se limitan a, IGLV1-36, IGLV1-40, IGLV1-41, IGLV1-44, IGLV1-47, IGLV1-50, IGLV1-51, IGLV1-62, IGLV2-5, IGLV2-8, IGLV2-11, IGLV2-14, IGLV2-18, IGLV2-23, IGLV2-28, IGLV2-33, IGLV2-34, IGLV3-1, IGLV3-2, IGLV3-4, IGLV3-6, IGLV3-7, IGLV3-9, IGLV3-10, IGLV3-12, IGLV3-13, IGLV3-15, IGLV3-16, IGLV3-17, IGLV3-19, IGLV3-21, IGLV3-22, IGLV3-24, IGLV3-25, IGLV3-26, IGLV3-27, IGLV3-29, IGLV3-30, IGLV3-31, IGLV3-32, IGLV4-3, IGLV4-60, IGLV4-69, IGLV5-37, IGLV5-39, IGLV5-45, IGLV5-48, IGLV5-52, IGLV6-57, IGLV7-35, IGLV7-43, IGLV7-46, IGLV8-61, IGLV9-49, IGLV10-54, IGLV10-67, IGLV11-55, IGLVI-20, IGLVI-38, IGLVI-42, IGLVI-56, IGLVI-63, IGLVI-68, IGLVI-70, IGLVIV-53, IGLVIV-59, IGLVIV-64, IGLVIV-65, IGLVIV-66-1, IGLVV-58, IGLVV-66, IGLWI-22-1, IGLVVI-25-1, IGLVII-41-1, IGLJ1, IGLJ2, IGLJ3, IGLJ4, IGLJ5, IGLJ6, IGLJ7, IGLC1, IGLC2, IGLC3, IGLC4, IGLC5, IGLC6 e IGLC7.

El receptor de linfocitos B (BCR) se compone de dos partes: i) una molécula de inmunoglobulina unida a la membrana de un isotipo (por ejemplo, IgD o IgM). Con la excepción de la presencia de un dominio de membrana integral, estos pueden ser idénticos a sus formas secretadas, ii) un resto de transducción de señales: un heterodímero llamado Ig- α /Ig- β (CD79), unido por puentes disulfuro. Cada miembro del dímero abarca la membrana plasmática y tiene una cola citoplásmica que porta un motivo de activación de inmunorreceptor basado en tirosina (ITAM).

El receptor de linfocitos T (TCR) se compone de dos cadenas de proteína diferentes (es decir, un heterodímero). En el 95 % de los linfocitos T, este consiste en una cadena alfa (α) y beta (β), mientras que en el 5 % de los linfocitos T este consiste en cadenas gamma y delta (γ/δ). Esta proporción puede cambiar durante la ontogenia y en estados de enfermedad. Los genes de receptor de linfocitos T son similares a los genes de inmunoglobulina en cuanto a que también contienen múltiples segmentos génicos V, D y J en sus cadenas beta y delta (y segmentos génicos V y J en sus cadenas alfa y gamma) que se reordenan durante el desarrollo del linfocito para proporcionar a esa célula un receptor de antígeno único.

El locus alfa del receptor de linfocitos T (TRA) es una región en el cromosoma humano 14 que contiene genes para las cadenas alfa de TCR. Por ejemplo, el locus TRA incluye, por ejemplo, los genes TRAV (variable), TRAJ (acoplamiento) y TRAC (constante). Los genes ejemplares que codifican las cadenas alfa de receptor de linfocitos T incluyen, pero no se limitan a, TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-5, TRAV8-6, TRAV8-7, TRAV9-1, TRAV9-2, TRAV10, TRAV11, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14DV4, TRAV15, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23DV6, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV28, TRAV29DV5, TRAV30, TRAV31, TRAV32, TRAV33, TRAV34, TRAV35, TRAV36DV7, TRAV37, TRAV38-1, TRAV38-2DV8, TRAV39, TRAV40, TRAV41, TRAJ1, TRAJ2, TRAJ3, TRAJ4, TRAJ5, TRAJ6, TRAJ7, TRAJ8, TRAJ9, TRAJ10, TRAJ11, TRAJ12, TRAJ13, TRAJ14, TRAJ15, TRAJ16, TRAJ17, TRAJ18, TRAJ19, TRAJ20, TRAJ21, TRAJ22, TRAJ23, TRAJ24, TRAJ25, TRAJ26, TRAJ27, TRAJ28, TRAJ29, TRAJ30, TRAJ31, TRAJ32, TRAJ33, TRAJ34, TRAJ35, TRAJ36, TRAJ37, TRAJ38, TRAJ39, TRAJ40, TRAJ41, TRAJ42, TRAJ43, TRAJ44, TRAJ45, TRAJ46, TRAJ47, TRAJ48, TRAJ49, TRAJ50, TRAJ51, TRAJ52, TRAJ53, TRAJ54, TRAJ55, TRAJ56, TRAJ57, TRAJ58, TRAJ59, TRAJ60, TRAJ61 y TRAC.

El locus beta del receptor de linfocitos T (TRB) es una región en el cromosoma humano 7 que contiene genes para las cadenas beta de TCR. Por ejemplo, el locus TRB incluye, por ejemplo, los genes TRBV (variable), TRBD

(diversidad), TRBJ (acoplamiento) e TRBC (constante). Los genes ejemplares que codifican las cadenas beta de receptor de linfocitos T incluyen, pero no se limitan a, TRBV1, TRBV2, TRBV3-1, TRBV3-2, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-2, TRBV5-3, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-7, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-7, TRBV6-8, TRBV6-9, TRBV7-1, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-5, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV8-1, TRBV8-2, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-1, TRBV12-2, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV17, TRBV18, TRBV19, TRBV20-1, TRBV21-1, TRBV22-1, TRBV23-1, TRBV24-1, TRBV25-1, TRBV26, TRBV27, TRBV28, TRBV29-1, TRBV30, TRBVA, TRBVB, TRBV5-8, TRBV6-1, TRBD1, TRBD2, TRBJ1-1, TRBJ1-2, TRBJ1-3, TRBJ1-4, TRBJ1-5, TRBJ1-6, TRBJ2-1, TRBJ2-2, TRBJ2-2P, TRBJ2-3, TRBJ2-4, TRBJ2-5, TRBJ2-6, TRBJ2-7, TRBC1 y TRBC2.

El locus delta del receptor de linfocitos T (TRD) es una región en el cromosoma humano 14 que contiene genes para las cadenas delta de TCR. Por ejemplo, el locus TRD incluye, por ejemplo, los genes TRDV (variable), TRDJ (acoplamiento) y TRDC (constante). Los genes ejemplares que codifican las cadenas delta de receptor de linfocitos T incluyen, pero no se limitan a, TRDV1, TRDV2, TRDV3, TRDD1, TRDD2, TRDD3, TRDJ1, TRDJ2, TRDJ3, TRDJ4 y TRDC.

El locus gamma del receptor de linfocitos T (TRG) es una región en el cromosoma humano 7 que contiene genes para las cadenas gamma de TCR. Por ejemplo, el locus TRG incluye, por ejemplo, los genes TRGV (variable), TRGJ (acoplamiento) y TRGC (constante). Los genes ejemplares que codifican las cadenas gamma de receptor de linfocitos T incluyen, pero no se limitan a, TRGV1, TRGV2, TRGV3, TRGV4, TRGV5, TRGV5P, TRGV6, TRGV7, TRGV8, TRGV9, TRGV10, TRGV11, TRGVA, TRGVB, TRGJ1, TRGJ2, TRGJP, TRGJP1, TRGJP2, TRGC1 y TRGC2.

Los cánceres ejemplares incluyen, pero no se limitan a, cáncer de linfocitos B, por ejemplo, mieloma múltiple, melanomas, cáncer de mama, cáncer de pulmón (tal como carcinoma de pulmón no microcítico o CPNM), cáncer de bronquios, cáncer colorrectal, cáncer de próstata, cáncer de páncreas, cáncer de estómago, cáncer de ovario, cáncer de vejiga urinaria, cáncer de cerebro o del sistema nervioso central, cáncer del sistema nervioso periférico, cáncer de esófago, cáncer cervicouterino, cáncer uterino o endometrial, cáncer de la cavidad oral o de faringe, cáncer de hígado, cáncer de riñón, cáncer testicular, cáncer del sistema biliar, cáncer del intestino delgado o del apéndice, cáncer de las glándulas salivales, cáncer de la glándula tiroidea, cáncer de las glándulas suprarrenales, osteosarcoma, condrosarcoma, cáncer de tejidos hematológicos, adenocarcinomas, tumores miofibroblásticos inflamatorios, tumor del estroma gastrointestinal (TEGI), cáncer de colon, mieloma múltiple (MM), síndrome mielodisplásico (SMD), trastorno mieloproliferativo (TMP), leucemia linfocítica aguda (LLA), leucemia mielocítica aguda (LMA), leucemia mielocítica crónica (LMC), leucemia linfocítica crónica (LLC), policitemia Vera, linfoma hodgkiniano, linfoma no hodgkiniano (LNH), sarcoma de tejidos blandos, fibrosarcoma, mixosarcoma, liposarcoma, sarcoma osteogénico, cordoma, angiosarcoma, endoteliosarcoma, linfangiosarcoma, linfangioendoteliosarcoma, sinovioma, mesotelioma, tumor de Ewing, leiomiomasarcoma, rhabdomiomasarcoma, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma de glándulas sebáceas, papilar carcinoma, adenocarcinomas papilares, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de las vías biliares, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, carcinoma de vejiga, carcinoma epitelial, glioma, astrocitoma, meduloblastoma, craneofaringioma,ependimoma, pinealoma, hemangioblastoma, neurinoma del acústico, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, linfoma folicular, linfoma difuso de linfocitos B grandes, linfoma de células del manto, carcinoma hepatocelular, cáncer tiroideo, cáncer gástrico, cáncer de cabeza y cuello, cánceres de células pequeñas, trombocitemia esencial, metaplasia mielocitoide agnógena, síndrome hipereosinofílico, mastocitosis sistémica, hipereosinofilia familiar, leucemia eosinofílica crónica, cánceres neuroendocrinos, tumores carcinoides y similares.

En una implementación, el cáncer es una neoplasia maligna (o premaligna) hematológica. Como se usa en el presente documento, una neoplasia maligna hematológica se refiere a un tumor de los tejidos hematopoyéticos o linfáticos, por ejemplo, un tumor que afecta la sangre, la médula ósea o los ganglios linfáticos. Las neoplasias malignas hematológicas ejemplares incluyen, pero no se limitan a, leucemia (por ejemplo, leucemia linfocítica aguda (LLA), leucemia mielógena aguda (LMA), leucemia linfocítica crónica (LLC), leucemia mielógena crónica (LMC), tricoleucemia, leucemia monocítica aguda (LMOA), leucemia mielomonocítica crónica (LMMC), leucemia mielomonocítica juvenil (LMMJ) o leucemia linfocítica granular de células grandes), linfoma (por ejemplo, linfoma relacionado con el SIDA, linfoma cutáneo de linfocitos T, linfoma hodgkiniano (por ejemplo, linfoma hodgkiniano clásico o linfoma hodgkiniano con predominio linfocitos nodular), micosis fungoide, linfoma no hodgkiniano (por ejemplo, linfoma no hodgkiniano de linfocitos B (por ejemplo, linfoma de Burkitt, linfoma linfocítico de células pequeñas (LLC/LLP), linfoma difuso de linfocitos B grandes, linfoma folicular, linfoma inmunoblástico de células grandes, linfoma linfoblástico de precursores de linfocitos B o linfoma de células del manto) o linfoma no hodgkiniano de linfocitos T (micosis fungoide, linfoma anaplásico de células grandes o linfoma linfoblástico de precursores de linfocitos B)), linfoma primario del sistema nervioso central, síndrome de Sézary, macroglobulinemia de Waldenström), neoplasia mieloproliferativa crónica, histiocitosis de células de Langerhans, mieloma múltiple/neoplasia de células plasmáticas, síndrome mielodisplásico o neoplasia mielodisplásica/mieloproliferativa. Neoplasia premaligna, como se usa en el presente documento, se refiere a un tejido que aún no es maligno pero

que está llamado a volverse maligno.

En una implementación, el gen o producto génico diana, o un fragmento del mismo, se selecciona de cualquiera de los genes o productos génicos descritos en las **tablas 1-4**.

5

Tabla 1. Genes ejemplares con cobertura exónica completa en un conjunto de cebos de ADN-seq ejemplar

ABL1	BTK	CTNNB1	FAS (TNFRSF6)	HIST1H1C	KDR	MYCN	PDK1	RPL13	SUFU
ACTB	BTLA	CUL4A	FAT3	HIST1H1D	KEAP1	MYD88	PHF6	RPL15	SUZ12
AKT1	c11orf30 (EMSY)	CUL4B	FBXO11	HIST1H1E	KIT	MYO18 A	PIK3C2G	RPL35A	SYK
AKT2	CAD	CUX1	FBXO31	HIST1H2A C	KLHL6	NBN	PIK3C3	RPS14	TAF1
AKT3	CARD11	CXCR4	FBXW7	HIST1H2A G	KMT2A (MLL)	NCOR1	PIK3CA	RPS19	TBLIXR1
ALK	CASP8	CYP17A 1	FGF10	HIST1H2A L	KMT2B (MLL2)	NCOR2	PIK3CG	RPS26	TBX3
ALOX12B	CBFB	DAXX	FGF12	HIST1H2A M	KMT2C (MLL3)	NCSTN	PIK3R1	RPTOR	TCF3
AMER1 (FAM123B o WTX)	CBL	DDR2	FGF14	HIST1H2B C	KRAS	NF1	PIK3R2	RUNX1	TCL1A
APC	CCND1	DDX3X	FGF19	HIST1H2B J	LEF1	NF2	PIM1	RUNXIT1	TET2
APCDD1	CCND2	DIS3	FGF23	HIST1H2B K	LMO1	NFE2L2	PLCG2	SIPR2	TGFR2
APH1A	CCND3	DKC1	FGF3	HIST1H2B O	LRP1B	NFKBIA	PMS2	SBDS	TIPARP
AR	CCNE1	DNM2	FGF4	HIST1H3B	LRRK2	NKX2-1	PNRC1	SDHA	TLL2
ARAF	CCT6B	DNMT3A	FGF6	HLA-A	MAF	NOD1	POT1	SDHB	TMEM30A
ARFRP1	CD22	DOT1L	FGF7	HNF1A	MAFB	NOTCH 1	PPP2R1A	SDHC	TMSB4XP8 (TMSL3)
ARHGAP2 6 (GRAF)	CD274 (PDL1)	DTX1	FGFR1	HRAS	MAGED1	NOTCH 2	PRDM1	SDHD	TNFAIP3
ARID1A	CD36	DUSP2	FGFR2	HSP90AA 1	MALT1	NOTCH 3	PRKAR1A	SERP2	TNFRSF11 A
ARID2	CD58	DUSP9	FGFR3	ICK	MAP2K1	NOTCH 4	PRKDC	SETBP1	TNFRSF14
ASMTL	CD70	EBF1	FGFR4	ID3	MAP2K2	NPM1	PRSS8	SETD2	TNFRSF17
ASXL1	CD79A	ECT2L	FHIT	IDH1	MAP2K4	ANR	PTCH1	SF3B1	TOP1
ATM	CD79B	EED	FLCN	IDH2	MAP3K1	NSD1	PTEN	SGK1	TP53
ATR	CDC73	EGFR	FLT1	IGF1	MAP3K1 3	NT5C2	PTPN11	SH2B3	TP63
ATRX	CDH1	ELP2	FLT3	IGF1R	MAP3K1 4	NTRK1	PTPN2	SMAD2	TRAF2
AURKA	CDK12	EP300	FLT4	IGF2	MAP3K6	NTRK2	PTPN6 (SHP-1)	SMAD4	TRAF3
AURKB	CDK4	EPHA3	FLYWCH1	IKBKE	MAP3K7	NTRK3	PTPRO	SMARCA 1	TRAF5
AXIN1	CDK6	EPHA5	FOXL2	IKZF1	MAPK1	NUP93	RAD21	SMARCA 4	TRRAP
AXL	CDK8	EPHA7	FOXO1	IKZF2	MCL1	NUP98	RAD50	SMARCB 1	TSC1
B2M	CDKN1 B	EPHB1	FOXO3	IKZF3	MDM2	P2RY8	RAD51	SMARCD 1	TSC2
BAP1	CDKN2	ERBB2	FOXP1	IL7R	MDM4	PAG1	RAD51B	SMC1A	TSHR

	A								
BARD1	CDKN2 B	ERBB3	FRS2	INHBA	MED12	PAK3	RAD51C	SMC3	TUSC3
BCL10	CDKN2 C	ERBB4	GADD45B	INPP4B	MEF2B	PAK7	RAD51D	SMO	TYK2
BCL11B	CEBPA	ERG	GATA1	INPP5D (SHIP)	MEF2C	PALB2	RAD52	SOCS1	U2AF1
BCL2	CHD2	ESR1	GATA2	IRF1	MEN1	PARP1	RAD54L	SOCS2	U2AF2
BCL2L2	CHEK1	ETS1	GATA3	IRF4	MET	PARP2	RAF1	SOCS3	BVS
BCL6	CHEK2	ETV6	GID4 (c17orf39)	IRF8	MIB1	PARP3	RARA	SOX10	WDR90
BCL7A	CHUK	EXOSC6	GNA11	IRS2	MITF	PARP4	RASGEF1 A	SOX2	WHSC1 (MMSET o NSD2)
BCOR	CIC	EZH2	GNA12	JAK1	MKI67	PASK	RB1	SPEN	WISP3
BCORL1	CIITA	FAF1	GNA13	JAK2	MLH1	PAX5	REL	SPOP	WT1
BIRC3	CKSIB	FAM46C	GNAQ	JAK3	MPL	PBRM1	RELN	SRC	XBP1
BLM	CPS1	FANCA	GNAS	JARID2	MRE11A	PC	RET	SRSF2	XPO1
BRAF	CRBN	FANCC	GPR124	JUN	MSH2	PCBP1	RHOA	STAG2	XRCC3
BRCA1	CREBB P	FANCD2	GRIN2A	KAT6A (MYST3)	MSH3	PCLO	RICTOR	STAT3	YYIAP1
BRCA2	CRKL	FANCE	GSK3B	KDM2B	MSH6	PDCD1	RMRP	STAT4	ZMYM3
BRD4	CRLF2	FANCF	GTSE1	KDM4C	MTOR	PDCD11	RNF43	STAT5A	ZNF217
BRIP1 (BACH1)	CSF1R	FANCG	HDAC1	KDM5A	MUTYH	PDCD1 LG2 (PDL2)	ROS1	STAT5B	ZNF24 (ZSCAN3)
BRSK1	CSF3R	FANC1	HDAC4	KDM5C	MYC	PDGFR A	RPA1	STAT6	ZNF703
BTG1	CTCF	FANCL	HDAC7	KDM6A	MYCL (MYCL1)	PDGFR B	RPL11	STK11	ZRSR2
BTG2	CTNNA 1	FANCM	HGF						

Tabla 2. Genes ejemplares con intrones seleccionados cubiertos en un conjunto de cebos de ADN-seq ejemplar

ALK	BRAF	EGFR	ETV4	EWSR1	IGK	JAK2	NTRK1	RAF1	ROS1
BCL2	CCND1	EPOR	ETV5	FGFR2	IGL	KMT2A (MLL)	PDGFRA	RARA	TMPRSS2
BCL6	CRLF2	ETV1	ETV6	IGH	JAK1	MYC	PDGFRB	RET	TRG
BCR									

5 **Tabla 3.** Genes ejemplares dirigidos en un conjunto de cebos de ARN-seq ejemplar

AB11	BRCA2	CTNNB1	FGFR3	IGH	MDS2	NKX2-1	PIK3R1	RPS14	TET1
ABL1	BTG1	DDIT3	FL11	IGK	MECO M	NOTCH1	PIK3R2	RPS15	TFE3
ABL2	CAMTA1	DDR1	FNBP1	IGL	MEF2C	NPM1	PIM1	RPS19	TFG
ACSL6	CARS	DDX10	FOXO1	IKZF1	MKL1	NR4A3	PLAG1	RPS26	TFPT
AFF1	CBFA2T 3	DDX6	FOXO3	IKZF3	MKL2	NSD1	PML	RUNX1	TFRC
AFF4	CBFB	DEK	FOXO4	IL21R	MLF1	NTRK1	POU2AF1	RUNX1T1 (ETC)	TLX1
ALK	CBL	DLEU2	FOXP1	IL3	MLLT1 (ENL)	NTRK2	PPP1CB	RUNX2	TLX3
ARHGAP2	CCND1	DNMT3A	FSTL3	INSR	MLLT10	NTRK3	PRDM1	SEC31A	TMPRSS2

6					(AF10)				
ARHGEF1 2	CCND2	DUSP22	FUS	IRF4	MLLT3	NUMA1	PRDM16	SEPT5	TNFRSF1 1A
ARID1A	CCND3	EGFR	GAS7	ITK	MLLT4 (AF6)	NUP214	PRRX1	SEPT6	TNFSF9
ARID1B	CD247	EIF4A2	GLI1	JAK1	MLLT6	NUP98	PSIP1	SEPT9	TOP1
ARNT	CD274 (PDL1)	ELF4	GLIS2	JAK2	MN1	NUTM2A	PTCH1	SET	TP53
ASXL1	CD70	ELL	GMPS	JAK3	MNX1	OLIG2	PTEN	SH3GL1	TP63
ATF1	CDC73	ELN	GPHN	JAZF1	MSH2	OMD	PTK7	SLCIA2	TPM3
ATG5	CDK6	EML4	HDAC4	KAT6A (MYST3)	MSH6	P2RY8	RABEP1	SMARCB1	TPM4
ATIC	CDKN2A	EP300	HERPUD1	KDM4C	MS12	PAFAHIB2	RAF1	SNX29 (RUNDC2A)	TRAF2
ATM	CDX2	EPHA7	HEY1	KDSR	MSN	PALB2	RALGDS	SRSF3	TRAF3
ATR	CEBPA	EPOR	HIP1	KIF5B	MTAP	PAX3	RANBP17	SS18	TRAF5
ATXN1	CHIC2	EPS15	HIST1H1A	KMT2A (MLL)	MTCP1	PAX5	RAPIGDS 1	SSX1	TRG
AXL	CHN1	ERBB2	HIST1H41	LASP1	MUC1	PAX7	RARA	SSX2	TRIM24
BAP1	CHTOP (C1orf77)	ERG	HLF	LCK	MYB	PBX1	RB1	SSX4	TRIP11
BCL10	CIC	ETS1	HMGA1	LCP1	MYC	PCM1	RBM15	STAT6	TSC1
BCL11A	CIITA	ETV1	HMGA2	LEF1	MYH11	PCSK7	RCOR1	STK11	TSC2
BCL11B	CKS1B	ETV4	HOXA11	LMO1	MYH9	PDCD1LG 2 (PDL2)	RET	STL	TTL
BCL2	CLP1	ETV5	HOXA13	LMO2	NACA	PDE4DIP	RHOH	SYK	TYK2
BCL3	CLTC	ETV6	HOXA3	LPP	NBEAP 1 (BCL8)	PDGFB	RNF213	TAF15	USP6
BCL6	CLTCL1	EWSR1	HOXA9	LTK	NCOA2	PDGFRA	ROS1	TAL1	WHSC1
BCL7A	CNTRL (CEP110)	FBXW7	HOXC11	LYL1	NDRG1	PDGFRB	RPA1	TAL2	WHSCIL1
BCL9	COLIA1	FCGR2B	HOXC13	MAF	NF1	PDK1	RPL13	TBL1XR1	YPEL5
BOOR	CREB3L 1	FCRL4	HOXD11	MAFB	NF2	PER1	RPL15	TCF3 (E2A)	ZBTB16
BCR	CREB3L 2	FEV	HOXD13	MAGEA 5	NFKB2	PGAM5	RPL22	TCL1A (TCL1)	ZMYM2
BIRC3	CREBBP	FGFR1	HSP90AA 1	MALT1	NFKBIE	PHF1	RPL35A	TCL6	ZNF384
BRAF	CRLF2	FGFR1O P	HSP90AB 1	MAP3K7	N1N	P1CALM	RPN1	TEC	ZNF521
BRCA1	CSF1	FGFR2							

Tabla 4. Genes ejemplares adicionales con cobertura exónica completa en un conjunto de cebos de ADN-seq ejemplar

ASMTL	CKS1B	EXOSC6	GNA12	IKZF2	MIB1	PDCD1	RPL15	SDHC	TCL1A
AXIN1	CPS1	FAF1	GTSE1	IKZF3	MKI67	PDCD11	RPL35A	SDHD	TLL2
BRD4	CSF3R	FBXO11	HDAC1	INPP4B	MSH3	PLCG2	RPS14	SERP2	TMEM30A
BRSK1	CXCR4	FBXO31	HIST1H1D	KMT2C (MLL3)	MYO18A	POT1	RPS19	SOCS2	TP63
BTLA	DDX3X	FHIT	HIST1H2AC	MAF	NCOR2	RASGEF1A	RPS26	SOCS3	TUSC3
CAD	DKC1	FLYWCH1	HIST1H2AM	MAFB	NOD1	RHOA	SIPR2	STAT5A	WDR90
CCT6B	EBF1	FOXP1	HIST1H2BJ	MALT1	PASK	RMRP	SBDS	STAT5B	WHSC1

									(MMSET o NSD2)
CD36	ELP2	FRS2	HNF1A	MAP3K6	PC	RPL11	SDHA	TAF1	YYIAP1
CHD2	ETS1	GADD45B	ICK	MAPK1	PCBP1	RPL13	SDHB	TBLIXR1	ZNF24 (ZSCAN3)
CALR									

En una implementación, el gen o producto génico diana, o un fragmento del mismo, tiene una o más sustituciones, indels o alteraciones en el número de copias que se asocian con cáncer, por ejemplo, una neoplasia maligna (o premaligna) hematológica. Los genes o productos génicos ejemplares incluyen, pero no se limitan a, ABL1, ACTB, AKT1, AKT2, AKT3, ALK, AMER1 (FAM123B o WTX), APC, APH1A, AR, ARAF, ARFRP1, ARHGAP26 (GRAF), ARID1A, ARID2, ASMTL, ASXL1, ATM, ATR, ATRX, AURKA, AURKB, AXIN1, AXL, B2M, BAP1, BARD1, BCL10, BCL11B, BCL2, BCL2L2, BCL6, BCL7A, BCOR, BCORL1, BIRC3, BLM, BRAF, BRCA1, BRCA2, BRIM, BRIP1 (BACH1), BRSK1, BTG2, BTK, BTLA, C11orf30 (EMSY), CAD, CARD11, CBFB, CBL, CCND1, CCND2, CCND3, CCNE1, CCT6B, CD22, CD274, (PDL 1), CD36, CD58, CD70, CD79A, CD79B, CDC73, CDH1, CDK12, CDK4, CDK6, CDK8, CDKN1B, CDKN2A, CDKN2B, CDKN2C, CEBPA, CHD2, CHEK1, CHEK2, CIC, CIITA, CKS1B, CPS1, CREBBP, CRKL, CRLF2, CSF1R, CSF3R, CTCF, CTNNA1, CTNNB1, CUX1, CXCR4, DAXX, DDR2, DDX3X, DNM2, DNMT3A, DOT1L, DTX1, DUSP2, DUSP9, EBF1, ECT2L, EED, EGFR, ELP2, EP300, EPHA3, EPHA5, EPHA7, EPHB1, ERBB2, ERBB3, ERBB4, ERG, ESR1, ETS1, ETV6, EXOSC6, EZH2, FAF1, FAM46C, FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCL, FAS (TNFRSF6), FBXO11, FBXO31, FBXW7, FGF10, FGF14, FGF19, FGF23, FGF3, FGF4, FGF6, FGFR1, FGFR2, FGFR3, FGFR4, FHIT, FLCN, FLT1, FLT3, FLT4, FLYWCH1, FOXL2, FOXO1, FOXO3, FOXP1, FRS2, GADD45B, GATA1, GATA2, GATA3, GID4 (C17orf39), GNA11, GNA12, GNA13, GNAQ, GNAS, GPR124, GRIN2A, GSK3B, GTSE1, HDAC1, HDAC4, HDAC7, HGF, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H2AC, HIST1H2AG, HIST1H2AL, HIST1H2AM, HIST1H2ABC, HIST1H2ABJ, HIST1H2ABK, HIST1H2ABO, HIST1H3B, HNF1A, HRAS, HSP90AA1, ICK, ID3, IDH1, IDH2, IGF1R, IKBKE, IKZF1, IKZF2, IKZF3, IL7R, INHBA, INPP4B, INPP5D (SHIP), IRF1, IRF4, IRF8, IRS2, JAK1, JAK2, JAK3, JARID2, JUN, KAT6A (MYST3), KDM2B, KDM4C, KDM5A, KDM5C, KDM6A, KDR, KEAP1, KIT, KLHL6, KMT2A (MLL), KMT2B (MLL2), KMT2C (MLL3), KRAS, LEF1, LRP1B, LRRK2, MAF, MAFB, MAGED1, MALT1, MAP2K1, MAP2K2, MAP2K4, MAP3K1, MAP3K14, MAP3K6, MAP3K7, MAPK1, MCL1, MDM2, MDM4, MED12, MEF2B, MEF2C, MEN1, MET, MIB1, MITF, MKI67, MLH1, MPL, MRE11A, MSH2, MSH3, MSH6, MTOR, MUTYH, MYC, MYCL (MYCL1), MYCN, MYD88, MYO18A, NCOR2, NCSTN, NF1, NF2, NFE2L2, NFKBIA, NKX2-1, NOD1, NOTCH1, NOTCH2, NPM1, NRAS, NT5C2, NTRK1, NTRK2, NTRK3, NUP93, NUP98, P2RY8, PAG1, PAK3, PALB2, PASK, PAX5, PBRM1, PC, PCBP1, PCLO, PDCD1, PDCD11, PDCD1LG2 (PDL2), PDGFRA, PDGFRB, PDK1, PHF6, PIK3CA, PIK3CG, PIK3R1, PIK3R2, PIM1, PLCG2, POT1, PPP2R1A, PRDM1, PRKAR1A, PRKDC, PRSS8, PTCH1, PTEN, PTPN11, PTPN2, PTPN6 (SHP-1), PTPRO, RAD21, RAD50, RAD51, RAF1, RARA, RASGEF1A, RB1, RELN, RET, RHOA, RICTOR, RNF43, ROS1, RPTOR, RUNX1, S1PR2, SDHA, SDHB, SDHC, SDHD, SERP2, SETBP1, SETD2, SF3B1, SGK1, SMAD2, SMAD4, SMARCA1, SMARCA4, SMARCB1, SMC1A, SMC3, SMO, SOCS1, SOCS2, SOCS3, SOX10, SOX2, SPEN, SPOP, SRC, SRSF2, STAG2, STAT3, STAT4, STAT5A, STAT5B, STAT6, STK11, SUFU, SUZ12, TAF1, TBL1XR1, TCF3, TCL1A, TET2, TGFB2, TLL2, TMEM30A, TMSB4XP8 (TMSL3), TNFAIP3, TNFRSF11A, TNFRSF14, TNFRSF17, TOP1, TP53, TP63, TRAF2, TRAF3, TRAF5, TSC1, TSC2, TSHR, TUSC3, TYK2, U2AF1, U2AF2, VHL, WDR90, WHSC1 (MMSET o NSD2), WISP3, WT1, XBP1, XPO1, YY1 AP1, ZMYM3, ZNF217, ZNF24 (ZSCAN3), ZNF703 o ZRSR2.

En una implementación, el gen o producto génico diana, o un fragmento del mismo, tiene una o más reordenamientos que se asocian con cáncer, por ejemplo, una neoplasia maligna (o premaligna) hematológica. Los genes o productos génicos ejemplares incluyen, pero no se limitan a, ALK, BCL6, BRAF, CRLF2, EPOR, ETV4, ETV6, FGFR2, IGK, BCL2, BCR, CCND1, EGFR, ETV1, ETV5, EWSR1, IGH, IGL, JAK1, KMT2A (MLL), NTRK1, PDGFRB, RARA, ROS1, TRG, JAK2, MYC, PDGFRA, RAF1, RET o TMPRSS2.

En otra implementación, el gen o producto génico diana, o un fragmento del mismo, tiene una o más fusiones que se asocian con cáncer. Los genes o productos génicos ejemplares incluyen, pero no se limitan a, ABI1, CBFA2T3, EIF4A2, FUS, JAK1, MUC1, PBX1, RNF213, TET1, ABL1, CBFB, ELF4, GAS7, JAK2, MYB, PCM1, ROS1, TFE3, ABL2, CBL, ELL, GLI1, JAK3, MYC, PCSK7, RPL22, TFG, ACSL6, CCND1, ELN, GMPS, JAZF1, MYH11, PDCD1LG2 (PDL2), RPN1, TFPT, AFF1, CCND2, EML4, GPHN, KAT6A (MYST3), MYH9, PDE4DIP, RUNX1, TFRC, AFF4, CCND3, EP300, HERPUD1, KDSR, NACA, PDGFB, RUNX1T1 (ETO), TLX1, ALK, CD274 (PDL1), EPOR, HEY1, KIF5B, NBEAP1 (BCL8), PDGFRA, RUNX2, TLX3, ARHGAP26 (GRAF), CDK6, EPS15, HIP1, KMT2A (MLL), NCOA2, PDGFRB, SEC31A, TMPRSS2, ARHGEF12, CDX2, ERBB2, HIST1H41, LASP1, NDRG1, PER1, SEPT5, TNFRSF11A, ARID1A, CFHC2, ERG, HLF, LCP1, NF1, PHF1, SEPT6, TOP1, ARNT, CHN1, ETS1, HMGA1, LMO1, NF2, PICALM, SEPT9, TP63, ASXL1, CIC, ETV1, HMGA2, LMO2, NFKB2, PIM1, SET, TPM3, ATF1, CIITA, ETV4, HOXA11, LPP, NIN, PLAG1, SH3GL1, TPM4, ATG5, CLP1, ETV5, HOXA13, LYL1, NOTCH1, PML, SLC1A2, TRIM24, ATIC, CLTC, ETV6, HOXA3, MAF, NPM1, POU2AF1, SNX29 (RUNDC2A), TRIP11, BCL10, CLTCL1, EWSR1, HOXA9, MAFB, NR4A3, PPP1CB, SRSF3, TTL, BCL11A, CNTRL (CEP110), FCGR2B, HOXC11, MALT1, NSD1, PRDM1, SS18, TYK2, BCL11B, COL1A1, FCRL4, HOXC13, MDS2, NTRK1, PRDM16, SSX1, USP6, BCL2, CREB3L1, FEV, HOXD11, MECOM, NTRK2, PRRX1, SSX2, WHSC1 (MMSET o NSD2),

BCL3, CREB3L2, FGFR1, HOXD13, MKL1, NTRK3, PSIP1, SSX4, WHSC1L1, BCL6, CREBBP, FGFR1OP, HSP90AA1, MLF1, NUMA1, PTCH1, STAT6, YPEL5, BCL7A, CRLF2, FGFR2, HSP90AB1, MLLT1 (ENL), NUP214, PTK7, STL, ZBTB16, BCL9, CSF1, FGFR3, IGH, MLLT10 (AF10), NUP98, RABEP1, SYK, ZMYM2, BCOR, CTNNB1, FLU, IGH, MLLT3, NUTM2A, RAF1, TAF15, ZNF384, BCR, DDIT3, FNBP1, IGL, MLLT4 (AF6), OMD, RALGDS, TAL1, ZNF521, BIRC3, DDX10, FOXO1, IKZF1, MLLT6, P2RY8, RAP1GDS1, TAL2, BRAF, DDX6, FOXO3, IL21R, MN1, PAFAH1B2, RARA, TBL1XR1, BTG1, DEK, FOXO4, IL3, MNX1, PAX3, RBM15, TCF3 (E2A), CAMTA1, DUSP22, FOXP1, IRF4, MSI2, PAX5, RET, TCL1A (TCL1), CARS, EGFR, FSTL3, ITK, MSN, PAX7, RHOH o TEC.

- 10 En una implementación, el gen o producto génico diana, o un fragmento del mismo, se selecciona de cualquiera de los genes o productos génicos descritos en las **tablas 5A-9**.

Tabla 5A. Lista de genes y variantes seleccionados ejemplares, tipos de cáncer asociados y codones de prioridad para el análisis múltiple

Gen Hugo	Categoría génica	Tipos de cáncer	Codones de prioridad
ABL1	Prioridad 1	Leucemia (por ejemplo, leucemia mielógena crónica (LMC), leucemia mielógena aguda (LMA), leucemia linfocítica aguda (LLA))	315
AKT1	Prioridad 1	Cáncer de mama, cáncer colorrectal, cáncer de ovario	
ALK	Prioridad 1	Linfoma (por ejemplo, linfoma no hodgkiniano, linfoma anaplásico de células grandes (LACG)), tumor miofibroblástico inflamatorio	
APC	Prioridad 1	Cáncer colorrectal, meduloblastoma, síndrome de cáncer con alteración de	1114, 1338, 1450, 1556
AR	Prioridad 1	Cáncer de próstata	
BRAF	Prioridad 1	Cáncer de pulmón, linfoma no hodgkiniano, cáncer colorrectal, cáncer tiroideo, melanoma	600
CDKN2A	Prioridad 1	Melanoma, cáncer de páncreas, síndrome de Li-Fraumeni, cáncer de pulmón (por ejemplo, carcinoma de pulmón no microcítico (CPNM)), carcinoma de células escamosas, retinoblastoma, astrocitoma	
CEBPA	Prioridad 1	Leucemia (por ejemplo, leucemia mielógena aguda (LMA), leucemia mielógena aguda (LMA), leucemia monoblástica), retinoblastoma	
CTNNB1	Prioridad 1	Cáncer colorrectal, cáncer de ovario, cáncer de próstata, cáncer de hígado (por ejemplo, hepatoblastoma (HB), carcinoma hepatocelular (CHC)), pilomatrixoma, meduloblastoma, adenomas pleomórficos de las	32, 33, 34, 37, 41, 45
EGFR	Prioridad 1	Cáncer de pulmón, carcinoma de células escamosas, glioblastoma, glioma, cáncer colorrectal	719, 746-750, 768, 790, 858, 861
ERBB2	Prioridad 1	Cáncer gástrico, glioma, cáncer de ovario, cáncer de pulmón	
ESR1	Prioridad 1	Cáncer de mama, cáncer de endometrio, adenocarcinoma de endometrio, liomioma, carcinoma canalicular de la mama	
FGFR1	Prioridad 1	Leucemia, linfoma	

FGFR2	Prioridad 1	Cáncer de mama, cáncer de próstata	
FGFR3	Prioridad 1	Cáncer de vejiga, cáncer cervicouterino, mieloma múltiple	
FLT3	Prioridad 1	Leucemia (por ejemplo, leucemia mielógena aguda (LMA), leucemia promielocítica aguda, leucemia linfocítica aguda)	835
HRAS	Prioridad 1	Carcinoma tiroideo de células de Hurthle, cáncer de vejiga, melanoma, cáncer colorrectal	12, 13, 61
JAK2	Prioridad 1	Leucemia (por ejemplo, leucemia linfoblástica crónica (LLC), leucemia linfocítica aguda (LLA), leucemia mielógena crónica (LMC), leucemia mielógena aguda (LMA))	617
KIT	Prioridad 1	Tumor del estroma gastrointestinal (TEGI), tumor testicular, leucemia (por ejemplo, leucemia mielógena aguda (LMA)), tumor de mastocitos, tumor mesenquimatoso, carcinoma quístico adenoideo, cáncer de pulmón (por ejemplo, carcinoma de pulmón microcítico), linfoma (por ejemplo, linfoma de Burkitt)	816
KRAS	Prioridad 1	Leucemia (por ejemplo, leucemia mielógena aguda (LMA), leucemia mielomonocítica juvenil (LMMJ)), cáncer colorrectal, cáncer de pulmón	12, 13, 61
MET	Prioridad 1	Cáncer gástrico, carcinoma hepatocelular (CHC), carcinoma renal papilar hereditario (CRPH), cáncer de pulmón (por ejemplo, carcinoma de pulmón no microcítico), carcinoma papilar tiroideo, glioma, adenocarcinoma esofágico, osteosarcoma, cáncer de endometrio, carcinoma de células escamosas, melanoma, cáncer de mama	
MLL	Prioridad 1	Leucemia (por ejemplo, leucemia linfocítica aguda (LLA), leucemia mielógena aguda (LMA))	
MYC	Prioridad 1	Leucemia linfocítica crónica (LLC), linfoma de Burkitt, plasmacitoma	
NF1	Prioridad 1	Leucemia (por ejemplo, leucemia mielomonocítica juvenil (LMMJ)), neurofibroma	
NOTCH1	Prioridad 1	Carcinoma de células escamosas, leucemia (por ejemplo, leucemia linfocítica aguda (LLA)), carcinoma medular tiroideo, linfoma (por ejemplo, linfoma tímico, linfoma de linfocitos T)	1575, 1601
NPM1	Prioridad 1	Linfoma (por ejemplo, linfoma no hodgkiniano, linfoma anaplásico de células grandes, linfoma anaplásico), leucemia (por ejemplo, leucemia promielocítica aguda, leucemia mielógena aguda (LMA))	

ANR	Prioridad 1	Leucemia (por ejemplo, leucemia mielomonocítica juvenil (LMMJ), leucemia mielógena aguda (LMA), leucemia linfocítica aguda), melanoma	12, 13, 61
PDGFRA	Prioridad 1	Tumor del estroma gastrointestinal (TEGI), leucemia (por ejemplo, leucemia eosinofílica crónica (LEC), leucemia linfocítica aguda (LLA)), tumor mesenquimatoso	
PIK3CA	Prioridad 1	Cáncer colorrectal, cáncer de mama, cáncer de ovario, carcinoma hepatocelular, carcinoma de células escamosas de cabeza y cuello (CCECC), carcinoma anaplásico tiroideo, cáncer de endometrio, adenocarcinoma de vesícula biliar, glioblastoma	88, 542, 545, 546, 1047, 1049
PTEN	Prioridad 1	Carcinomas de células escamosas de cabeza y cuello (CCECC), cáncer de endometrio, glioma, cáncer de próstata, glioblastoma	130, 173, 233, 267
RB1	Prioridad 1	Retinoblastoma, cáncer de vejiga, osteosarcoma, cáncer de pulmón (por ejemplo, carcinoma de pulmón microcítico, carcinoma de pulmón no microcítico), leucemia (por ejemplo, leucemia linfocítica aguda (LLA))	
RET	Prioridad 1	Cáncer colorrectal, carcinoma medular tiroideo, neoplasia múltiple tipo 2B, feocromocitoma, neoplasia múltiple tipo 2A, carcinoma papilar tiroideo, cáncer tiroideo, retinoblastoma	918
TP53	Prioridad 1	TP53 está mutado o inactivado con frecuencia en aproximadamente el 60 % de los cánceres, por ejemplo, carcinoma esofágico de células escamosas, síndrome de Li-Fraumeni, carcinomas de células escamosas de cabeza y cuello (CCECC), cáncer de pulmón, carcinoma corticosuprarrenal hereditario, astrocitoma, carcinoma de células escamosas, cáncer de vejiga, cáncer colorrectal, glioblastoma, retinoblastoma	175, 245, 248, 273, 306
ABL2	Gen de cáncer	Leucemia mielógena aguda (LMA)	
AKT2	Gen de cáncer	Cáncer de ovario, cáncer de páncreas	
AKT3	Gen de cáncer	Melanoma, glioma, cáncer uterino, cáncer de próstata, cáncer oral, cáncer de ovario	
ARAF	Gen de cáncer	Linfoma angioinmunoblástico de linfocitos T, tumor ascítico de Ehrlich	
ARFRP1	Gen de cáncer	Cáncer de mama	
ARID1A	Gen de cáncer	Neuroblastoma, leucemia linfocítica aguda (LLA), tumor neuroendocrino	
ATM	Gen de cáncer	Leucemia (por ejemplo, leucemia prolinfocítica de linfocitos T (T-PLL)), linfoma, meduloblastoma, glioma	

ATR	Gen de cáncer	Linfoma asociado a pitorax, linfoma de linfocitos T	
AURKA	Gen de cáncer	Carcinoma laríngeo de células escamosas, cáncer de ovario, cáncer de vejiga, carcinoma de células escamosas de cabeza y cuello (CCECC), carcinoma de laringe, carcinoma esofágico de células escamosas (CECE), cáncer de páncreas	
AURKB	Gen de cáncer	Cáncer colorrectal, astrocitoma, tumor ependimario, glioma, carcinoma esofágico de células escamosas (CECE), leucemia mielógena aguda (LMA)	
BCL2	Gen de cáncer	Linfoma, adenocarcinoma colorrectal, carcinoma esofágico de células escamosas (CECE), sarcoma sinovial, leucemia	
BCL2A1	Gen de cáncer	Granuloma pulmonar, adenoma gástrico, linfoma de Burkitt, adenoma parotídeo, sarcoma de Kaposi, cáncer gástrico, cáncer de colon	
BCL2L1	Gen de cáncer	Carcinoma de células escamosas de cabeza y cuello, glioblastoma, mesotelioma, cáncer de páncreas, adenocarcinoma de pulmón	
BCL2L2	Gen de cáncer	Cáncer cerebral, leucemia, linfoma, adenocarcinoma colorrectal, cáncer colorrectal, adenoma, carcinoma cervicouterino de células escamosas	
BCL6	Gen de cáncer	Linfoma, leucemia	
BRCA1	Gen de cáncer	Cáncer de mama, cáncer de ovario	
BRCA2	Gen de cáncer	Cáncer de mama, cáncer de ovario, cáncer de páncreas	
CARD11	Gen de cáncer	Linfoma	
CBL	Gen de cáncer	Linfoma, leucemia	
CCND1	Gen de cáncer	Leucemia linfocítica crónica (LLC), leucemia linfocítica aguda de linfocitos B (LLA-B), cáncer de mama	
CCND2	Gen de cáncer	Retinoblastoma, linfoma de células del manto, leucemia linfocítica aguda de linfocitos T (LLA-T), linfoma de Burkitt, tumor testicular de células reproductoras, tumor ovárico de células de la granulosa, mieloma múltiple	
CCND3	Gen de cáncer	Retinoblastoma, linfoma de células del manto, linfoma anaplásico de células grandes, linfoma (no hodgkiniano), linfoma de linfocitos B, carcinoma laríngeo de células escamosas, linfoma de escasa malignidad, adenoma de linfocitos nulos	
CCNE1	Gen de cáncer	Cáncer de mama, cáncer de ovario, cáncer de vejiga, retinoblastoma	

CDH1	Gen de cáncer	Cáncer gástrico, carcinoma lobular, carcinoma de células escamosas, carcinoma canalicular invasivo, carcinoma lobular invasivo	
CDH2	Gen de cáncer	Melanoma, mesotelioma maligno, mesotelioma pleural, melanoma desmoplásico, adenocarcinoma de pulmón, tumor endometrioide, mesotelioma, cáncer de vejiga, carcinoma esofágico de células escamosas (CECE)	
CDH20	Gen de cáncer	Cáncer de mama	
CDH5	Gen de cáncer	Granuloma, sarcoma epiteliode	
CDK4	Gen de cáncer	Melanoma	
CDK6	Gen de cáncer	Leucemia linfocítica aguda (LLA)	
CDK8	Gen de cáncer	Cáncer de colon, cáncer de pulmón, cáncer rectal, leucemia linfocítica aguda (LLA)	
CDKN2B	Gen de cáncer	Leucemia, retinoblastoma, carcinoma laríngeo de células escamosas	
CDKN2C	Gen de cáncer	Carcinoma tiroideo, adenoma hipofisario, oligodendroglioma, tumor endocrino pancreático, mieloma múltiple, hepatoblastoma, tumor linfático, neoplasia endocrina múltiple tipo 1, oligodendroglioma anaplásico	
CHEK1	Gen de cáncer	Leucemia, cáncer de colon	
CHEK2	Gen de cáncer	Cáncer de mama	
CRKL	Gen de cáncer	Leucemia, linfoma	
CRLF2	Gen de cáncer	Leucemia	
DNMT3A	Gen de cáncer	Tumor testicular de células reproductoras, linfosarcoma, carcinoma hepatocelular, tumor de glándulas salivales	
DOT1L	Gen de cáncer	Leucemia	
EPHA3	Gen de cáncer	Rabdomiosarcoma, linfoma, cáncer de próstata, carcinoma hepatocelular, leucemia, melanoma	
EPHA5	Gen de cáncer	Glioblastoma, cáncer de mama, astrocitoma, tumor de Wilms, glioma	
EPHA6	Gen de cáncer	Cáncer de mama	
EPHA7	Gen de cáncer	Glioblastoma multiforme (GBM), cáncer de colon, cáncer duodenal, tumor paratiroideo, cáncer de próstata	
EPHB1	Gen de cáncer	Cáncer colorrectal, carcinoma embrionario, cáncer gástrico, teratocarcinoma, carcinoma mucinoso	
EPHB4	Gen de cáncer	Carcinoma de células escamosas de cabeza y cuello (CCECC), cáncer cerebral, cáncer de endometrio, cáncer de ovario	
EPHB6	Gen de cáncer	Neuroblastoma, melanoma, carcinoma de pulmón no microcítico (CPNM)	

ERBB3	Gen de cáncer	Cáncer de mama, carcinoma de pulmón no microcítico (CPNM), cáncer de páncreas, carcinoma canalicular invasivo, adenocarcinoma de pulmón, carcinoma endometriode, astrocitoma pilocítico	
ERBB4	Gen de cáncer	Cáncer de mama, meduloblastoma, carcinoma cervicouterino de células escamosas, cáncer de próstata, leucemia	
ERG	Gen de cáncer	Cáncer de próstata, sarcoma de Ewing, leucemia, cáncer de próstata	
ETV1	Gen de cáncer	Cáncer de próstata, cáncer de mama, sarcoma de Ewing, tumor desmoplásico de células redondas pequeñas, liposarcoma mixoide, sarcoma de células claras	
ETV4	Gen de cáncer	Cáncer de mama, cáncer de ovario, carcinoma de células escamosas de la lengua, sarcoma de Ewing	
ETV5	Gen de cáncer	Ganglioglioma, tumor cerebral	
ETV6	Gen de cáncer	Leucemia, fibrosarcoma congénito, carcinoma secretor, síndrome mielodisplásico	
EWSR1	Gen de cáncer	Sarcoma de Ewing, sarcoma de células claras, tumor desmoplásico de células redondas pequeñas, condrosarcoma mixoide extraesquelético, liposarcoma mixoide, histiocitoma fibroso angiomatoide	
EZH2	Gen de cáncer	Cáncer de próstata, adenocarcinoma de vesícula biliar, cáncer de mama, cáncer de vejiga, cáncer gástrico, sarcoma de Ewing	
FANCA	Gen de cáncer	Leucemia	
FBXW7	Gen de cáncer	Cáncer colorrectal, cáncer de endometrio, leucemia linfocítica aguda de linfocitos T (LLA-T)	
FGFR4	Gen de cáncer	Tumor hipofisario, cáncer de próstata, cáncer de pulmón, astrocitoma, rabdomiosarcoma, adenoma hipofisario, fibroadenoma	
FLT1	Gen de cáncer	Cáncer de mama, cáncer de próstata	
FLT4	Gen de cáncer	Cáncer de pulmón, sarcoma de Kaposi, cáncer gástrico, linfangioma, carcinoma de células escamosas	
FOXP4	Gen de cáncer	Linfoma, tumor cerebral	
GATA1	Gen de cáncer	Leucemia megacarioblástica del síndrome de Down	
GNA11	Gen de cáncer	Cáncer de mama	
GNAQ	Gen de cáncer	Melanoma maligno uveal	
GNAS	Gen de cáncer	Adenoma hipofisario	
GPR124	Gen de cáncer	Cáncer de colon	
GUCY1A2	Gen de cáncer	Cáncer de mama	

HOXA3	Gen de cáncer	Cáncer de mama	
HSP90AA1	Gen de cáncer	Linfoma, mieloma	
IDH1	Gen de cáncer	Glioblastoma multiforme (GBM)	
IDH2	Gen de cáncer	Glioblastoma multiforme (GBM)	
IGF1R	Gen de cáncer	Sarcoma de Ewing, cáncer de mama, melanoma maligno uveal, carcinoma corticosuprarrenal, cáncer de páncreas	
IGF2R	Gen de cáncer	Tumor gastrointestinal, cáncer de hígado	
IKBKE	Gen de cáncer	Cáncer de mama	
IKZF1	Gen de cáncer	Linfoma, leucemia	
INHBA	Gen de cáncer	Eritroleucemia, metaplasia de Barrett, adenocarcinoma esofágico, tumor de células de la granulosa, tumor del estroma de los cordones sexuales, adenocarcinoma de pulmón, feocromocitoma, tumor de Krukenberg, cáncer de ovario	
IRS2	Gen de cáncer	Hiperinsulinemia, leiomiomasarcoma uterino	
JAK1	Gen de cáncer	Leucemia, cáncer de ovario, cáncer de mama	
JAK3	Gen de cáncer	Leucemia linfocítica aguda (LLA)	
JUN	Gen de cáncer	Cáncer de piel, leucemia	
KDR	Gen de cáncer	Carcinoma de pulmón no microcítico (CPNM), angiosarcoma	
LRP1B	Gen de cáncer	Cáncer de pulmón, cáncer gástrico, cáncer esofágico	
LTK	Gen de cáncer	Linfoma, cáncer de mama	
MAP2K1	Gen de cáncer	Cáncer de próstata, cáncer gástrico	
MAP2K2	Gen de cáncer	Cáncer de páncreas, tumor intestinal	
MAP2K4	Gen de cáncer	Cáncer de páncreas, cáncer de mama, cáncer colorrectal	
MCL1	Gen de cáncer	Mieloma múltiple, leucemia, linfoma	
MDM2	Gen de cáncer	Sarcoma, glioma, cáncer colorrectal	
MDM4	Gen de cáncer	Glioblastoma multiforme (GBM), cáncer de vejiga, retinoblastoma	
MEN1	Gen de cáncer	Tumor paratiroideo	
MITF	Gen de cáncer	Melanoma	
MLH1	Gen de cáncer	Cáncer colorrectal, cáncer de endometrio, cáncer de ovario, cáncer del SNC	
MPL	Gen de cáncer	Trastorno mieloproliferativo (TMP)	
MRE11A	Gen de cáncer	Cáncer de mama, linfoma	
MSH2	Gen de cáncer	Cáncer colorrectal, cáncer de endometrio, cáncer de ovario	
MSH6	Gen de cáncer	Cáncer colorrectal	
MTOR	Gen de cáncer	Linfoma, cáncer de pulmón, cáncer renal, carcinoma de células claras, glioma	

MUTYH	Gen de cáncer	Cáncer colorrectal	
MYCL1	Gen de cáncer	Carcinoma de pulmón microcítico (CPM)	
MYCN	Gen de cáncer	Neuroblastoma	
NF2	Gen de cáncer	Meningioma, neuroma del acústico, cáncer renal	
NKX2-1	Gen de cáncer	Cáncer de pulmón, cáncer tiroideo, adenocarcinoma	
NTRK1	Gen de cáncer	Cáncer papilar tiroideo	
NTRK3	Gen de cáncer	Fibrosarcoma congénito, cáncer de mama secretor	
PAK3	Gen de cáncer	Cáncer de pulmón	
PAX5	Gen de cáncer	Linfoma no hodgkiniano (LNH), leucemia linfocítica aguda (LLA, por ejemplo, LLA de linfocitos B)	
PDGFRB	Gen de cáncer	Trastorno mieloproliferativo (TMP), leucemia mielógena aguda (LMA), leucemia mielógena crónica (LMC), leucemia mielomonocítica crónica (LMMC)	
PIK3R1	Gen de cáncer	Glioblastoma, cáncer de ovario, cáncer colorrectal	
PKHD1	Gen de cáncer	Cáncer de páncreas	
PLCG1	Gen de cáncer	Cáncer de cabeza y cuello, leucemia	
PRKDC	Gen de cáncer	Glioma, glioblastoma, cáncer gástrico, cáncer de ovario	
PTCH1	Gen de cáncer	Células basales de la piel, meduloblastoma	
PTPN11	Gen de cáncer	Leucemia mielomonocítica juvenil (LMMJ), leucemia mielógena aguda (LMA), síndromes mielodisplásicos (SMD)	
PTPRD	Gen de cáncer	Cáncer de pulmón, carcinoma cutáneo de células escamosas, glioblastoma, neuroblastoma	
RAF1	Gen de cáncer	Astrocitoma pilocítico	
RARA	Gen de cáncer	Leucemia	
RICTOR	Gen de cáncer	Cáncer de colon, linfoma, glioma, cáncer de mama	
RPTOR	Gen de cáncer	Cáncer de mama, cáncer de próstata	
RUNX1	Gen de cáncer	Leucemia mielógena aguda (LMA), leucemia linfocítica aguda de linfocitos preB (LLA-preB), leucemia linfocítica aguda de linfocitos T (LLA-T)	
SMAD2	Gen de cáncer	Carcinoma esofágico de células escamosas (CECE)	
SMAD3	Gen de cáncer	Cáncer de piel, coriocarcinoma	
SMAD4	Gen de cáncer	Cáncer de páncreas, cáncer de colon	
SMARCA4	Gen de cáncer	Carcinoma de pulmón no microcítico (CPNM)	
SMARCB1	Gen de cáncer	Rabdoide maligno	
SMO	Gen de cáncer	Cáncer de células basales de la piel	

SOX10	Gen de cáncer	Oligodendroglioma	
SOX2	Gen de cáncer	Carcinoma embrionario, tumor de células reproductoras	
SRC	Gen de cáncer	Sarcoma, cáncer de colon, cáncer de mama	
STK11	Gen de cáncer	Carcinoma de pulmón no microcítico (CPNM), cáncer de páncreas	
TBX22	Gen de cáncer	Cáncer de mama	
TET2	Gen de cáncer	Síndromes mielodisplásicos (SMD)	
TGFBR2	Gen de cáncer	Cáncer de pulmón, cáncer gástrico, cáncer de colon	
TMPRSS2	Gen de cáncer	Cáncer de próstata	
TOP1	Gen de cáncer	Leucemia mielógena aguda (LMA)	
TSC1	Gen de cáncer	Hamartoma, cáncer de células renales	
TSC2	Gen de cáncer	Hamartoma, cáncer de células renales	
USP9X	Gen de cáncer	Leucemia	
BVS	Gen de cáncer	Cáncer renal, hemangioma, feocromocitoma	
WT1	Gen de cáncer	Tumor de Wilms, tumor desmoplásico de células redondas pequeñas	
ABCB1	Gen PGx		
ABCC2	Gen PGx		
ABCC4	Gen PGx		
ABCG2	Gen PGx		
C1orf144	Gen PGx		
CYP1B1	Gen PGx		
CYP2C19	Gen PGx		
CYP2C8	Gen PGx		
CYP2D6	Gen PGx		
CYP3A4	Gen PGx		
CYP3A5	Gen PGx		
DPYD	Gen PGx		
ERCC2	Gen PGx		
ESR2	Gen PGx		
FCGR3A	Gen PGx		
GSTP1	Gen PGx		
ITPA	Gen PGx		
LRP2	Gen PGx		
MAN1B1	Gen PGx		
MTHFR	Gen PGx		
NQO1	Gen PGx		
NRP2	Gen PGx		
SLC19A1	Gen PGx		
SLC22A2	Gen PGx		
SLCO1B3	Gen PGx		
SOD2	Gen PGx		
SULT1A1	Gen PGx		

TPMT	Gen PGx		
TYMS	Gen PGx		
UGT1A1	Gen PGx		
UMPS	Gen PGx		

"Prioridad 1" se refiere a la prioridad más alta de los genes o productos génicos seleccionados.

5 "Genes de cáncer" se refiere a genes o productos génicos asociados al cáncer de menor prioridad en relación con la Prioridad 1.

"Genes PGx" se refiere a genes que son importantes para farmacogenética y farmacogenómica (PGx).

10 **Tabla 5B.** Genes y variantes seleccionados ejemplares adicionales, tipos de cáncer asociados, codones de prioridad, categoría de aprovechabilidad y tratamientos potenciales

Gen Hugo	Categoría génica	Tipos de cáncer	Codones de prioridad	Categoría de aprovechabilidad	Motivo
ASXL1	Prioridad 1	Mieloma múltiple (MM)		D	Pronóstico (SMD neg.)
BACH1	Prioridad 1	Mama		C	Inhibidores de PARP
BAP1	Prioridad 1	Melanoma maligno uveal, mama, CPNM		C	Inhibidores de PARP
BARD1	Prioridad 1	Mama		C	Inhibidores de PARP
BLM	Prioridad 1	Leucemia, linfoma, cáncer de piel de células escamosas, otros cánceres		C	
BRIP1	Prioridad 1	Leucemia mielógena aguda (LMA), leucemia, mama		C	Inhibidores de PARP
CDKN1B	Prioridad 1	Mama		D	
CREBBP	Prioridad 1	Leucemia linfocítica aguda (LLA), LMA, LDLBG, linfoma no hodgkiniano de linfocitos B (LNH-B)		D	
DDR2	Prioridad 1	CPNM		C	Dasatinib
EMSY	Prioridad 1	Mama		C	Inhibidores de PARP
FANCC	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
FANCD2	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
FANCE	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
FANCF	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
FANCG	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
FANCL	Prioridad 1	LMA, leucemia		C	Inhibidor DE PARP
HGF	Prioridad 1	MM		C	Resistencia
NFKB1	Prioridad 1	Mama		D	Posible MAL PRONÓSTICO

NOTCH2	Prioridad 1	Linfoma de zona marginal, LDLBG		D	-
PALB2	Prioridad 1	Tumor de Wilms, meduloblastoma, LMA, mama		C	Inhibidores de PARP
PBRM1	Prioridad 1	Carcinoma renal de células claras, mama		E	¿Inhibidores de HDAC?
PDK1	Prioridad 1	CPNM		C	Inhibidores de PDK1
PIK3R2	Prioridad 1	CPNM		C	Inhibidores de la vía de PI3K
RAD50	Prioridad 1	Mama		C	Inhibidores de PARP
RAD51	Prioridad 1	Mama		C	Inhibidores de PARP
ROS1	Prioridad 1	Glioblastoma, CPNM		C	
SF3B1	Prioridad 1	SMD, LMC, LLA, pancreático, mama		E	
SPOP	Prioridad 1	Melanoma maligno		E	
ACVR1B	Gen de cáncer	Páncreas, mama		E	
ALOX12B	Gen de cáncer	Mieloma múltiple (MM)		E	
ATRX	Gen de cáncer	Tumores neuroendocrinos pancreáticos		E	
AXL	Gen de cáncer	Carcinoma de pulmón microcítico (CPNM), MM		E	
BCOR	Gen de cáncer	Mama		E	
BCORL1	Gen de cáncer	Mama		E	
C17orf39	Gen de cáncer	Mama		E	
CASP8	Gen de cáncer	Mama		E	
CBFB	Gen de cáncer	LMA		E	
CD22	Gen de cáncer	CPNM, mama		E	
CD79A	Gen de cáncer	Linfoma difuso de linfocitos B grandes (LDLBG)		E	
CD79B	Gen de cáncer	LDLBG		E	
CDC73	Gen de cáncer	Paratiroides		E	
CDK12	Gen de cáncer	Ovario		E	
CHUK	Gen de cáncer	Colorrectal		E	
CRBN	Gen de cáncer	Tubo aerodigestivo superior		E	
CSF1R	Gen de cáncer	CPNM		E	
CTCF	Gen de cáncer	Mama		E	

CTNNA1	Gen de cáncer	Mama		E	
CUL4A	Gen de cáncer	Leucemia		E	
CUL4B	Gen de cáncer	Leucemia		E	
CYP17A1	Gen de cáncer	Mama		E	
DAXX	Gen de cáncer	Tumores neuroendocrinos pancreáticos		E	
DIS3	Gen de cáncer	MM		E	
EP300	Gen de cáncer	Colorrectal, mamario, pancreático, LMA, LLA, LDLBG		E	
ERCC2	Gen de cáncer	Células basales de la piel, células escamosas de la piel, melanoma		E	
FAM46C	Gen de cáncer	MM		E	
FGF1	Gen de cáncer	Mama		E	
FGF10	Gen de cáncer	Mama		E	
FGF12	Gen de cáncer	Mama		E	
FGF14	Gen de cáncer	Mama		E	
FGF19	Gen de cáncer	Mama		E	
FGF23	Gen de cáncer	Mama		E	
FGF3	Gen de cáncer	Mama		E	
FGF4	Gen de cáncer	Mama		E	
FGF6	Gen de cáncer	Mama		E	
FGF7	Gen de cáncer	Mama		E	
FOXL2	Gen de cáncer	Tumor ovárico de células de la granulosa	134	E	
GATA2	Gen de cáncer	LMA, leucemia mielógena crónica (LMC, transformación linfoblástica)		E	
GATA3	Gen de cáncer	Mama		E	
GRAF	Gen de cáncer	LMA, síndrome mielodisplásico (SMD)		E	
GRIN2A	Gen de cáncer	Melanoma maligno		E	
GSK3B	Gen de cáncer	CPNM		E	
HLA-A	Gen de cáncer	MM		E	
IGF1	Gen de cáncer	Mama		E	

IGF2	Gen de cáncer	Mama		E	
IL7R	Gen de cáncer	Leucemia linfocítica aguda de linfocitos T (LLA-T)		E	
INSR	Gen de cáncer	CPNM, glioblastoma, gástrico		E	
IRF4	Gen de cáncer	Mieloma múltiple (MM)		E	
KDM4C	Gen de cáncer	Ovario, mama		E	
KDM5A	Gen de cáncer	LMA		E	
KDM6A	Gen de cáncer	Renal, carcinoma esofágico de células escamosas (CCE), MM		E	
KEAP1	Gen de cáncer	CPNM		E	
KLHL6	Gen de cáncer	Leucemia linfocítica crónica (LLC)		E	
LMO1	Gen de cáncer	Leucemia linfocítica aguda de linfocitos T (LLA-T), neuroblastoma		E	
LRP6	Gen de cáncer	CPNM, melanoma maligno		E	
LRRK2	Gen de cáncer	Ovario, CPNM		E	
MAGED1	Gen de cáncer	MM		E	
MAP3K1	Gen de cáncer	Mama		E	
MAP3K13	Gen de cáncer	Mama		E	
MLL2	Gen de cáncer	Meduloblastoma, renal		E	
MLST8	Gen de cáncer	Mama		E	
MYD88	Gen de cáncer	LDLBG de tipo linfocitos B activados (LDLBG-LBA)		E	
MYST3	Gen de cáncer	Mama		E	
NCOR1	Gen de cáncer	Mama		E	
NFE2L2	Gen de cáncer	CPNM, carcinoma de células escamosas de cabeza y cuello (CCECC)		E	
NFKBIA	Gen de cáncer	Mama		E	
NOTCH3	Gen de cáncer	CPNM, mama		E	
NOTCH4	Gen de cáncer	CPNM, mama		E	
NSD1	Gen de cáncer	LMA		E	
NTRK2	Gen de cáncer	Renal, CPNM		E	

NUP93	Gen de cáncer	Mama		E	
PAK7	Gen de cáncer	CPNM, melanoma maligno		E	
PHLPP2	Gen de cáncer	Ovario, glioblastoma, CPNM		E	
PHOX2B	Gen de cáncer	Neuroblastoma		E	
PIK3C2G	Gen de cáncer	CPNM		E	
PIK3C3	Gen de cáncer	CPNM		E	
PIK3CG	Gen de cáncer	CPNM		E	
PNRC1	Gen de cáncer	MM		E	
PRDM1	Gen de cáncer	LDLBG		E	
PRKAR1A	Gen de cáncer	Glándula suprarrenal, tiroides		E	
PRSS8	Gen de cáncer	Mama		E	
PTCH2	Gen de cáncer	Melanoma maligno		E	
PTK2	Gen de cáncer	CPNM, glioblastoma		E	
PTK2B	Gen de cáncer	CPNM, mama		E	
REL	Gen de cáncer	Linfoma hodgkiniano		E	
RHEB	Gen de cáncer	CPNM, colorrectal		E	
ROCK1	Gen de cáncer	Mama		E	
RUNXT1	Gen de cáncer	CPNM, colorrectal		E	
SETD2	Gen de cáncer	Carcinoma renal de células claras		E	
SH2B3	Gen de cáncer	Síndrome mielodisplásico (SMD)		E	
SOCS1	Gen de cáncer	LDLBG		E	
SPEN	Gen de cáncer	Carcinoma quístico adenoideo		E	
STAG2	Gen de cáncer	Glioblastoma		E	
STAT3	Gen de cáncer	Mama		E	
STAT4	Gen de cáncer	Mama		E	
STK12	Gen de cáncer	TNEP, CPNM		E	
SUFU	Gen de cáncer	Meduloblastoma		E	
TBX23	Gen de cáncer	Mama		E	

TBX3	Gen de cáncer	Mama		E	
TNFAIP3	Gen de cáncer	Linfomas de linfocitos B de zona marginal, linfoma hodgkiniano, linfoma mediastínico primario de linfocitos B		E	
TNFRSF14	Gen de cáncer	Linfoma folicular		E	
TNFRSF17	Gen de cáncer	Linfoma intestinal de linfocitos T		E	
TNKS	Gen de cáncer	CPNM		E	
TNKS2	Gen de cáncer	Melanoma, mama		E	
TRRAP	Gen de cáncer	Colorrectal, glioblastoma		E	
TYK2	Gen de cáncer	CPNM, mama		E	
XBP1	Gen de cáncer	MM		E	
XPO1	Gen de cáncer	Leucemia linfocítica crónica (LLC)		E	
ZNF217	Gen de cáncer	Mama		E	
ZNF703	Gen de cáncer	Mama		E	

Las categorías de aprovechabilidad se clasifican como se describe a continuación. La **tabla 6** proporciona un resumen de la aplicación de las diferentes categorías a alteraciones ejemplares en diferentes tipos de cáncer.

- 5 Categoría A: alteraciones aprobadas / estándar que predicen la sensibilidad o resistencia a tratamientos aprobados / estándar

KRAS G13D en cáncer de colon metastásico

- 10 Amplificación de *ERBB2* en cáncer de mama

EGFR L858R en carcinoma de pulmón microcítico

- 15 Categoría B: alteraciones que son criterios de inclusión o exclusión para tratamientos experimentales específicos

KRAS G13D en cáncer de colon, cáncer de pulmón o cáncer de mama

BRAF V600E en melanoma, cáncer de colon o cáncer de pulmón

- 20 *NRAS* Q61K en melanoma

PIK3CA H1047R en cáncer de mama

- 25 Amplificación de *FGFR1* en cáncer de mama

Inactivación bialélica de *PTEN* en cáncer de mama

Inactivación bialélica de *BRCA1* en cáncer de mama o cáncer de páncreas

- 30 Categoría C: alteraciones con pruebas limitadas (datos clínicos tempranos, datos clínicos en conflicto, datos preclínicos, teóricos) que predicen sensibilidad o resistencia a tratamientos estándar o experimentales

KRAS Q61H en cáncer de colon (clínicos tempranos)

- 35 *PIK3CA* H1047R en cáncer de mama (clínicos en conflicto)

BRAF V600E en cáncer de colon (clínicos en conflicto)

Mutación o amplificación de *ERBB2* en cáncer de pulmón (informes de casos)

BRAF D594G en cáncer de pulmón (preclínicos)

Amplificación de *FGFR1* en cáncer de mama (preclínicos)

Inactivación bialélica de *ATM* en cáncer de mama (preclínicos)

Inactivación bialélica de *TSC1* en cáncer de colon (preclínicos)

Inactivación bialélica de *ATR* en cáncer de mama (teóricos)

Mutación *BRAF* V600E en sarcoma (teóricos)

Categoría D: alteraciones con utilidad pronóstica o diagnóstica en un subtipo particular de cáncer

Inactivación bialélica de *MSH2* en cáncer de colon (pruebas clínicas sólidas)

BRAF V600E en cáncer de pruebas colon (pruebas clínicas sólidas)

KRAS G13D en cáncer de pulmón pruebas (pruebas clínicas sólidas)

Inactivación de *BRCA1* en cáncer de mama (pruebas clínicas sólidas)

Categoría E: alteraciones con clara importancia biológica en el cáncer (es decir, mutaciones oncoincitadoras) sin claras implicaciones clínicas

Inactivación bialélica de *APC* en cáncer de colon

Inactivación bialélica de *TP53* en cáncer de mama

Amplificación de *MITF* en melanoma

ARID1A en cáncer de ovario

Categoría F: alteraciones sin importancia biológica conocida en el cáncer

Alteraciones novedosas en genes de cáncer conocidos

Dianas terapéuticas

Ortólogos de genes de cáncer conocidos

Tabla 6. Clasificación ejemplar de alteraciones en diferentes tipos de cáncer

		A	B	C	D	E
KRAS G13D	Cáncer de colon	x	x		x	x
KRAS G13D	Cáncer de pulmón		x		x	x
KRAS G13D	Cáncer de mama		x			x
NRAS Q61K	Melanoma		x	x		x
KRAS Q61H	Cáncer de colon		x	x		x
BRAF V600E	Melanoma		x			x
BRAF V600E	Cáncer de colon		x	x	x	x
BRAF V600E	Cáncer de pulmón		x			x
BRAF D594G	Cáncer de pulmón			x		x
PIK3CA H1047R	Cáncer de mama		x	x		x

PIK3CA H1047R	Cáncer de colon		x	x		x
EGFR L858R	Cáncer de pulmón	x				x
EGFR T790M	Cáncer de pulmón	x	x			x
Amplificación de ERBB2	Cáncer de mama	x				x
Inactivación bialélica de BRCA1	Cáncer de mama		x	x	x	x
Inactivación bialélica de BRCA2	Cáncer de páncreas		x	x	x	x
Inactivación bialélica de ATM	Cáncer de mama			x		x
Inactivación bialélica de TSC	Cáncer de colon			x		x
Inactivación bialélica de PTEN	Cáncer de colon			x		x
Inactivación bialélica de PTEN	Cáncer de mama		x	x		x
Inactivación bialélica de VHL	Cáncer de riñón				x	x
Inactivación bialélica de MSH2	Cáncer de colon				x	x
Inactivación bialélica de ATR	Cáncer de mama			x		x
Amplificación de MYC	Cáncer de mama			x		x

Tabla 7. Genes seleccionados ejemplares asociados con farmacogenética y farmacogenómica (PGx).

Gen	Locus	Mutación	Efecto
ABCB1	cr7:86976581	3853C>T	Mejor supervivencia en LMA asiática tratada con Ida/AraC; supervivencia en pacientes con cáncer de mama tratadas con paclitaxel
ABCB1	cr7:86998554	2677G>T/A	Respuesta a taxanos, platinos y toxicidad GI; mejor supervivencia en LMA asiática tratada con Ida/AraC
ABCC2	cr10:101610761		Leucopenia inducida por docetaxel
ABCC4	cr13:94613416		Toxicidad de 6MP
ABCG2	cr4:89252551		MTX
ABCG2	cr4:89271347	q141K	Diarrea después de gefitinib
ABCG2	cr4:89274403		MTX
C1orf144	cr1:16578662		Toxicidad de daunorrubicina
CYP1B1	cr2:38151707	CYP1B1*3	Toxicidad de daunorrubicina; supervivencia en pacientes con cáncer de mama tratados con paclitaxel
CYP2C19	cr10:96509051	CYP2C19*17	Beneficio mejorado por tamoxifeno
CYP2C19	cr10:96511647	CYP2C19*17	Beneficio mejorado por tamoxifeno
CYP2C8	cr10:96786964	461delV	Metabolismo del paclitaxel
CYP2C8	cr10:96788739	K399R	Metabolismo del paclitaxel
CYP2C8	cr10:96808096		Metabolismo del paclitaxel
CYP2C8	cr10:96808109		Metabolismo del paclitaxel
CYP2C8	cr10:96817020		Metabolismo del paclitaxel
CYP2D6	cr22:40853554	CYP2D6: 3183 G>A	CYP2D6*29, presente en tanzanos
CYP2D6	cr22:40853749	CYP2D6: 2988 G>A	CYP2D6*41 (IM)
CYP2D6	cr22:40853887	CYP2D6: 2850 C>T	CYP2D6*2 (EM)
CYP2D6	cr22:40854122	CYP2D6: 2613-2615 del AGA	CYP2D6*9 (¿función no clara?)
CYP2D6	cr22:40854188	CYP2D6: 2549 del A	CYP2D6*3

ES 2 989 276 T3

CYP2D6	cr22:40854891	CYP2D6: 1846 G>A	CYP2D6*4
CYP2D6	cr22:40855030	CYP2D6: 1707 del T	CYP2D6*6
CYP2D6	cr22:40855078	CYP2D6: 1659G>A	CYP2D6*29, presente en tanzanos
CYP2D6	cr22:40855716	CYP2D6: 1023 C>T	Presente en CYP2D6*17
CYP2D6	cr22:40856638	CYP2D6: 100C>T	Presente en CYP2D6*10 (casual) y *4 (asociado)
CYP3A4	cr7:99196395		
CYP3A4	cr7:99196460		
CYP3A4	cr7:99197606		
CYP3A4	cr7:99204017		
CYP3A4	cr7:99204029	CYP3A4*16B	Metabolismo del paclitaxel en japoneses
CYP3A4	cr7:99205328		
CYP3A4	cr7:99205363		
CYP3A4	cr7:99219597		
CYP3A4	cr7:99220032	CYP3A4*1B	Mayor eliminación de docetaxel
CYP3A5	cr7:99088330		
CYP3A5	cr7:99100771		
CYP3A5	cr7:99108475		
DPYD	cr1:97688202	DPYD*2A	Toxicidad de 5FU
DPYD	cr1:97753983	DPYD*5	Toxicidad de 5FU
DPYD	cr1:97937679	496A>G	5FU, toxicidad de Xeloda
DPYD	cr1:98121473	DPYD*9A	Toxicidad de 5FU
ERCC2	cr19:50546759	2251A>C	Recidiva después de 5FU en asiáticos
ESR1	cr6:152205074		Hipercolesterolemia inducida por tamoxifeno
ESR2	cr14:63769569		Hipercolesterolemia inducida por tamoxifeno
FCGR3A	cr1:159781166	V158F	Respuesta a cetuximab
FGFR4	cr5:176452849	GLY388ARG	
GSTP1	cr11:67109265	I105V	Resistencia a múltiples quimioterapias
GSTP1	cr11:67110155	A114V	No claro, desequilibrio de acoplamiento con I105V
ITPA	cr20:3141842		Toxicidad de 6MP
LRP2	cr2:169719231		Asociado con ototoxicidad del cisplatino
MAN1B1	cr9:139102689		Toxicidad de daunorrubicina
MTHFR	cr1:11777044		MTX
MTHFR	cr1:11777063		MTX
MTHFR	cr1:11778965	677C>T	MTX
NQO1	cr16:68302646	NQO1*2	Degradación rápida (cisplatino, doxorubicina); mala supervivencia en cáncer de mama tratado con antraciclinas
NRP2	cr2:206360545		Toxicidad de daunorrubicina
SLC19A1	cr21:45782222		MTX
SLC22A2	cr6:160590272	Ala270Ser	Reducida nefrotoxicidad del cisplatino
SLCO1B3	cr12:20936961		Leucopenia inducida por docetaxel
SOD2	cr6:160033862	V16A	Supervivencia inferior en cáncer de mama tratado con ciclofosfamida
SULT1A1	cr16:28524986		
SULT1A1	cr16:28525015		
SULT1A1	cr16:28528073		
SULT1A1	cr16:28528301		

TMPT	cr6:18247207	TPMT*3B	Toxicidad de purinas
TPMT	cr6:18238897		Toxicidad de 6MP
TPMT	cr6:18238991		Toxicidad de 6MP
TPMT	cr6:18251934		Toxicidad de 6MP
TYMS	cr18:647646	Repetición en tándem de 28 pb	Toxicidad de 5FU
TYMS	cr18:663451	Delección de 6 pb	Toxicidad de 5FU
UGT1A1	cr2:234255266		Anemia por irinotecán
UGT1A1	cr2:234255709		Trombocitopenia por irinotecán
UGT1A1	cr2:234330398		UGT1A1*60
UGT1A1	cr2:234330521		UGT1A1*93
UGT1A1	cr2:234333620		UGT1A1*28
UGT1A1	cr2:234333883		UGT1A1*6
UGT1A1	cr2:234334358		UGT1A1*27
UMPS	cr3:125939432	Gly213Ala	Toxicidad de 5FU

Tabla 8. Genes seleccionados ejemplares asociados con mutaciones de translocación en tumores sólidos

Gen Hugo	Categoría génica	Socio de translocación	Tipos de cáncer
ACSL3	Prioridad 1	ETV1	próstata
ALK	Prioridad 1	NPM1, TPM3, TFG, TPM4, ATIC, CLTC, MSN, ALO17, CARS, EML4	LACG, CPNM, neuroblastoma
BRAF	Prioridad 1	AKAP9, KIAA1549	melanoma, colorrectal, papilar tiroideo, ovario de escasa malignidad, carcinoma de pulmón no microcítico (CPNM), colangiocarcinoma, astrocitoma pilocítico
C15orf21	Prioridad 1	ETV1	próstata
CANT1	Prioridad 1	ETV4	próstata
CCND1	Prioridad 1	IGH, FSTL3	LLC, LLA-B, mama
DDX5	Prioridad 1	ETV4	próstata
ELK4	Prioridad 1	SLC45A3	próstata
EML4	Prioridad 1	ALK	CPNM
EP300	Prioridad 1	MLL, RUNXBP2	colorrectal, mama, pancreático, LMA
ERG	Prioridad 1	EWSR1, TMPRSS2, ELF4, FUS, HERPUD1	sarcoma de Ewing, próstata, LMA
ETV1	Prioridad 1	EWSR1, TMPRSS2, SLC45A3, C15orf21, HNRNPA2B1 ACSL3	sarcoma de Ewing, próstata
ETV4	Prioridad 1	EWSR1, TMPRSS2, DDX5, KLK2, CANT1	sarcoma de Ewing, carcinoma de próstata
ETV5	Prioridad 1	TMPRSS2, SCL45A3	próstata
FGFR3	Prioridad 1	IGH@, ETV6	vejiga, MM, linfoma de linfocitos T
HERPUD1	Prioridad 1	ERG	próstata
HNRNPA2B1	Prioridad 1	ETV1	próstata
KLK2	Prioridad 1	ETV4	próstata

RET	Prioridad 1	H4, PRKAR1A, NCOA4, PCM1, GOLGA5, TRIM33, KTN1, TRIM27, HOOK3	medular tiroideo, papilar tiroideo, feocromocitoma
ROS1	Prioridad 1	GOPC, ROS1	glioblastoma, CPNM
SLC45A3	Prioridad 1	ETV1, ETV5, ELK4, ERG	próstata
TMPRSS2	Prioridad 1	ERG, ETV1, ETV4, ETV5	próstata
AKAP9		BRAF	papilar tiroideo
ASPCR1		TFE3	sarcoma alveolar de tejidos blandos
ATF1		EWSR1, FUS	melanoma maligno de tejidos blandos, histiocitoma fibroso angiomatoide
BRD3		NUT	carcinoma letal de la línea media en jóvenes
BRD4		NUT	carcinoma letal de la línea media en jóvenes
C12orf9		LPP	lipoma
CD74		ROS1	CPNM
CDH11		USP6	quistes óseos aneurismáticos
CHCHD7		PLAG1	adenoma salival
CHN1		TAF15	condrosarcoma mixoide extraesquelético
CIC		DUX4	sarcoma de tejidos blandos
CMKOR1		HMGA2	lipoma
COL1A1		PDGFB, USP6	dermatofibrosarcoma protuberante, quiste óseo aneurismático
COX6C		HMGA2	liomioma uterino
CREB1		EWSR1	sarcoma de células claras, histiocitoma fibroso angiomatoide
CREB3L2		FUS	sarcoma fibromixoide
CRTC3		MAML2	mucoepidermoide de las glándulas salivales
CTNNB1		PLAG1	colorrectal, ovario, hepatoblastoma, otros, adenoma salival pleomórfico
D10S170		RET, PDGFRB	papilar tiroideo, LMC
DDIT3		FUS	liposarcoma
DUX4		CIC	sarcoma de tejidos blandos
ELKS		RET	papilar tiroideo
ETV6		NTRK3, RUNX1, PDGFRB, ABL1, MN1, ABL2, FACL6, CHIC2, ARNT, JAK2, EVI1, CDX2, STL, HLXB9, MDS2, PERI, SYK, TTL, FGFR3, PAX5	fibrosarcoma congénito, leucemia y linfoma múltiple, mama secretor, SMD, LLA
EWSR1		FLI1, ERG, ZNF278, NR4A3, FEV, ATF1, ETV1, ETV4, WT1, ZNF384, CREB1, POU5F1, PBX1	sarcoma de Ewing, tumor desmoplásico de células redondas pequeñas, LLA, sarcoma de células claras, sarcoma, mioepitelioma
FEV		EWSR1, FUS	sarcoma de Ewing
FLI1		EWSR1	sarcoma de Ewing
F0X01A		PAX3	rabdomiosarcomas alveolares
FUS		DDIT3, ERG, FEV, ATF1, CREB3L2	liposarcoma, LMA, sarcoma de Ewing, histiocitoma fibroso angiomatoide, sarcoma fibromixoide
GOLGA5		RET	papilar tiroideo
HEI10		HMGA2	liomioma uterino
HMGA1		?	adenoma tiroideo microfolicular, varios tumores mesenquimatosos

			benignos
HMGA2		LHFP, RAD51L1, LPP, HEI10, COX6C, CMKOR1, NFIB	lipoma
HOOK3		RET	papilar tiroideo
JAZF1		SUZ12	tumores del estroma endometrial
KTN1		RET	papilar tiroideo
LHFP		HMGA2	lipoma
LIFR		PLAG1	adenoma salival
LPP		HMGA2, MLL, C12orf9	lipoma, leucemia
MAML2		MECT1, CRTC3	mucoepidermoide de las glándulas salivales
MECT1		MAML2	mucoepidermoide de las glándulas salivales
MN1		ETV6	LMA, meningioma
MYB		NFIB	carcinoma quístico adenoideo
MYC		IGK, BCL5, BCL7A, BTG1, TRA, IGH	linfoma de Burkitt, amplificado en otros cánceres, LLC-B
NCOA1		PAX3	rabdomiosarcoma alveolar
NCOA4		RET	papilar tiroideo
NFIB		MYB, HGMA2	carcinoma quístico adenoideo, lipoma
NONO		TFE3	cáncer renal papilar
NR4A3		EWSR1	condrosarcoma mixoide extraesquelético
NTRK1		TPM3, TPR, TFG	papilar tiroideo
NTRK3		ETV6	fibrosarcoma congénito, mama secretor
NUT		BRD4, BRD3	carcinoma letal de la línea media en jóvenes
OMD		USP6	quistes óseos aneurismáticos
PAX3		FOXO1A, NCOA1	rabdomiosarcoma alveolar
PAX7		FOXO1A	rabdomiosarcoma alveolar
PAX8		PPARG	folicular tiroideo
PBX1		TCF3, EWSR1	LLA-preB, mioepitelioma
PCM1		RET, JAK2	papilar tiroideo, LMC, TMP
PDGFB		COL1A1	DFSP
PDGFRA		FIP1L1	TEGI, síndrome hipereosinofílico idiopático
PLAG1		TCEA1, LIFR, CTNNB1, CHCHD7	adenoma salival
POU5F1		EWSR1	sarcoma
PPARG		PAX8	folicular tiroideo
PRCC		TFE3	papilar renal
PRKAR1A		RET	papilar tiroideo
PRO1073		TFEB	carcinoma de células renales (epitelioide infantil)
RAD51L1		HMGA2	lipoma, liomioma uterino
RAF1		SRGAP3	astrocitoma pilocítico
SFPQ		TFE3	células papilares renales
SRGAP3		RAF1	astrocitoma pilocítico
SS18		SSX1, SSX2	sarcoma sinovial
SS18L1		SSX1	sarcoma sinovial
SSX1		SS18	sarcoma sinovial
SSX2		SS18	sarcoma sinovial
SSX4		SS18	sarcoma sinovial
SUZ12		JAZF1	tumores del estroma endometrial
TAF15		TEC, CHN1, ZNF384	condrosarcomas mixoides extraesqueléticos, LLA

TCEA1		PLAG1	adenoma salival
TCF12		TEC	condrosarcoma mixoide extraesquelético
TFE3		SFPQ, ASPSCR1, PRCC, NONO, CLTC	papilar renal, sarcoma alveolar de tejidos blandos, renal
TFEB		ALPHA	renal (epitelioide infantil)
TFG		NTRK1, ALK	papilar tiroideo, LACG, CPNM
THRAP3		USP6	quistes óseos aneurismáticos
TPM3		NTRK1, ALK	papilar tiroideo, LACG
TPR		NTRK1	papilar tiroideo
TRIM27		RET	papilar tiroideo
TRIM33		RET	papilar tiroideo
USP6		COL1A1, CDH11, ZNF9, OMD	quistes óseos aneurismáticos
ZNF278		EWSR1	sarcoma de Ewing
ZNF331		?	adenoma folicular tiroideo
ZNF9		USP6	quistes óseos aneurismáticos

Tabla 9. Genes seleccionados ejemplares asociados con mutaciones de translocación en neoplasias malignas hematológicas

Gen Hugo	Categoría génica	Socio de translocación	Tipos de cáncer
ABL1	Prioridad 1	BCR, ETV6, NUP214	LMC, LLA, LLA-T
ALK	Prioridad 1	NPM1, TPM3, TFG, TPM4, ATIC, CLTC, MSN, ALO17, CARS, EML4	LACG, CPNM, neuroblastoma
BCL2	Prioridad 1	IGH	LNH, LLC
BCL6	Prioridad 1	locus IG, ZNFN1A1, LCP1, PIM1, TFRC, MHC2TA, NACA, HSPCB, HSPCA, HIST1H4I, IL21R, POU2AF1, ARHH, EIF4A2, SFRS3	LNH, LLC
CCND1	Prioridad 1	IGH, FSTL3	LLC, LLA-B, mama
CREBBP	Prioridad 1	MLL, MORF, RUNXBP2	leucemia aguda (LA), LMA
FGFR1	Prioridad 1	BCR, FOP, ZNF198, CEP1	TMP, LNH
FGFR3	Prioridad 1	IGH, ETV6	vejiga, MM, linfoma de linfocitos T
JAK2	Prioridad 1	ETV6, PCM1, BCR	LLA, LMA, TMP, LMC
MLL	Prioridad 1	MLL, MLLT1, MLLT2, MLLT3, MLLT4, MLLT7, MLLT10, MLLT6, ELL, EPS15, AF1Q, CREBBP, SH3GL1, FBNP1, PNUTL1, MSF, GPHN, GMPS, SSH3BP1, ARHGEF12, GAS7, FOXO3A, LAF4, LCX, SEPT6, LPP, CBFA2T1, GRAF, EP300, PICALM, HEAB	LMA, LLA
PDGFRA	Prioridad 1	FIP1L1	TEGI, síndrome hipereosinofílico idiopático
RARA	Prioridad 1	PML, ZNF145, TIF1, NUMA1, NPM1	leucemia promielocítica aguda (LPA)
SEPT6		MLL	LMA
ABL2		ETV6	LMA
AF15Q14		MLL	LMA
AF1Q		MLL	LLA
AF3p21		MLL	LLA
AF5q31		MLL	LLA
ALO17		ALK	LACG

ARHGEF12		MLL	LMA
ARHH		BCL6	LNH
ARNT		ETV6	LMA
ATIC		ALK	LACG
BCL10		IGH	tejido linfático asociado a mucosas (TLAM)
BCL11A		IGH	LLC-B
BCL11B		TLX3	LLA-T
BCL3		IGH	LLC
BCL5		MYC	LLC
BCL7A		MYC	LNH-B
BCL9		IGH, IGL	LLA-B
BCR		ABL1, FGFR1, JAK2	LMC, LLA, LMA
BIRC3		MALT1	tejido linfático asociado a mucosas (TLAM)
BTG1		MYC	LLC-B
CARS		ALK	LACG
CBFA2T1		MLL, RUNX1	LMA
CBFA2T3		RUNX1	LMA
CBFB		MYH11	LMA
CBL		MLL	LMA, LMMJ, SMD
CCND2		IGL	LNH, LLC
CCND3		IGH	MM
CDK6		MLLT10	LLA
CDX2		ETV6	LMA
CEP1		FGFR1	TMP, LNH
CHIC2		ETV6	LMA
CLTC		ALK, TFE3	LACG, renal
CLTCL1		?	LACG
DDX10		NUP98	LMA*
DDX6		IGH	LNH-B
DEK		NUP214	LMA
EIF4A2		BCL6	LNH
ELF4		ERG	LMA
ELL		MLL	LA
ELN		PAX5	LLA-B
EP300		MLL, RUNXBP2	colorrectal, mama, pancreático, LMA
EPS15		MLL	LLA
ERG		EWSR1, TMPRSS2, ELF4, FUS, HERPUD1	sarcoma de Ewing, próstata, LMA
ETV6		NTRK3, RUNX1, PDGFRB, ABL1, MN1, ABL2, FACL6, CHIC2, ARNT, JAK2, EVI1, CDX2, STL, HLXB9, MDS2, PERI, SYK, TTL, FGFR3, PAX5	fibrosarcoma congénito, leucemia y linfoma múltiple, mama secretor, SMD, LLA
EVI1		RUNX1, ETV6, PRDM16, RPN1	LMA, LMC
EWSR1		FL11, ERG, ZNF278, NR4A3, FEV, ATF1, ETV1, ETV4, WT1, ZNF384, CREB1, POU5F1, PBX1	sarcoma de Ewing, tumor desmoplásico de células redondas pequeñas, LLA, sarcoma de células claras, sarcoma, mioepitelioma
FACL6		ETV6	LMA, eritroleucemia aguda (ELA)
FCGR2B		?	LLA
FGFR1OP		FGFR1	TMP, LNH
FIP1L1		PDGFRA	síndrome hipereosinofílico idiopático
FNBP1		MLL	LMA
FOXO3A		MLL	LA
FOXP1		PAX5	LLA

FSTL3		CCND1	LLC-B
FUS		DDIT3, ERG, FEV, ATF1, CREB3L2	liposarcoma, LMA, sarcoma de Ewing, histiocitoma fibroso angiomatoide, sarcoma fibromixioide
FVT1		IGK	LNH-B
GAS7		MLL	LMA*
GMPS		MLL	LMA
GPHN		MLL	LA
GRAF		MLL	LMA, SMD
HCMOGT-1		PDGFRB	LMMJ
HEAB		MLL	LMA
HIP1		PDGFRB	LMMC
HIST1H4I		BCL6	LNH
HLF		TCF3	LLA
HLXB9		ETV6	LMA
HOXA11		NUP98	LMC
HOXA13		NUP98	LMA
HOXA9		NUP98, MSI2	LMA*
HOXC11		NUP98	LMA
HOXC13		NUP98	LMA
HOXD11		NUP98	LMA
HOXD13		NUP98	LMA*
HSPCA		BCL6	LNH
HSPCB		BCL6	LNH
IGH		MYC, FGFR3,PAX5, IRTA1, IRF4, CCND1, BCL9, BCL8, BCL6, BCL2, BCL3, BCL10, BCL11A, LHX4, DDX6, NFKB2, PAFAH1B2, PCSK7	MM, linfoma de Burkitt, LNH, LLC, LLA-B, TLAM, linfoma mediastínico de células grandes con esclerosis (LMCGE)
IGK		MYC, FVT1	linfoma de Burkitt, LNH-B
IGL		BCL9, MYC, CCND2	linfoma de Burkitt
IL2		TNFRSF17	linfoma intestinal de linfocitos T
IL21R		BCL6	LNH
IRF4		IGH	MM
IRTA1		IGH	LNH-B
ITK		SYK	linfoma periférico de linfocitos T
KDM5A		NUP98	LMA
LAF4		MLL, RUNX1	LLA, LLA-T
LASP1		MLL	LMA
LCK		TRB	LLA-T
LCP1		BCL6	LNH
LCX		MLL	LMA
LMO1		TRD	LLA-T
LMO2		TRD	LLA-T
LPP		HMGA2, MLL, C12orf9	lipoma, leucemia
LYL1		TRB	LLA-T
MAF		IGH	MM
MAFB		IGH	MM
MALT1		BIRC3	tejido linfático asociado a mucosas (TLAM)
MDS1		RUNX1	SMD, LMA
MDS2		ETV6	SMD
MHC2TA		BCL6	LNH
MKL1		RBM15	leucemia megacariocítica aguda
MLF1		NPM1	LMA
MLLT1		MLL	LA
MLLT10		MLL, PICALM, CDK6	LA
MLLT2		MLL	LA

ES 2 989 276 T3

MLLT3		MLL	LLA
MLLT4		MLL	LA
MLLT6		MLL	LA
MLLT7		MLL	LA
MN1		ETV6	LMA, meningioma
MSF		MLL	LMA*
MSI2		HOXA9	LMC
MSN		ALK	LACG
MTCP1		TRA	leucemia prolinfocítica de linfocitos T
MUC1		IGH	LNH-B
MYC		IGK, BCL5, BCL7A, BTG1, TRA, IGH	linfoma de Burkitt, amplificado en otros cánceres, LLC-B
MYH11		CBFB	LMA
MYH9		ALK	LACG
MYST4		CREBBP	LMA
NACA		BCL6	LNH
NCOA2		RUNXBP2	LMA
NFKB2		IGH	LNH-B
NIN		PDGFRB	TMP
NOTCH1		TRB	LLA-T
NPM1		ALK, RARA, MLF1	LNH, LPA, LMA
NSD1		NUP98	LMA
NUMA1		RARA	leucemia promielocítica aguda (LPA)
NUP214		DEK, SET, ABL1	LMA, LLA-T
NUP98		HOXA9, NSD1, WHSC1L1, DDX10, TOPI, HOXD13, PMX1, HOXA13, HOXD11, HOXA11, RAP1GDS1, HOXC11	LMA
OLIG2		TRA	LLA-T
PAFAH1B2		IGH	LMCGE
PAX5		IGH, ETV6, PML, FOXP1, ZNF521, ELN	LNH, LLA, LLA-B
PBX1		TCF3, EWSR1	LLA-preB, mioepitelioma
PCM1		RET, JAK2	papilar tiroideo, LMC, TMP
PCSK7		IGH	LMCGE
PDE4DIP		PDGFRB	TMP
PDGFRB		ETV6, TRIP11, HIP1, RAB5EP, H4, NIN, HCMOGT-1, PDE4DIP	TMP, LMA, LMMC, LMC
PER1		ETV6	LMA, LMMC
PICALM		MLLT10, MLL	LLA-T, LMA
PIM1		BCL6	LNH
PML		RARA, PAX5	LPA, LLA
PMX1		NUP98	LMA
PNUTL1		MLL	LMA
POU2AF1		BCL6	LNH
PRDM16		EVI1	SMD, LMA
PSIP2		NUP98	LMA
RAB5EP		PDGFRB	LMMC
RANBP17		TRD	LLA
RAP1GDS1		NUP98	LLA-T
RBM15		MKL1	leucemia megacariocítica aguda
RPL22		RUNX1	LMA, LMC
RPN1		EVI1	LMA
RUNX1		RPL22, MDS1, EVI1, CBFA2T3, CBFA2T1, ETV6, LAF4	LMA, LLA-preB, LLA-T
RUNXBP2		CREBBP, NCOA2, EP300	LMA

SET		NUP214	LMA
SFRS3		BCL6	linfoma folicular
SH3GL1		MLL	LA
SIL		TAL1	LLA-T
SSH3BP1		MLL	LMA
STL		ETV6	LLA-B
SYK		ETV6, ITK	SMD, linfoma periférico de linfocitos T
TAF15		TEC, CHN1, ZNF384	condrosarcomas mixoides extraesqueléticos, LLA
TAL1		TRD, SIL	leucemia linfoblástica/bifásica
TAL2		TRB	LLA-T
TCF3		PBX1, HLF, TFPT	LLA-preB
TCL1A		TRA	LLC-T
TCL6		TRA	LLA-T
TFG		NTRK1, ALK	papilar tiroideo, LACG, CPNM
TFPT		TCF3	LLA-preB
TFRC		BCL6	LNH
TIF1		RARA	leucemia promielocítica aguda (LPA)
TLX1		TRB, TRD	LLA-T
TLX3		BCL11B	LLA-T
TNFRSF17		IL2	linfoma intestinal de linfocitos T
TOP1		NUP98	LMA*
TPM3		NTRK1, ALK	papilar tiroideo, LACG
TPM4		ALK	LACG
TRA		ATL, OLIG2, MYC, TCL1A, TCL6, MTCP1, TCL6	LLA-T
TRB		HOX11, LCK, NOTCH1, TAL2, LYL1	LLA-T
TRD		TAL1, HOX11, TLX1, LMO1, LMO2, RANBP17	leucemia de linfocitos T
TRIP11		PDGFRB	LMA
TTL		ETV6	LLA
WHSC1		IGH	MM
WHSC1L1		NUP98	LMA
ZNF145		RARA	leucemia promielocítica aguda (LPA)
ZNF198		FGFR1	TMP, LNH
ZNF384		EWSR1, TAF15	LLA
ZNF521		PAX5	LLA
ZNFN1A1		BCL6	LLA, LDLBG

Las aplicaciones de los procedimientos anteriores incluyen el uso de una colección de oligonucleótidos que contiene todas las variantes de secuencia conocidas (o un subconjunto de las mismas) de un gen o genes particulares para la secuenciación en muestras médicas.

5

En determinadas implementaciones, el procedimiento o ensayo incluye además uno o más de:

- (i) realizar una identificación genética de la muestra de ácido nucleico;
- 10 (ii) cuantificar la abundancia de un gen o producto génico (por ejemplo, un gen o producto génico como se describe en el presente documento) en la muestra de ácido nucleico;
- (iii) cuantificar la abundancia relativa de un transcrito en la muestra;
- 15 (iv) identificar la muestra de ácido nucleico como perteneciente a un sujeto particular (por ejemplo, un control normal o un paciente con cáncer);
- (v) identificar un rasgo genético en la muestra de ácido nucleico (por ejemplo, composición genética de uno o más sujetos (por ejemplo, etnia, raza, rasgos familiares));

20

(vi) determinar la ploidía en la muestra de ácido nucleico; determinar una pérdida de heterocigosidad en la muestra de ácido nucleico;

(vii) determinar la presencia o ausencia de un acontecimiento de duplicación génica en la muestra de ácido nucleico;

(viii) determinar la presencia o ausencia de un acontecimiento de amplificación génica en la muestra de ácido nucleico; o

(ix) determinar el nivel de mezcla de células tumorales/normales en la muestra de ácido nucleico.

Muestras de ácido nucleico

Una variedad de muestras tisulares puede ser la fuente de las muestras de ácido nucleico usadas en los presentes procedimientos. El ácido nucleico genómico o subgenómico (por ejemplo, ADN o ARN) se puede aislar de una muestra de sujeto (por ejemplo, una muestra tumoral, un tejido contiguo normal (TCN), una muestra de sangre, una muestra que contenga células tumorales circulantes (CTC) o cualquier control normal). En determinadas implementaciones, la muestra tisular se conserva como una muestra congelada o como una preparación tisular fijada con formaldehído o paraformaldehído e incluida en parafina (FFIP). Por ejemplo, la muestra se puede incluir en una matriz, por ejemplo, un bloque FFIP o una muestra congelada. En determinadas implementaciones, la muestra tisular es una muestra de sangre. En otras implementaciones, la muestra tisular es una muestra de aspirado de médula ósea (AMO). La etapa de aislamiento puede incluir la separación por flujo de cromosomas individuales; y/o microdissección de una muestra de sujeto (por ejemplo, una muestra tumoral, un TCN, una muestra de sangre).

Una molécula de ácido nucleico "aislada" es una que está separada de otras moléculas de ácido nucleico que están presentes en la fuente natural de la molécula de ácido nucleico. En determinadas implementaciones, una molécula de ácido nucleico "aislada" está libre de secuencias (tales como secuencias que codifican proteínas) que naturalmente flanquean el ácido nucleico (es decir, secuencias localizadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del que se deriva el ácido nucleico. Por ejemplo, en diversas implementaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kB, menos de aproximadamente 4 kB, menos de aproximadamente 3 kB, menos de aproximadamente 2 kB, menos de aproximadamente 1 kB, menos de aproximadamente 0,5 kB o menos de aproximadamente 0,1 kB de secuencias de nucleótidos que naturalmente flanquean la molécula de ácido nucleico en el ADN genómico de la célula de la que se deriva el ácido nucleico. Además, una molécula de ácido nucleico "aislada", tal como una molécula de ARN o una molécula de ADNc, puede estar sustancialmente libre de otro material celular o medio de cultivo, por ejemplo, cuando se produce por técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos, por ejemplo, cuando se sintetiza químicamente.

El lenguaje "sustancialmente libre de otro material celular o medio de cultivo" incluye preparaciones de molécula de ácido nucleico en las que la molécula se separa de los componentes celulares de las células de las que se aísla o se produce de forma recombinante. Por tanto, la molécula de ácido nucleico que está sustancialmente libre de material celular incluye preparaciones de molécula de ácido nucleico que tienen menos de aproximadamente un 30 %, menos de aproximadamente un 20 %, menos de aproximadamente un 10 % o menos de aproximadamente un 5 % (en peso seco) de otro material celular o medio de cultivo.

En determinadas implementaciones, el ácido nucleico se aísla de una muestra almacenada, por ejemplo, una muestra FFIP almacenada. La muestra almacenada puede tener, por ejemplo, 1 año, 2 años, 3 años, 4 años, 5 años, 10 años, 15 años, 20 años, 25 años, 50 años, 75 años o 100 años o más.

Se puede obtener una muestra de ácido nucleico a partir de muestras tisulares (por ejemplo, una biopsia, una muestra FFIP, una muestra de sangre o una muestra de aspirado de médula ósea) de diversos tamaños. Por ejemplo, el ácido nucleico se puede aislar de una muestra tisular de 5 a 200 µm, o mayor. Por ejemplo, la muestra tisular puede medir 5 µm, 10 µm, 20 µm, 30 µm, 40 µm, 50 µm, 70 µm, 100 µm, 110 µm, 120 µm, 150 µm o 200 µm o más.

Los protocolos para el aislamiento de ADN de una muestra tisular se proporcionan en el ejemplo 1. Se divulgan procedimientos adicionales para aislar ácidos nucleicos (por ejemplo, ADN) de tejidos fijados con formaldehído o paraformaldehído e incluidos en parafina (FFIP), por ejemplo, en Cronin M. *et al.*, (2004) *Am J Pathol.* 164(1):35-42; Masuda N. *et al.*, (1999) *Nucleic Acids Res.* 27(22):4436-4443; Specht K. *et al.*, (2001) *Am J Pathol.* 158(2):419-429, el protocolo RecoverAll™ Total Nucleic Acid Isolation de Ambion (Ambion, n.º de cat. AM1975, septiembre de 2008), el manual técnico del kit Maxwell® 16 FFPE Plus LEV DNA Purification (Promega, literatura n.º TM349, febrero de 2011), el manual del E.Z.N.A.® FFPE DNA Kit (OMEGA bio-tek, Norcross, GA, números de producto D3399-00, D3399-01 y D3399-02; junio de 2009) y el manual del kit QIAamp® DNA FFPE Tissue (Qiagen, n.º de cat. 37625, octubre de 2007). El kit RecoverAll™ Total Nucleic Acid Isolation usa xileno a temperaturas elevadas para solubilizar muestras incluidas en parafina y un filtro de fibra de vidrio para capturar ácidos nucleicos. El kit

Maxwell® 16 FFPE Plus LEV DNA Purification se usa con el instrumento Maxwell® 16 para la purificación de ADN genómico de secciones de 1 a 10 µm de tejido FFIP. El ADN se purifica usando partículas paramagnéticas (PMP) recubiertas de sílice y se eluye en un volumen de elución bajo. El kit E.Z.N. A.® FFEP DNA usa una columna de centrifugación y un sistema de tampón para el aislamiento del ADN genómico. El kit QIAamp® DNA FFIP Tissue usa la tecnología QIAamp® DNA Micro para la purificación de ADN genómico y mitocondrial. Los protocolos para el aislamiento de ADN de la sangre se divulgan, por ejemplo, en el manual técnico del kit Maxwell® 16 LEV Blood DNA y del kit Maxwell 16 Buccal Swab LEV DNA Purification (Promega, literatura n.º TM333, 1 de enero de 2011).

Los protocolos para el aislamiento de ARN se divulgan, por ejemplo, en el boletín técnico del kit Maxwell® 16 Total RNA Purification (Promega, literatura n.º TB351, agosto de 2009).

Las muestras de ácido nucleico aislado (por ejemplo, muestras de ADN genómico) se pueden fragmentar o cortar poniendo en práctica técnicas rutinarias. Por ejemplo, el ADN genómico se puede fragmentar por procedimientos de corte físico, procedimientos de escisión enzimática, procedimientos de escisión química y otros procedimientos bien conocidos para los expertos en la técnica. La colección de ácidos nucleicos puede contener toda o sustancialmente toda la complejidad del genoma. El término "sustancialmente toda" en este contexto se refiere a la posibilidad de que, en la práctica, pueda existir una pérdida no deseada de la complejidad genómica durante las etapas iniciales del procedimiento. Los procedimientos descritos en el presente documento también son útiles en los casos donde la colección de ácidos nucleicos es una porción del genoma, es decir, donde la complejidad del genoma se reduce por diseño. En algunas implementaciones, cualquier porción seleccionada del genoma se puede usar con los procedimientos descritos en el presente documento. En determinadas implementaciones, se aísla todo el exoma o un subconjunto del mismo.

Los procedimientos presentados en el presente documento pueden incluir además aislar una muestra de ácido nucleico para proporcionar una colección (por ejemplo, una colección de ácidos nucleicos como se describe en el presente documento). En determinadas implementaciones, la muestra de ácido nucleico incluye fragmentos completos genómicos, subgenómicos o ambos. Las muestras de ácido nucleico aislado se pueden usar para preparar colecciones de ácidos nucleicos. Por tanto, en una implementación, los procedimientos presentados en el presente documento incluyen además aislar una muestra de ácido nucleico para proporcionar una colección (por ejemplo, una colección de ácidos nucleicos como se describe en el presente documento). Los protocolos para aislar y preparar colecciones a partir de fragmentos genómicos o subgenómicos completos son conocidos en la técnica (por ejemplo, el kit de preparación de muestras de ADN genómico de Illumina). En determinadas implementaciones, el fragmento de ADN genómico o subgenómico se aísla de una muestra de sujeto (por ejemplo, una muestra tumoral, un tejido adyacente normal (TCN), una muestra de sangre o cualquier control normal). En una implementación, la muestra (por ejemplo, la muestra tumoral o de TCN) es una muestra conservada. Por ejemplo, la muestra se incluye en una matriz, por ejemplo, un bloque FFIP o una muestra congelada. En determinadas implementaciones, la etapa de aislamiento incluye la separación por flujo de cromosomas individuales; y/o microdissección de una muestra de sujeto (por ejemplo, una muestra tumoral, un TCN, una muestra de sangre). En determinadas implementaciones, la muestra de ácido nucleico usada para generar la colección de ácidos nucleicos es de menos de 5 microgramos, de menos de 1 microgramo, de menos de 500 ng, de menos de 200 ng, de menos de 100 ng, de menos de 50 ng o de menos de 20 ng, de menos de 10 ng, de menos de 5 ng o de menos de 1 ng.

Todavía en otras implementaciones, la muestra de ácido nucleico usada para generar la colección incluye ARN o ADNc derivado de ARN. En algunas implementaciones, el ARN incluye ARN celular total. En otras implementaciones, se han disminuido determinadas secuencias de ARN abundantes (por ejemplo, ARN ribosómicos). En algunas implementaciones, se ha enriquecido la fracción de ARNm con cola de poli(A) en la preparación de ARN total. En algunas implementaciones, el ADNc se produce por procedimientos de síntesis de ADNc con cebado aleatorio. En otras implementaciones, la síntesis de ADNc se inicia en la cola de poli(A) de los ARNm maduros mediante cebado con oligonucleótidos que contienen oligo(dT). Los procedimientos para la disminución, el enriquecimiento de poli(A) y la síntesis de ADNc son bien conocidos para los expertos en la técnica.

El procedimiento puede incluir además la amplificación de la muestra de ácido nucleico por procedimientos de amplificación de ácidos nucleicos específicos o no específicos que son bien conocidos para los expertos en la técnica. En algunas implementaciones, en determinadas implementaciones, la muestra de ácido nucleico se amplifica, por ejemplo, por procedimientos de amplificación hologenómica, tales como amplificación por desplazamiento de hebra con cebado aleatorio.

En otras implementaciones, la muestra de ácido nucleico se fragmenta o corta por procedimientos físicos o enzimáticos y se liga a adaptadores sintéticos, se selecciona por tamaño (por ejemplo, por electroforesis en gel preparativa) y se amplifica (por ejemplo, por PCR). En otras implementaciones, el grupo de ácidos nucleicos fragmentados y ligados a adaptador se usa sin selección de tamaño o amplificación explícitas antes de la selección de híbridos.

En otras implementaciones, el ADN aislado (por ejemplo, el ADN genómico) se fragmenta o corta. En algunas implementaciones, la colección incluye menos de un 50 % de ADN genómico, tal como una subfracción de ADN

genómico que es una representación reducida o una porción definida de un genoma, por ejemplo, que se ha subfraccionado por otros medios. En otras implementaciones, la colección incluye todo o sustancialmente todo el ADN genómico.

En algunas implementaciones, la colección incluye menos de un 50 % de ADN genómico, tal como una subfracción de ADN genómico que es una representación reducida o una porción definida de un genoma, por ejemplo, que se ha subfraccionado por otros medios. En otras implementaciones, la colección incluye todo o sustancialmente todo el ADN genómico. Los protocolos para aislar y preparar colecciones a partir de fragmentos genómicos o subgenómicos completos son conocidos en la técnica (por ejemplo, el kit de preparación de muestras de ADN genómico de Illumina) y se describen en el presente documento en los ejemplos. Los procedimientos alternativos para el corte del ADN se describen en el presente documento como ejemplo 4. Por ejemplo, los procedimientos de corte de ADN alternativos pueden ser más automatizables y/o más eficaces (por ejemplo, con muestras FFIP degradadas). También se pueden usar alternativas a los procedimientos de corte de ADN para evitar una etapa de ligación durante la preparación de colecciones.

Los procedimientos descritos en el presente documento se pueden realizar usando una pequeña cantidad de ácidos nucleicos, por ejemplo, cuando la cantidad de ADN o ARN fuente es limitante (por ejemplo, incluso después de la amplificación hologenómica). En una implementación, el ácido nucleico comprende menos de aproximadamente 5 µg, 4 µg, 3 µg, 2 µg, 1 µg, 0,8 µg, 0,7 µg, 0,6 µg, 0,5 µg o 400 ng, 300 ng, 200 ng, 100 ng, 50 ng, 10 ng, 5 ng, 1 ng o menos de muestra de ácido nucleico. Por ejemplo, se puede comenzar típicamente con 50-100 ng de ADN genómico. Sin embargo, se puede empezar con menos si se amplifica el ADN genómico (por ejemplo, usando PCR) antes de la etapa de hibridación, por ejemplo, hibridación en solución. Por tanto, es posible, pero no esencial, amplificar el ADN genómico antes de la hibridación, por ejemplo, hibridación en solución.

La muestra de ácido nucleico usada para generar la colección también puede incluir ARN o ADNc derivado de ARN. En algunas implementaciones, el ARN incluye ARN celular total. En otras implementaciones, se han disminuido determinadas secuencias de ARN abundantes (por ejemplo, ARN ribosómicos). En otras implementaciones, se ha enriquecido la fracción de ARNm con cola de poli(A) en la preparación de ARN total. En algunas implementaciones, el ADNc se produce por procedimientos de síntesis de ADNc con cebado aleatorio. En otras implementaciones, la síntesis de ADNc se inicia en la cola de poli(A) de los ARNm maduros mediante cebado con oligonucleótidos que contienen oligo(dT). Los procedimientos para la disminución, el enriquecimiento de poli(A) y la síntesis de ADNc son bien conocidos para los expertos en la técnica.

El procedimiento puede incluir además la amplificación de la muestra de ácido nucleico por procedimientos de amplificación de ácidos nucleicos específicos o no específicos que son conocidos para los expertos en la técnica. La muestra de ácido nucleico se puede amplificar, por ejemplo, por procedimientos de amplificación hologenómica, tales como amplificación por desplazamiento de hebra con cebado aleatorio.

La muestra de ácido nucleico se puede fragmentar o cortar por procedimientos físicos o enzimáticos, como se describe en el presente documento, y se liga a adaptadores sintéticos, se selecciona por tamaño (por ejemplo, por electroforesis en gel preparativa) y se amplifica (por ejemplo, por PCR). El grupo de ácidos nucleicos fragmentados y ligados a adaptador se usa sin selección de tamaño o amplificación explícitas antes de la selección de híbridos.

En una implementación, la muestra de ácido nucleico comprende ADN, ARN (o ADNc derivado de ARN) o ambos de una célula no cancerosa o una célula no maligna, por ejemplo, un linfocito infiltrante del tumor. En una implementación, la muestra de ácido nucleico comprende ADN, ARN (o ADNc derivado de ARN) o ambos de una célula no cancerosa o una célula no maligna, por ejemplo, un linfocito infiltrante del tumor, y no comprende, o está esencialmente libre de, ADN, ARN (o ADNc derivado de ARN) o ambos de una célula cancerosa o una célula maligna.

En una implementación, la muestra de ácido nucleico comprende ADN, ARN (o ADNc derivado de ARN) de una célula cancerosa o una célula maligna. En una implementación, la muestra de ácido nucleico comprende ADN, ARN (o ADNc derivado de ARN) de una célula cancerosa o una célula maligna y no comprende, o está esencialmente libre de, ADN, ARN (o ADNc derivado de ARN) o ambos de una célula no cancerosa o una célula no maligna, por ejemplo, un linfocito infiltrante del tumor.

En una implementación, la muestra de ácido nucleico comprende ADN, ARN (o ADNc derivado de ARN) o ambos de una célula no cancerosa o una célula no maligna, por ejemplo, un linfocito infiltrante del tumor, y ADN, ARN (o ADNc derivado de ARN) o ambos de una célula cancerosa o una célula maligna.

Diseño y construcción de cebos

Un cebo puede ser una molécula de ácido nucleico, por ejemplo, una molécula de ADN o ARN, que se puede hibridar (por ejemplo, ser complementaria a) y, de este modo, permitir la captura de un ácido nucleico diana. En determinadas implementaciones, el ácido nucleico diana es una molécula de ADN genómico. En otras implementaciones, el ácido nucleico diana es una molécula de ARN o una molécula de ADNc derivada de una

molécula de ARN. En una implementación, un cebo es una molécula de ARN. En otras implementaciones, un cebo incluye una entidad de unión, por ejemplo, una marca de afinidad, que permite la captura y separación, por ejemplo, uniéndose a una entidad de unión, de un híbrido formado por un cebo y un ácido nucleico hibridado al cebo. En una implementación, un cebo es adecuado para la hibridación en fase de solución.

Típicamente, se usan moléculas de ARN como secuencias de cebo. Un dúplex ARN-ADN es más estable que un dúplex ADN-ADN y, por lo tanto, proporciona una captura potencialmente mejor de los ácidos nucleicos.

Los cebos de ARN se pueden preparar como se describe en otra parte en el presente documento usando procedimientos conocidos en la técnica que incluyen, pero no se limitan a, síntesis química *de novo* y transcripción de moléculas de ADN usando una ARN polimerasa dependiente de ADN. En una implementación, la secuencia de cebo se produce usando procedimientos de amplificación de ácidos nucleicos conocidos, tales como PCR, por ejemplo, usando ADN humano o muestras de ADN humano agrupadas como molde. A continuación, los oligonucleótidos se pueden convertir en cebos de ARN. En una implementación, se usa la transcripción *in vitro*, por ejemplo, basada en la adición de una secuencia de promotor de ARN polimerasa a un extremo del oligonucleótido. En una implementación, la secuencia de promotor de ARN polimerasa se añade al final del cebo amplificando o reamplificando la secuencia del cebo, por ejemplo, usando PCR u otros procedimientos de amplificación de ácidos nucleicos, por ejemplo, añadiendo una cola de un cebador de cada par de cebadores específicos de la diana con una secuencia promotora de ARN. En una implementación, la ARN polimerasa es una polimerasa T7, una polimerasa SP6 o una polimerasa T3. En una implementación, el cebo de ARN se marca con una marca, por ejemplo, una marca de afinidad. En una implementación, el cebo de ARN se prepara mediante transcripción *in vitro*, por ejemplo, usando UTP biotinilado. En otra implementación, el cebo de ARN se produce sin biotina y, a continuación, se reticula la biotina a la molécula de ARN usando procedimientos bien conocidos en la técnica, tales como reticulación con psoraleno. En una implementación, el cebo de ARN es una molécula de ARN resistente a RNasa, que se puede preparar, por ejemplo, usando nucleótidos modificados durante la transcripción para producir una molécula de ARN que resista la degradación por RNasa. En una implementación, el cebo de ARN corresponde a solo una cadena de la diana de ADN bicatenario. Típicamente, dichos cebos de ARN no son autocomplementarios y son más eficaces como impulsores de hibridación.

Los conjuntos de cebos se pueden diseñar a partir de secuencias de referencia, de modo que los cebos sean óptimos para seleccionar dianas de las secuencias de referencia. En algunas implementaciones, las secuencias de cebo se diseñan usando una base mixta (por ejemplo, degeneración). Por ejemplo, la(s) base(s) mixta(s) se puede(n) incluir en la secuencia de cebo en la(s) posición(es) de un SNP o mutación común para optimizar las secuencias de cebo para atrapar ambos alelos (por ejemplo, SNP y no SNP; mutante y no mutante). En algunas implementaciones, todas las variaciones de secuencia conocidas (o un subconjunto de las mismas) se pueden dirigir con múltiples cebos de oligonucleótidos, en lugar de usar oligonucleótidos degenerados mixtos.

En determinadas implementaciones, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 100 nucleótidos y 300 nucleótidos de longitud. Típicamente, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 130 nucleótidos y 230 nucleótidos, o aproximadamente 150 y 200 nucleótidos, de longitud. En otras implementaciones, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 300 nucleótidos y 1000 nucleótidos de longitud.

En algunas implementaciones, las secuencias específicas del miembro diana en el oligonucleótido son de entre aproximadamente 40 y 1000 nucleótidos, aproximadamente 70 y 300 nucleótidos, aproximadamente 100 y 200 nucleótidos de longitud, típicamente de entre aproximadamente 120 y 170 nucleótidos de longitud.

En algunas implementaciones, el conjunto de cebos incluye una entidad de unión. La entidad de unión puede ser una marca de afinidad en cada secuencia de cebo. En algunas implementaciones, la marca de afinidad es una molécula de biotina o un hapteno. En determinadas implementaciones, la entidad de unión permite la separación de los híbridos cebo/miembro de una mezcla de hibridación uniéndose a un compañero de unión, tal como una molécula de avidina, o un anticuerpo que se une al hapteno o a un fragmento de unión a antígeno del mismo.

En otras implementaciones, los oligonucleótidos en el conjunto de cebos contienen secuencias complementadas directa e inversa para la secuencia de misma miembro diana, con lo que los oligonucleótidos con secuencias específicas de miembro complementario inverso también llevan colas universales complementarias inversas. Esto puede dar lugar a transcritos de ARN que son la misma cadena, es decir, no son complementarios entre sí.

En otras implementaciones, el conjunto de cebos incluye oligonucleótidos que contienen bases degeneradas o mixtas en una o más posiciones. Todavía en otras implementaciones, el conjunto de cebos incluye múltiples o sustancialmente todas las variantes de secuencia conocidas presentes en una población de una única especie o comunidad de organismos. En una implementación, el conjunto de cebos incluye múltiples o sustancialmente todas las variantes de secuencia conocidas presentes en una población humana.

En otras implementaciones, el conjunto de cebos incluye secuencias de ADNc o se deriva de secuencias de ADNc.

En otras implementaciones, el conjunto de cebos incluye productos de amplificación (por ejemplo, productos de PCR) que se amplifican a partir de ADN genómico, ADNc o ADN clonado.

En otras implementaciones, el conjunto de cebos incluye moléculas de ARN. En algunas implementaciones, el conjunto incluye moléculas de ARN química o enzimáticamente modificadas o transcritas *in vitro* que incluyen, pero no se limitan a, aquellas que son más estables y resistentes a RNasa.

Aún en otras implementaciones, los cebos se producen mediante procedimientos descritos en el documento US 2010/0029498 y Gnirke, A. *et al.* (2009) *Nat Biotechnol.* 27(2):182-189. Por ejemplo, se pueden producir cebos de ARN biotinilado obteniendo un grupo de oligonucleótidos largos sintéticos, sintetizados originalmente en una micromatriz, y amplificando los oligonucleótidos para producir las secuencias de cebo. En algunas implementaciones, los cebos se producen añadiendo una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de cebo y sintetizando secuencias de ARN usando ARN polimerasa. En una implementación, las colecciones de oligodesoxinucleótidos sintéticos se pueden obtener de proveedores comerciales, tales como Agilent Technologies, Inc., y amplificar usando procedimientos de amplificación de ácidos nucleicos conocidos.

En consecuencia, se proporciona un procedimiento para preparar el conjunto de cebos mencionado anteriormente. El procedimiento incluye seleccionar una o más secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, una o más secuencias de oligonucleótidos de captura, referencia o control de mutaciones como se describe en el presente documento); obtener un grupo de secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, sintetizar el grupo de secuencias de oligonucleótidos de cebo específicas de diana, por ejemplo, mediante síntesis en micromatrices); y, opcionalmente, amplificar los oligonucleótidos para producir el conjunto de cebos.

En otras implementaciones, los procedimientos incluyen además amplificar (por ejemplo, mediante PCR) los oligonucleótidos usando uno o más cebadores biotinilados. En algunas implementaciones, los oligonucleótidos incluyen una secuencia universal en el extremo de cada oligonucleótido unido a la micromatriz. Los procedimientos pueden incluir además retirar las secuencias universales de los oligonucleótidos. Dichos procedimientos también pueden incluir retirar la cadena complementaria de los oligonucleótidos, hibridar los oligonucleótidos y extender los oligonucleótidos. En algunas de estas implementaciones, los procedimientos para amplificar (por ejemplo, mediante PCR) los oligonucleótidos usan uno o más cebadores biotinilados. En algunas implementaciones, el procedimiento incluye además seleccionar por tamaño los oligonucleótidos amplificados.

En una implementación se prepara un conjunto de cebos de ARN. Los procedimientos incluyen producir un conjunto de secuencias de cebo de acuerdo con los procedimientos descritos en el presente documento, añadir una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de cebo y sintetizar secuencias de ARN usando ARN polimerasa. La ARN polimerasa se puede elegir de una ARN polimerasa T7, una ARN polimerasa SP6 o una ARN polimerasa T3. En otras implementaciones, la secuencia de promotor de ARN polimerasa se añade en los extremos de las secuencias de cebo amplificando (por ejemplo, mediante PCR) las secuencias de cebo. En implementaciones en las que las secuencias de cebo se amplifican mediante PCR con pares de cebadores específicos de ADN genómico o ADNc, la adición de una secuencia de promotor de ARN en el extremo 5' de uno de los dos cebadores específicos en cada par dará lugar a un producto de PCR que se puede transcribir en un cebo de ARN usando procedimientos estándar.

En otras implementaciones, se pueden producir conjuntos de cebos usando muestras de ADN humano o muestras de ADN humano agrupadas como molde. En dichas implementaciones, los oligonucleótidos se amplifican mediante reacción en cadena de la polimerasa (PCR). En otras implementaciones, los oligonucleótidos amplificados se reamplifican mediante amplificación por círculo rodante o amplificación por círculo rodante hiperramificado. También se pueden usar los mismos procedimientos para producir secuencias de cebo usando muestras de ADN humano o muestras de ADN humano agrupadas como molde. Los mismos procedimientos también se pueden usar para producir secuencias de cebo usando subfracciones de un genoma obtenido mediante otros procedimientos, que incluyen, pero no se limitan a, digestión restrictiva, electroforesis en gel de campo pulsado, separación por flujo, centrifugación en gradiente de densidad de CsCl, reasociación cinética selectiva, microdissección de preparaciones cromosómicas y otros procedimientos de fraccionamiento conocidos por los expertos en la técnica.

En determinadas implementaciones, el número de cebos en el conjunto de cebos es de menos de 1000. En otras implementaciones, el número de cebos en el conjunto de cebos es mayor que 1000, mayor que 5000, mayor que 10 000, mayor que 20 000, mayor que 50 000, mayor que 100 000 o mayor que 500 000.

La longitud de la secuencia de cebo puede ser de entre aproximadamente 70 nucleótidos y 1000 nucleótidos. En una implementación, la longitud del cebo es de entre aproximadamente 100 y 300 nucleótidos, 110 y 200 nucleótidos o 120 y 170 nucleótidos de longitud. Además de las mencionadas anteriormente, en los procedimientos descritos en el presente documento se pueden usar longitudes de oligonucleótidos intermedias de aproximadamente 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300,

400, 500, 600, 700, 800 y 900 nucleótidos de longitud. En algunas implementaciones, se pueden usar oligonucleótidos de aproximadamente 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220 o 230 bases.

Cada secuencia de cebo puede incluir una secuencia de cebo específica de diana (por ejemplo, una específica de miembro) y colas universales en uno o ambos extremos. Como se usa en el presente documento, el término "secuencia de cebo" se puede referir a la secuencia de cebo específica de diana o a todo el oligonucleótido, incluyendo la "secuencia de cebo" específica de diana y otros nucleótidos del oligonucleótido. Las secuencias específicas de diana de los cebos son de entre aproximadamente 40 nucleótidos y 1000 nucleótidos de longitud. En una implementación, la secuencia específica de diana es de entre aproximadamente 70 nucleótidos y 300 nucleótidos de longitud. En otra implementación, la secuencia específica de diana es de entre aproximadamente 100 nucleótidos y 200 nucleótidos de longitud. Aún en otra implementación, la secuencia específica de diana es de entre aproximadamente 120 nucleótidos y 170 nucleótidos de longitud, típicamente de 120 nucleótidos de longitud. En los procedimientos descritos en el presente documento también se pueden usar longitudes intermedias además de las mencionadas anteriormente, tales como secuencias específicas de diana de aproximadamente 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 400, 500, 600, 700, 800 y 900 nucleótidos de longitud, así como secuencias específicas de diana de longitudes entre las longitudes mencionadas anteriormente.

En una implementación, el cebo es un oligómero (por ejemplo, compuesto de oligómeros de ARN, oligómeros de ADN o una combinación de los mismos) de aproximadamente 50 a 200 nucleótidos de longitud (por ejemplo, de aproximadamente 50, 60, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 190 o 200 nucleótidos de longitud). En una implementación, cada oligómero de cebo incluye aproximadamente 120 a 170, o típicamente, aproximadamente 120 nucleótidos, que son una secuencia de cebo específica de diana. El cebo puede comprender secuencias de nucleótidos no específicas de diana adicionales en uno o ambos extremos. Las secuencias de nucleótidos adicionales se pueden usar, por ejemplo, para amplificación PCT o como un identificador de cebo. En determinadas implementaciones, el cebo comprende adicionalmente una entidad de unión como se describe en el presente documento (por ejemplo, una marca de captura tal como una molécula de biotina). La entidad de unión, por ejemplo, la molécula de biotina, se puede unir al cebo, por ejemplo, en el extremo 5', 3', o internamente (por ejemplo, mediante la incorporación de un nucleótido biotinilado), del cebo. En una implementación, la molécula de biotina se une al extremo 5' del cebo.

En una implementación ejemplar, el cebo es un oligonucleótido de aproximadamente 150 nucleótidos de longitud, de los que 120 nucleótidos son una "secuencia de cebo" específica de diana. Los otros 30 nucleótidos (por ejemplo, 15 nucleótidos en cada extremo) son colas arbitrarias universales usadas para la amplificación por PCR. Las colas pueden ser cualquier secuencia seleccionada por el usuario. Por ejemplo, el grupo de oligonucleótidos sintéticos puede incluir oligonucleótidos de la secuencia de 5'-ATCGCACCAGCGTGTG₁₂₀CACTGCGGCTCCTCA-3' (SEQ ID NO: 1) con N₁₂₀ indicando las secuencias de cebo específicas de diana.

Las secuencias de cebo descritas en el presente documento se pueden usar para la selección de exones y secuencias diana cortas. En una implementación, el cebo es de entre aproximadamente 100 nucleótidos y 300 nucleótidos de longitud. En otra implementación, el cebo es de entre aproximadamente 130 nucleótidos y 230 nucleótidos de longitud. Aún en otra implementación, el cebo es de entre aproximadamente 150 nucleótidos y 200 nucleótidos de longitud. Las secuencias específicas de diana en los cebos, por ejemplo, para la selección de exones y secuencias diana cortas, son de entre aproximadamente 40 nucleótidos y 1000 nucleótidos de longitud. En una implementación, la secuencia específica de diana es de entre aproximadamente 70 nucleótidos y 300 nucleótidos de longitud. En otra implementación, la secuencia específica de diana es de entre aproximadamente 100 nucleótidos y 200 nucleótidos de longitud. Aún en otra implementación, la secuencia específica de diana es de entre aproximadamente 120 nucleótidos y 170 nucleótidos de longitud.

En algunas implementaciones, los oligonucleótidos largos pueden minimizar la cantidad de oligonucleótidos necesarios para capturar las secuencias diana. Por ejemplo, se puede usar un oligonucleótido por exón. Es conocido en la técnica que la longitud media y la mediana de longitud de los exones codificantes de proteínas en el genoma humano son de aproximadamente 164 y 120 pares de bases, respectivamente. Los cebos más largos pueden ser más específicos y capturar mejor que los más cortos. Como resultado, la tasa de éxito por secuencia de cebo oligonucleotídico es mayor que con oligonucleótidos cortos. En una implementación, la secuencia mínima cubierta por el cebo es del tamaño de un cebo (por ejemplo, 120-170 bases), por ejemplo, para capturar dianas del tamaño de un exón. Al determinar la longitud de las secuencias de cebo, también se puede tomar en consideración que los cebos innecesariamente largos atrapan más ADN no deseado directamente contiguo a la diana. Los cebos oligonucleotídicos más largos también pueden ser más tolerantes a los polimorfismos en la región diana en las muestras de ADN que los más cortos. Típicamente, las secuencias de cebo se derivan de una secuencia de genoma de referencia. Si la secuencia diana en la muestra de ADN real se desvía de la secuencia de referencia, por ejemplo, si contiene un polimorfismo mononucleotídico (SNP), puede hibridar menos eficazmente con el cebo y, por lo tanto, puede estar subrepresentada o completamente ausente en las secuencias hibridadas con las secuencias de cebo. Las pérdidas alélicas debidas a los SNP pueden ser menos probables con las

moléculas de cebos sintéticos más largas por el motivo de que un único apareamiento incorrecto en, por ejemplo, 120 a 170 bases puede tener menos efecto sobre la estabilidad del híbrido que un único apareamiento incorrecto en 20 o 70 bases, que son las longitudes típicas de los cebos o cebadores en la amplificación múltiple y la captura en micromatrices, respectivamente.

Para la selección de dianas que son largas en comparación con la longitud de los cebos de captura, tales como regiones genómicas, las longitudes de secuencia de cebo están típicamente en el mismo intervalo de tamaños que los cebos para dianas cortas mencionados anteriormente, excepto que no exista la necesidad de limitar el tamaño máximo de las secuencias de cebo con el único propósito de minimizar el reconocimiento de secuencias contiguas. De forma alternativa, se pueden titular oligonucleótidos en una ventana mucho más amplia (típicamente 600 bases). Este procedimiento se puede usar para capturar fragmentos de ADN que sean mucho más grandes (por ejemplo, aproximadamente 500 bases) que un exón típico. Como resultado, se seleccionan muchas más secuencias no diana flanqueantes no deseadas.

Síntesis de cebos

Los cebos pueden ser cualquier tipo de oligonucleótido, por ejemplo, ADN o ARN. Los cebos de ADN o ARN ("oligo-cebos") se pueden sintetizar individualmente o se pueden sintetizar en una matriz como un conjunto de cebos de ADN o ARN ("cebos de matriz"). Un cebo oligo, ya sea proporcionado en un formato de matriz o como un oligo aislado, es típicamente monocatenario. El cebo puede comprender adicionalmente una entidad de unión como se describe en el presente documento (por ejemplo, una marca de captura tal como una molécula de biotina). La entidad de unión, por ejemplo, la molécula de biotina, se puede unir al cebo, por ejemplo, en el extremo 5' o 3' del cebo, típicamente, en el extremo 5' del cebo. Los conjuntos de cebos se pueden sintetizar mediante procedimientos descritos en la técnica, por ejemplo, como se describe en la publicación de solicitud de patente internacional n.º WO 2012/092426.

Condiciones de hibridación

Los procedimientos presentados en el presente documento incluyen la etapa de poner en contacto la colección (por ejemplo, la colección de ácidos nucleicos) con una pluralidad de cebos para proporcionar una captura de la colección seleccionada. La etapa de contacto se puede efectuar en la hibridación en solución. En determinadas implementaciones, el procedimiento incluye repetir la etapa de hibridación en uno o más ciclos adicionales de hibridación en solución. En algunas implementaciones, los procedimientos incluyen además someter la captura de la colección a uno o más ciclos adicionales de hibridación en solución con el mismo o diferente conjunto de cebos. Los procedimientos de hibridación que se pueden adaptar para su uso en los procedimientos en el presente documento se describen en la técnica, por ejemplo, como se describe en la publicación de solicitud de patente internacional n.º WO 2012/092426.

Las implementaciones o rasgos característicos adicionales de la presente divulgación son como sigue:

La divulgación también presenta un procedimiento para preparar los conjuntos de cebos mencionados anteriormente. El procedimiento incluye seleccionar una o más secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, cualquiera de las secuencias de cebo correspondientes a los intervalos de sujeto (por ejemplo, intervalos subgenómicos, intervalos subgenómicos expresados o ambos) de los genes o productos génicos como se describe en el presente documento); obtener un grupo de secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, sintetizar el grupo de secuencias de oligonucleótidos de cebo específicas de diana, por ejemplo, mediante síntesis en micromatrices); y, opcionalmente, amplificar los oligonucleótidos para producir los conjuntos de cebos.

La divulgación también presenta un procedimiento para determinar la presencia o ausencia de una alteración asociada, por ejemplo, positiva o negativamente, con un fenotipo canceroso (por ejemplo, al menos 10, 20, 30, 50 o más de las alteraciones en los genes o productos génicos descritos en el presente documento) en una muestra de ácido nucleico. El procedimiento incluye poner en contacto los ácidos nucleicos en la muestra con la selección basada en solución de acuerdo con cualquiera de los procedimientos y cebos descritos en el presente documento para obtener una captura de ácidos nucleicos; y secuenciar (por ejemplo, mediante secuenciación de nueva generación) toda o un subconjunto de la captura de ácidos nucleicos, determinando de este modo la presencia o ausencia de la alteración en los genes o productos génicos descritos en el presente documento).

En determinadas implementaciones, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 100 nucleótidos y 300 nucleótidos de longitud. Típicamente, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 130 nucleótidos y 230 nucleótidos, o aproximadamente 150 y 200 nucleótidos, de longitud. En otras implementaciones, el conjunto de cebos incluye un oligonucleótido (o una pluralidad de oligonucleótidos) de entre aproximadamente 300 nucleótidos y 1000 nucleótidos de longitud.

En algunas implementaciones, las secuencias específicas del miembro diana en el oligonucleótido son de entre

aproximadamente 40 y 1000 nucleótidos, aproximadamente 70 y 300 nucleótidos, aproximadamente 100 y 200 nucleótidos de longitud, típicamente de entre aproximadamente 120 y 170 nucleótidos de longitud.

- 5 En algunas implementaciones, el conjunto de cebos incluye una entidad de unión. La entidad de unión puede ser una marca de afinidad en cada secuencia de cebo. En algunas implementaciones, la marca de afinidad es una molécula de biotina o un hapteno. En determinadas implementaciones, la entidad de unión permite la separación de los híbridos cebo/miembro de una mezcla de hibridación uniéndose a un compañero de unión, tal como una molécula de avidina, o un anticuerpo que se une al hapteno o a un fragmento de unión a antígeno del mismo.
- 10 En otras implementaciones, los oligonucleótidos en el conjunto de cebos contienen secuencias complementadas directa e inversa para la secuencia de misma miembro diana, con lo que los oligonucleótidos con secuencias específicas de miembro complementario inverso también llevan colas universales complementarias inversas. Esto puede dar lugar a transcritos de ARN que son la misma cadena, es decir, no son complementarios entre sí.
- 15 En otras implementaciones, el conjunto de cebos incluye oligonucleótidos que contienen bases degeneradas o mixtas en una o más posiciones. Todavía en otras implementaciones, el conjunto de cebos incluye múltiples o sustancialmente todas las variantes de secuencia conocidas presentes en una población de una única especie o comunidad de organismos. En una implementación, el conjunto de cebos incluye múltiples o sustancialmente todas las variantes de secuencia conocidas presentes en una población humana.
- 20 En otras implementaciones, el conjunto de cebos incluye secuencias de ADNc o se deriva de secuencias de ADNc. En una implementación, el ADNc se prepara a partir de una secuencia de ARN, por ejemplo, un ARN derivado de células tumorales o cancerosas, por ejemplo, un ARN obtenido de una muestra tumoral FFIP, una muestra de sangre o una muestra de aspirado de médula ósea. En otras implementaciones, el conjunto de cebos incluye
- 25 productos de amplificación (por ejemplo, productos de PCR) que se amplifican a partir de ADN genómico, ADNc o ADN clonado.
- En otras implementaciones, el conjunto de cebos incluye moléculas de ARN. En algunas implementaciones, el conjunto incluye moléculas de ARN química o enzimáticamente modificadas o transcritas *in vitro* que incluyen, pero
- 30 no se limitan a, aquellas que son más estables y resistentes a RNasa.
- Aún en otras implementaciones, los cebos se producen mediante procedimientos descritos en el documento US 2010/0029498 y Gnirke, A. *et al.* (2009) *Nat Biotechnol.* 27(2): 182-189. Por ejemplo, se pueden producir cebos de ARN biotinilado obteniendo un grupo de oligonucleótidos largos sintéticos, sintetizados originalmente en una
- 35 micromatriz, y amplificando los oligonucleótidos para producir las secuencias de cebo. En algunas implementaciones, los cebos se producen añadiendo una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de cebo y sintetizando secuencias de ARN usando ARN polimerasa. En una implementación, las colecciones de oligodesoxinucleótidos sintéticos se pueden obtener de proveedores comerciales, tales como Agilent Technologies, Inc., y amplificar usando procedimientos de amplificación de ácidos nucleicos conocidos.
- 40 En consecuencia, se proporciona un procedimiento para preparar el conjunto de cebos mencionado anteriormente. El procedimiento incluye seleccionar una o más secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, una o más secuencias de oligonucleótidos de captura, referencia o control de mutaciones como se describe en el presente documento); obtener un grupo de secuencias de oligonucleótidos de cebo específicas de diana (por ejemplo, sintetizar el grupo de secuencias de oligonucleótidos de cebo específicas de diana, por ejemplo, mediante síntesis en micromatrices); y, opcionalmente, amplificar los oligonucleótidos para producir el conjunto de cebos.
- 45 En otras implementaciones, los procedimientos incluyen además amplificar (por ejemplo, mediante PCR) los oligonucleótidos usando uno o más cebadores biotinilados. En algunas implementaciones, los oligonucleótidos incluyen una secuencia universal en el extremo de cada oligonucleótido unido a la micromatriz. Los procedimientos pueden incluir además retirar las secuencias universales de los oligonucleótidos. Dichos procedimientos también pueden incluir retirar la cadena complementaria de los oligonucleótidos, hibridar los oligonucleótidos y extender los
- 50 oligonucleótidos. En algunas de estas implementaciones, los procedimientos para amplificar (por ejemplo, mediante PCR) los oligonucleótidos usan uno o más cebadores biotinilados. En algunas implementaciones, el procedimiento incluye además seleccionar por tamaño los oligonucleótidos amplificados.
- 55 En una implementación se prepara un conjunto de cebos de ARN. Los procedimientos incluyen producir un conjunto de secuencias de cebo de acuerdo con los procedimientos descritos en el presente documento, añadir una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de cebo y sintetizar secuencias de ARN usando ARN polimerasa. La ARN polimerasa se puede elegir de una ARN polimerasa T7, una ARN polimerasa SP6 o una ARN polimerasa T3. En otras implementaciones, la secuencia de promotor de ARN polimerasa se añade en los extremos de las secuencias de cebo amplificando (por ejemplo, mediante PCR) las
- 60 secuencias de cebo. En implementaciones en las que las secuencias de cebo se amplifican mediante PCR con pares de cebadores específicos de ADN genómico o ADNc, la adición de una secuencia de promotor de ARN en
- 65

el extremo 5' de uno de los dos cebadores específicos en cada par dará lugar a un producto de PCR que se puede transcribir en un cebo de ARN usando procedimientos estándar.

En otras implementaciones, se pueden producir conjuntos de cebos usando muestras de ADN humano o muestras de ADN humano agrupadas como molde. En dichas implementaciones, los oligonucleótidos se amplifican mediante reacción en cadena de la polimerasa (PCR). En otras implementaciones, los oligonucleótidos amplificados se reamplifican mediante amplificación por círculo rodante o amplificación por círculo rodante hiperramificado. También se pueden usar los mismos procedimientos para producir secuencias de cebo usando muestras de ADN humano o muestras de ADN humano agrupadas como molde. Los mismos procedimientos también se pueden usar para producir secuencias de cebo usando subfracciones de un genoma obtenido mediante otros procedimientos, que incluyen, pero no se limitan a, digestión restrictiva, electroforesis en gel de campo pulsado, separación por flujo, centrifugación en gradiente de densidad de CsCl, reasociación cinética selectiva, microdissección de preparaciones cromosómicas y otros procedimientos de fraccionamiento conocidos por los expertos en la técnica.

En determinadas implementaciones, el número de cebos en el conjunto de cebos es de menos de 1000, por ejemplo, 2, 3, 4, 5, 10, 50, 100, 500 cebos. En otras implementaciones, el número de cebos en el conjunto de cebos es mayor que 1000, mayor que 5000, mayor que 10 000, mayor que 20 000, mayor que 50 000, mayor que 100 000 o mayor que 500 000.

En determinadas implementaciones, una colección (por ejemplo, una colección de ácidos nucleicos) incluye un conjunto de miembros. Como se describe en el presente documento, los miembros de la colección pueden incluir un miembro diana (por ejemplo, un miembro tumoral, un miembro de referencia y/o un miembro de control; también denominado en el presente documento como un primer, segundo y/o tercer miembro, respectivamente). Los miembros de la colección pueden proceder de un único individuo. En implementaciones, una colección puede comprender miembros de más de un sujeto (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 o más sujetos), por ejemplo, se pueden combinar dos o más colecciones de diferentes sujetos a partir de una colección que tiene miembros de más de un sujeto. En una implementación, el sujeto es un ser humano que tiene, o está en riesgo de tener, un cáncer o tumor.

"Miembro" o "miembro de la colección" u otro término similar, como se usa en el presente documento, se refiere a una molécula de ácido nucleico, por ejemplo, ADN o ARN, que es un miembro de una colección. Típicamente, un miembro es una molécula de ADN, por ejemplo, ADN genómico o ADNc. Un miembro puede ser ADN genómico cortado. En otras implementaciones, el miembro puede ser un ADNc. En otras implementaciones, el miembro puede ser un ARN. Los miembros comprenden una secuencia de un sujeto y también pueden comprender una secuencia no derivada del sujeto, por ejemplo, cebadores o secuencias que permitan la identificación, por ejemplo, secuencias de "código de barras".

Aún en otra implementación, los procedimientos presentados en el presente documento incluyen además aislar una muestra de ácido nucleico para proporcionar una colección (por ejemplo, una colección de ácidos nucleicos como se describe en el presente documento). En determinadas implementaciones, la muestra de ácido nucleico incluye fragmentos completos genómicos, subgenómicos o ambos. Los protocolos para aislar y preparar colecciones a partir de fragmentos genómicos o subgenómicos completos son conocidos en la técnica (por ejemplo, el kit de preparación de muestras de ADN genómico de Illumina). En determinadas implementaciones, el fragmento de ADN genómico o subgenómico se aísla de una muestra de sujeto (por ejemplo, una muestra tumoral, un tejido adyacente normal (TCN), una muestra de sangre o cualquier control normal). En una implementación, la muestra (por ejemplo, la muestra tumoral o de TCN) es una conservada. Por ejemplo, la muestra se incluye en una matriz, por ejemplo, un bloque FFIP o una muestra congelada. En determinadas implementaciones, la etapa de aislamiento incluye la separación por flujo de cromosomas individuales; y/o microdissección de una muestra de sujeto (por ejemplo, una muestra tumoral, un TCN, una muestra de sangre). En determinadas implementaciones, la muestra de ácido nucleico usada para generar la colección de ácidos nucleicos es de menos de 5, menos de 1 microgramo o menos de 500 ng (por ejemplo, 200 ng o menos).

Todavía en otras implementaciones, la muestra de ácido nucleico usada para generar la colección incluye ARN o ADNc derivado de ARN. En algunas implementaciones, el ARN incluye ARN celular total. En otras implementaciones, se han disminuido determinadas secuencias de ARN abundantes (por ejemplo, ARN ribosómicos). En algunas implementaciones, se ha enriquecido la fracción de ARNm con cola de poli(A) en la preparación de ARN total. En algunas implementaciones, el ADNc se produce por procedimientos de síntesis de ADNc con cebado aleatorio. En otras implementaciones, la síntesis de ADNc se inicia en la cola de poli(A) de los ARNm maduros mediante cebado con oligonucleótidos que contienen oligo(dT). Los procedimientos para la disminución, el enriquecimiento de poli(A) y la síntesis de ADNc son bien conocidos para los expertos en la técnica.

El procedimiento puede incluir además la amplificación de la muestra de ácido nucleico por procedimientos de amplificación de ácidos nucleicos específicos o no específicos que son bien conocidos para los expertos en la técnica.

En algunas implementaciones, en determinadas implementaciones, la muestra de ácido nucleico se amplifica, por ejemplo, por procedimientos de amplificación hologenómica, tales como amplificación por desplazamiento de hebra con cebado aleatorio.

En otras implementaciones, la muestra de ácido nucleico se fragmenta o corta por procedimientos físicos o enzimáticos y se liga a adaptadores sintéticos, se selecciona por tamaño (por ejemplo, por electroforesis en gel preparativa) y se amplifica (por ejemplo, por PCR). En otras implementaciones, el grupo de ácidos nucleicos fragmentados y ligados a adaptador se usa sin selección de tamaño o amplificación explícitas antes de la selección de híbridos.

En otras implementaciones, el ADN aislado (por ejemplo, el ADN genómico) se fragmenta o corta. En algunas implementaciones, la colección incluye menos de un 50 % de ADN genómico, tal como una subfracción de ADN genómico que es una representación reducida o una porción definida de un genoma, por ejemplo, que se ha subfraccionado por otros medios. En otras implementaciones, la colección incluye todo o sustancialmente todo el ADN genómico.

En determinadas implementaciones, los miembros de la colección incluyen un intervalo subgenómico que incluye una región intragénica o una región intergénica. En otra implementación, el intervalo subgenómico incluye un exón o un intrón, o un fragmento de los mismos, típicamente una secuencia de exón o un fragmento de la misma. En una implementación, el intervalo subgenómico incluye una región codificante o una región no codificante, por ejemplo, un promotor, un potenciador, una región 5' no traducida (5'-UTR) o una región 3' no traducida (3'-UTR), o un fragmento de los mismos. En otras implementaciones, el intervalo subgenómico incluye un ADNc o un fragmento del mismo (por ejemplo, ADNc obtenido de un ARN tumoral (por ejemplo, ARN extraído de una muestra tumoral, por ejemplo, muestra tumoral FFIP). En otras implementaciones, el intervalo subgenómico incluye un SNP, por ejemplo, como se describe en el presente documento. En otras implementaciones, los miembros diana incluyen sustancialmente todos los exones en un genoma. En otras implementaciones, los miembros diana incluyen un intervalo subgenómico como se describe en el presente documento, por ejemplo, intervalos subgenómicos, por ejemplo, exones de genes o productos génicos seleccionados de interés (por ejemplo, genes o productos génicos asociados con un fenotipo canceroso como se describe en el presente documento).

En una implementación, el intervalo subgenómico incluye una mutación somática, una mutación de estirpe germinal o ambas. En una implementación, el intervalo subgenómico incluye una alteración, por ejemplo, una mutación puntual o una única, una mutación de delección (por ejemplo, una delección sin cambio de pauta de lectura, una delección intragénica, una delección génica completa), una mutación de inserción (por ejemplo, inserción intragénica), una mutación de inversión (por ejemplo, una inversión intracromosómica), una mutación de enlace, una mutación de inserción enlazada, una mutación de duplicación invertida, una duplicación en tándem (por ejemplo, una duplicación en tándem intracromosómica), una translocación (por ejemplo, una translocación cromosómica, una translocación no recíproca), un reordenamiento (por ejemplo, un reordenamiento genómico), un cambio en el número de copias del gen o una combinación de los mismos. En determinadas implementaciones, el intervalo subgenómico constituye menos de un 5, 1, 0,5, 0,1 %, 0,01 %, 0,001 % de la región codificante del genoma de las células tumorales en una muestra. En otras implementaciones, los intervalos subgenómicos no están involucrados en una enfermedad, por ejemplo, no se asocian con un fenotipo canceroso como se describe en el presente documento.

Los procedimientos en el presente documento incluyen la etapa de poner en contacto una o una pluralidad de colecciones (por ejemplo, una o una pluralidad de colecciones de ácidos nucleicos) con una pluralidad de cebos para proporcionar un subgrupo seleccionado de ácidos nucleicos, por ejemplo, una captura de la colección. En una implementación, la etapa de poner en contacto se efectúa en un soporte sólido, por ejemplo, una matriz. Los soportes sólidos adecuados para la hibridación se describen, por ejemplo, en, Albert, T.J. *et al.* (2007) *Nat. Methods* 4(11):903-5; Hodges, E. *et al.* (2002) *Nat. Genet.* 39(12):1522-7; Okou, D.T. *et al.* (2007) *Nat. Methods* 4(11):907-9. En otras implementaciones, la etapa de poner en contacto se efectúa en la hibridación en solución. En determinadas implementaciones, el procedimiento incluye repetir la etapa de hibridación en uno o más ciclos adicionales de hibridación. En algunas implementaciones, los procedimientos incluyen además someter la captura de la colección a uno o más ciclos adicionales de hibridación con el mismo o diferente conjunto de cebos.

En otras implementaciones, los procedimientos presentados en el presente documento incluyen además amplificar la captura de la colección (por ejemplo, mediante PCR). En otras implementaciones, la captura de la colección no se amplifica.

Aún en otras implementaciones, los procedimientos incluyen además analizar la captura de la colección. En una implementación, la captura de la colección se analiza mediante un procedimiento de secuenciación, por ejemplo, un procedimiento de secuenciación de nueva generación como se describe en el presente documento. Los procedimientos incluyen aislar una captura de la colección mediante hibridación en solución y someter la captura de la colección a secuenciación de ácidos nucleicos. En determinadas implementaciones, la captura de la colección se puede volver a secuenciar. Son conocidos procedimientos de secuenciación de nueva generación en la técnica y se describen, por ejemplo, en Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46.

Aún en otras implementaciones, los procedimientos incluyen además la etapa de someter la captura de la colección a genotipado, identificando de este modo el genotipo de los ácidos nucleicos seleccionados.

5 En determinadas implementaciones, el procedimiento incluye además uno o más de:

i) realizar una identificación genética de la muestra de ácido nucleico;

10 ii) cuantificar la abundancia de un gen o producto génico (por ejemplo, un gen o producto génico como se describe en el presente documento) en la muestra de ácido nucleico (por ejemplo, cuantificar la abundancia relativa de un transcrito en la muestra);

15 iii) identificar la muestra de ácido nucleico como perteneciente a un sujeto particular (por ejemplo, un control normal o un paciente con cáncer);

iv) identificar un rasgo genético en la muestra de ácido nucleico (por ejemplo, composición genética de uno o más sujetos (por ejemplo, etnia, raza, rasgos familiares));

20 v) determinar la ploidía en la muestra de ácido nucleico; determinar una pérdida de heterocigosidad en la muestra de ácido nucleico;

vi) determinar la presencia o ausencia de un acontecimiento de duplicación génica en la muestra de ácido nucleico;

25 vii) determinar la presencia o ausencia de un acontecimiento de amplificación génica en la muestra de ácido nucleico; o

viii) determinar el nivel de mezcla de células tumorales/normales en la muestra de ácido nucleico.

30 Cualquiera de los procedimientos descritos en el presente documento se puede combinar con una o más de las implementaciones a continuación.

35 En una implementación, el procedimiento comprende adquirir una lectura de secuencia de nucleótidos obtenida de una muestra de ácido nucleico tumoral y/o de control (por ejemplo, una muestra de ácido nucleico derivada de FFIP o una muestra de ácido nucleico derivada de una muestra de sangre o una muestra de aspirado de médula ósea).

40 En una implementación, las lecturas son proporcionadas por un procedimiento de secuenciación de nueva generación.

45 En una implementación, el procedimiento incluye proporcionar una colección de miembros de ácido nucleico y secuenciar un intervalo subgenómico preseleccionado de una pluralidad de miembros de dicha colección. En implementaciones, el procedimiento puede incluir una etapa de selección de un subconjunto de dicha colección para secuenciación, por ejemplo, una selección basada en solución.

50 En determinadas implementaciones, un procedimiento comprende procedimientos de captura de híbridos que se diseñan para capturar dos o más categorías diana diferentes, cada una con estrategias de diseño de cebos diferentes. Los procedimientos y composiciones de captura de híbridos pretenden capturar un subconjunto definido de secuencias diana (por ejemplo, miembros diana) y proporcionar una cobertura homogénea de la secuencia diana, al tiempo que minimizan la cobertura fuera de ese subconjunto. En una implementación, las secuencias diana incluyen el exoma completo del ADN genómico, o un subconjunto seleccionado del mismo. Los procedimientos y composiciones divulgados en el presente documento proporcionan diferentes conjuntos de cebos para lograr diferentes profundidades y patrones de cobertura para secuencias de ácidos nucleicos diana complejas (por ejemplo, colecciones).

55 En una determinada implementación, las diferentes categorías de conjuntos de cebos y dianas son como sigue.

60 A. Un primer conjunto de cebos que selecciona una diana de alto nivel (por ejemplo, uno o más miembros tumorales y/o miembros de referencia, tales como genes, exones o bases) para la que se requiere la cobertura más profunda para permitir un alto nivel de sensibilidad a las mutaciones que aparecen con baja frecuencia. Por ejemplo, la detección de mutaciones puntuales que aparecen con una frecuencia de aproximadamente un 5 % o menos (es decir, un 5 % de las células a partir de las cuales se preparó la muestra albergan esta mutación en su genoma). Típicamente, el primer conjunto de cebos requiere una profundidad de secuenciación de aproximadamente 500X o mayor para garantizar una alta fiabilidad de detección. En una implementación, el primer conjunto de cebos selecciona uno o más intervalos subgenómicos (por ejemplo, exones) que mutan con frecuencia en determinados tipos de cáncer, por ejemplo, un gen o producto génico de acuerdo con las **tablas 1-4**.

B. Un segundo conjunto de cebos que selecciona una diana de nivel intermedio (por ejemplo, uno o más miembros tumorales y/o miembros de referencia, tales como genes, exones o bases) para la que se requiere una alta cobertura para permitir un alto nivel de sensibilidad a las mutaciones que aparecen con una frecuencia mayor que la diana de nivel alto, por ejemplo, una frecuencia de aproximadamente un 10 %. Por ejemplo, la detección de una alteración (por ejemplo, una mutación puntual) que aparece con una frecuencia de un 10 % requiere una profundidad de secuenciación de aproximadamente 200X o mayor para garantizar una alta fiabilidad de detección. En una implementación, el segundo conjunto de cebos selecciona uno o más intervalos subgenómicos (por ejemplo, exones) que se eligen de los genes o productos génicos de acuerdo con las **tablas 1-4**.

C. Un tercer conjunto de cebos que selecciona una diana de bajo nivel (por ejemplo, uno o más miembros PGx, tales como genes, exones o bases) para la que se requiere una cobertura baja-media para permitir un alto nivel de sensibilidad, por ejemplo, para detectar alelos heterocigóticos. Por ejemplo, la detección de alelos heterocigóticos requiere una profundidad de secuenciación de 10-100X para garantizar una alta fiabilidad de detección. En una implementación, el tercer conjunto de cebos selecciona uno o más intervalos subgenómicos (por ejemplo, exones) que se eligen de: a) SNP farmacogenómicos que pueden explicar la capacidad del paciente para metabolizar diferentes fármacos, b) SNP genómicos que se pueden usar para identificar de forma única (huella) a un paciente, c) SNP/locus genómicos que se pueden usar para evaluar las ganancias/pérdidas en el número de copias del ADN genómico y la pérdida de heterocigosidad (PDH).

D. Un cuarto conjunto de cebos que selecciona una diana de intrón (por ejemplo, un miembro de intrón) para la que se requiere una cobertura baja-media para detectar puntos de ruptura estructurales tales como translocaciones genómicas o indels. Por ejemplo, la detección de un punto de ruptura intrónico requiere una profundidad que abarque 5-50X pares de secuencias para garantizar una alta fiabilidad de detección. Dichos cuartos conjuntos de cebos se pueden usar para detectar, por ejemplo, genes de cáncer propensos a translocación/indel.

E. Un quinto conjunto de cebos que selecciona una diana de intrón (por ejemplo, un miembro de intrón) para la que se requiere una escasa cobertura para mejorar la capacidad de detectar cambios en el número de copias. Por ejemplo, la detección de la delección de 1 copia de varios exones terminales requiere una cobertura de 0,1-10X para garantizar una alta fiabilidad de detección. Dichos quintos conjuntos de cebos se pueden usar para detectar, por ejemplo, genes de cáncer propensos a amplificación/delección.

Los procedimientos y composiciones presentados en el presente documento implican ajustar la cobertura relativa de secuencia de cada conjunto de cebos/categoría diana. Los procedimientos para implementar diferencias en la cobertura relativa de secuencia en el diseño de cebos incluyen uno o más de:

(i) Representación diferencial de diferentes conjuntos de cebos: el diseño del conjunto de cebos para capturar una diana dada (por ejemplo, un miembro diana) se puede incluir en un mayor/menor número de copias para potenciar/reducir las profundidades de coberturas relativas de la diana;

(ii) Solapamiento diferencial de subconjuntos de cebos: el diseño del conjunto de cebos para capturar una diana dada (por ejemplo, un miembro diana) puede incluir un solapamiento más largo o más corto entre cebos vecinos para potenciar/reducir las profundidades de coberturas relativas de la diana;

(iii) Parámetros diferenciales del cebo: el diseño del conjunto de cebos para capturar una diana dada (por ejemplo, un miembro diana) puede incluir modificaciones de secuencia/longitud más corta para reducir la eficacia de captura y disminuir las profundidades de coberturas relativas de la diana;

(iv) Mezcla de diferentes conjuntos de cebos: los conjuntos de cebos que se diseñan para capturar los conjuntos de dianas se pueden mezclar en diferentes proporciones molares para potenciar/reducir las profundidades de coberturas relativas de la diana;

(v) Uso de diferentes tipos de conjuntos de cebos oligonucleotídicos: en determinadas implementaciones, el conjunto de cebos puede incluir:

(a) uno o más cebos sintetizados químicamente (por ejemplo, no enzimáticamente) (por ejemplo, sintetizados individualmente),

(b) uno o más cebos sintetizados en una matriz,

(c) uno o más cebos preparados enzimáticamente, por ejemplo, transcritos *in vitro*;

(d) cualquier combinación de (a), (b) y/o (c),

(e) uno o más oligonucleótidos de ADN (por ejemplo, un oligonucleótido de ADN natural o no natural),

- (f) uno o más oligonucleótidos de ARN (por ejemplo, un oligonucleótido de ARN natural o no natural),
- (g) una combinación de (e) y (f), o
- 5 (h) una combinación de cualquiera de los anteriores.

Las diferentes combinaciones de oligonucleótidos se pueden mezclar en diferentes proporciones, por ejemplo, una proporción elegida de 1: 1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:50; 1:100, 1:1000 o similares. En una implementación, la proporción entre cebo sintetizado químicamente y cebo generado por matriz se elige de 1:5, 1:10 o 1:20. Los oligonucleótidos de ADN o ARN pueden ser naturales o no naturales. En determinadas implementaciones, los cebos incluyen uno o más nucleótidos no naturales para, por ejemplo, incrementar la temperatura de fusión. Los oligonucleótidos no naturales ejemplares incluyen nucleótidos de ADN o ARN modificados. Un nucleótido de ARN modificado ejemplar es un ácido nucleico bloqueado (ANB), en el que el resto de ribosa de un nucleótido de ANB se modifica con un puente adicional que conecta el oxígeno 2' y el carbono 4' (Kaur, H; Arora, A; Wengel, J; Maiti, S; Arora, A.; Wengel, J.; Maiti, S. (2006). "Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes". *Biochemistry* 45 (23): 7347-55). Otros nucleótidos de ADN y ARN ejemplares modificados incluyen, pero no se limitan a, ácido nucleico peptídico (ANP) compuesto por unidades de N-(2-aminoetil)-glicina repetidas unidas por enlaces peptídicos (Egholm, M. *et al.* (1993) *Nature* 365 (6446): 566-8); un oligonucleótido de ADN o ARN modificado para capturar regiones de bajo contenido en GC; un ácido nucleico bicíclico (ANB) o un oligonucleótido reticulado; una 5-metil desoxicitidina modificada; y 2,6-diaminopurina. Se conocen en la técnica otros nucleótidos de ADN y ARN modificados.

En determinadas implementaciones se obtiene una cobertura sustancialmente uniforme u homogénea de una secuencia diana (por ejemplo, un miembro diana). Por ejemplo, dentro de cada conjunto de cebos/categoría diana, se puede optimizar la uniformidad de la cobertura modificando los parámetros de cebo, por ejemplo, mediante uno o más de:

(i) se puede usar el incremento/disminución de la representación o solapamiento de los cebos para potenciar/reducir la cobertura de las dianas (por ejemplo, miembros diana) que están insuficientemente/excesivamente cubiertas en relación con otras dianas de la misma categoría;

(ii) para una cobertura baja, las secuencias diana difíciles de capturar (por ejemplo, secuencias con alto contenido en GC) expanden la región que es la diana de los conjuntos de cebos para cubrir, por ejemplo, secuencias contiguas (por ejemplo, secuencias contiguas menos ricas en GC);

(iii) se puede realizar una modificación de una secuencia de cebo para reducir la estructura secundaria del cebo y potenciar su eficacia de selección;

(iv) se puede usar la modificación de la longitud de un cebo para equilibrar la cinética de hibridación de fusión de diferentes cebos dentro de la misma categoría. La longitud del cebo se puede modificar directamente (produciendo cebos con longitudes variables) o indirectamente (produciendo cebos de longitud constante y reemplazando los extremos del cebo con una secuencia arbitraria);

(v) modificar cebos de diferente orientación para la misma región diana (es decir, cadena directa e inversa) puede tener diferentes eficacias de unión. Se puede seleccionar el conjunto de cebos con cualquier orientación que proporcione una cobertura óptima para cada diana;

(vi) modificar la cantidad de una entidad de unión, por ejemplo, una marca de captura (por ejemplo, biotina), presente en cada cebo puede afectar a su eficacia de unión. Se puede usar un incremento/disminución del nivel de marca de los cebos dirigidos a una diana específica para potenciar/reducir la cobertura relativa de la diana;

(vii) se puede alterar la modificación del tipo de nucleótido usado para diferentes cebos para afectar la afinidad de unión a la diana y potenciar/reducir la cobertura relativa de la diana; o

(viii) se pueden usar cebos de oligonucleótidos modificados, por ejemplo, que tengan un emparejamiento de bases más estable, para equilibrar la cinética de hibridación de fusión entre áreas con contenido en GC bajo o normal en relación con las de alto contenido en GC.

Por ejemplo, se pueden usar diferentes tipos de conjuntos de cebos oligonucleotídicos.

En una implementación, el valor para la eficacia de selección se modifica mediante el uso de diferentes tipos de cebos oligonucleotídicos para abarcar regiones diana preseleccionadas. Por ejemplo, se puede usar un primer conjunto de cebos (por ejemplo, un conjunto de cebos basado en matriz que comprende 10 000-50 000 cebos de ARN o ADN) para cubrir una gran zona diana (por ejemplo, zona diana total de 1-2 MB). El primer conjunto de cebos se puede enriquecer con un segundo conjunto de cebos (por ejemplo, conjunto de cebos de ARN o ADN sintetizado individualmente que comprende menos de 5000 cebos) para cubrir una región diana preseleccionada

(por ejemplo, intervalos subgenómicos seleccionados de interés que abarcan, por ejemplo, 250 kb o menos, de una zona diana) y/o regiones con estructura secundaria superior, por ejemplo, mayor contenido de GC. Los intervalos subgenómicos de interés seleccionados pueden corresponder a uno o más de los genes o productos génicos descritos en el presente documento, o un fragmento de los mismos. El segundo conjunto de cebos puede incluir aproximadamente 2000-5000 cebos dependiendo del solapamiento de cebos deseado. Aún en otras implementaciones, el segundo conjunto de cebos puede incluir oligo-cebos seleccionados (por ejemplo, menos de 400, 200, 100, 50, 40, 30, 20, 10 cebos) añadidos al primer conjunto de cebos. El segundo conjunto de cebos se puede mezclar en cualquier proporción de oligo-cebos individuales. Por ejemplo, el segundo conjunto de cebos puede incluir cebos individuales presentes en una proporción equimolar 1:1. De forma alternativa, el segundo conjunto de cebos puede incluir cebos individuales presentes en diferentes proporciones (por ejemplo, 1:5, 1:10, 1:20), por ejemplo, para optimizar la captura de determinadas dianas (por ejemplo, determinadas dianas pueden tener 5-10 veces más del segundo cebo que de otras dianas).

Secuenciación

La presente divulgación también incluye procedimientos de secuenciación de ácidos nucleicos. En estos procedimientos, los miembros de la colección de ácidos nucleicos se aíslan usando los procedimientos descritos en el presente documento, por ejemplo, usando hibridación en solución, proporcionando de este modo una captura de la colección. La captura de la colección o un subgrupo de la misma se puede secuenciar. En consecuencia, los procedimientos presentados en el presente documento incluyen además analizar la captura de la colección. En una implementación, la captura de la colección se analiza mediante un procedimiento de secuenciación, por ejemplo, un procedimiento de secuenciación de nueva generación como se describe en el presente documento. Los procedimientos incluyen aislar una captura de la colección mediante hibridación en solución y someter la captura de la colección a secuenciación de ácidos nucleicos. En determinadas implementaciones, la captura de la colección se puede volver a secuenciar.

Se puede usar cualquier procedimiento de secuenciación conocido en la técnica. La secuenciación de ácidos nucleicos aislados mediante procedimientos de selección se lleva a cabo típicamente usando secuenciación de nueva generación (NGS). Los procedimientos de secuenciación adecuados para su uso en los procedimientos en el presente documento se describen en la técnica, por ejemplo, como se describe en la publicación de solicitud de patente internacional n.º WO 2012/092426.

Después de que se hayan generado las lecturas de NGS, se pueden alinear con una secuencia de referencia conocida o se pueden ensamblar *de novo*. Por ejemplo, se puede lograr la identificación de variaciones genéticas, tales como un polimorfismo mononucleotídico, y de variantes estructurales en una muestra (por ejemplo, una muestra tumoral) alineando las lecturas de NGS con una secuencia de referencia (por ejemplo, una secuencia natural). Se describen procedimientos de alineación de secuencias para NGS, por ejemplo, en Trapnell C. y Salzberg S.L. *Nature Biotech.*, 2009, 27:455-457. Se describen ejemplos de ensamblajes *de novo*, por ejemplo, en Warren R. *et al.*, *Bioinformatics*, 2007, 23:500-501; Butler J. *et al.*, *Genome Res.*, 2008, 18:810-820; y Zerbino D.R. y Birney E., *Genome Res.*, 2008, 18:821-829. La alineación o el ensamblaje de secuencias se puede realizar usando datos de lectura de una o más plataformas de NGS, por ejemplo, mezclando datos de lectura de Roche/454 y de Illumina/Solexa.

Alineación

La alineación es el procedimiento de emparejamiento de una lectura con una localización, por ejemplo, una localización genómica. La alineación incorrecta (por ejemplo, la colocación de pares de bases de una lectura corta en localizaciones incorrectas en el genoma), por ejemplo, la alineación incorrecta debido al contexto de la secuencia (por ejemplo, presencia de una secuencia repetitiva) de lecturas alrededor de una mutación de cáncer real puede dar lugar a una reducción de la sensibilidad de detección de mutaciones, ya que las lecturas del alelo alternativo pueden estar desplazadas fuera de la acumulación principal de lecturas de alelos alternativos. Si el contexto de secuencia problemático se produce cuando no hay una mutación real presente, la alineación incorrecta puede introducir lecturas artefactuales de alelos "mutados" al colocar lecturas reales de bases del genoma de referencia en la localización incorrecta. Debido a que los algoritmos de llamada de mutaciones para el análisis multigénico múltiple deben ser sensibles incluso a mutaciones de baja abundancia, estas alineaciones incorrectas pueden incrementar las tasas de descubrimiento de positivos falsos/reducir la especificidad.

Como se analiza en el presente documento, la sensibilidad reducida por las mutaciones reales se puede abordar evaluando la calidad de las alineaciones (manualmente o de forma automatizada) alrededor de los sitios de mutación esperados en los genes que se están analizando. Los sitios que se van a evaluar se pueden obtener de bases de datos de mutaciones de cáncer (por ejemplo, COSMIC). Las regiones que se identifican como problemáticas se pueden remediar con el uso de un algoritmo seleccionado para proporcionar un mejor rendimiento en el contexto de secuencia relevante, por ejemplo, mediante la optimización de la alineación (o realineación) usando algoritmos de alineación más lentos pero más exactos, tales como la alineación de Smith-Waterman. En los casos en que los algoritmos de alineación generales no pueden remediar el problema, se pueden crear enfoques de alineación personalizados mediante, por ejemplo: el ajuste de los parámetros de penalización por

emparejamiento erróneo de diferencia máxima para genes con una alta probabilidad de contener sustituciones; ajustando los parámetros de penalización por emparejamiento erróneo específicos en base a los tipos de mutación específicos que son comunes en determinados tipos de tumores (por ejemplo, C→T en melanoma); o ajustando los parámetros de penalización por emparejamiento erróneo específicos en base a tipos de mutación específicos que son comunes en determinados tipos de muestra (por ejemplo, sustituciones que son comunes en muestras FFIP). La especificidad reducida (tasa incrementada de positivos falsos) en las regiones génicas evaluadas debido a la alineación incorrecta se puede evaluar mediante un examen manual o automatizado de todas las llamadas de mutación en las muestras secuenciadas. Aquellas regiones que son propensas a llamadas de mutación espurias debido a alineación incorrecta se pueden someter a los mismos remedios de alineación que anteriormente. En los casos en que se encuentre que ningún remedio algorítmico resulta posible, las "mutaciones" de las regiones problemáticas se pueden clasificar o descartar del panel de prueba.

Los procedimientos descritos en el presente documento permiten el uso de múltiples procedimientos o algoritmos de alineación ajustados individualmente para optimizar el rendimiento en la secuenciación de intervalos subgenómicos asociados con reordenamientos, por ejemplo, indels, en particular en procedimientos que se basan en secuenciación masiva en paralelo de un gran número de acontecimientos genéticos diversos en un gran número de genes diversos, por ejemplo, de muestras tumorales. En implementaciones, para analizar las lecturas se usan múltiples procedimientos de alineación que se personalizan individualmente o se ajustan a cada uno de una serie de reordenamientos en diferentes genes. En implementaciones, el ajuste puede ser una función de (uno o más de) el gen (u otro intervalo subgenómico) que se está secuenciando, el tipo de tumor en la muestra, la variante que se está secuenciando o una característica de la muestra o del sujeto. Esta selección o uso de condiciones de alineación finamente ajustadas a una serie de intervalos subgenómicos que se van a secuenciar permite la optimización de la velocidad, la sensibilidad y la especificidad. El procedimiento es en particular eficaz cuando se optimiza la alineación de lecturas para un número relativamente grande de intervalos subgenómicos diversos. En implementaciones, el procedimiento incluye el uso de procedimientos de alineación optimizados para reordenamientos y de otros optimizados para intervalos subgenómicos no asociados con reordenamientos.

Por tanto, en una implementación, un procedimiento descrito en el presente documento, por ejemplo, un procedimiento para analizar una muestra tumoral, comprende un procedimiento de alineación para los reordenamientos descritos en el presente documento.

En general, la detección exacta de mutaciones indel es un ejercicio de alineación, ya que la tasa de indels espurios en las plataformas de secuenciación deshabilitadas en el presente documento es relativamente baja (por tanto, incluso una pocas observaciones de indels correctamente alineados puede ser una sólida prueba de mutación). Sin embargo, la alineación exacta en presencia de indels puede ser difícil (especialmente a medida que se incrementa la longitud del indel). Además de los problemas generales asociados con la alineación, por ejemplo, de las sustituciones, el indel en sí mismo puede causar problemas con la alineación. (Por ejemplo, una delección de 2 pb de una repetición de dinucleótido no se puede colocar fácilmente de forma definitiva). Tanto la sensibilidad como la especificidad se pueden reducir debido a la colocación incorrecta de lecturas que contienen indels aparentemente más cortos (<15 pb). Los indels más grandes (que se acercan en magnitud a la longitud de las lecturas individuales: 36 pb en nuestro procedimiento actual) pueden causar un fallo de alineación de la lectura, haciendo imposible la detección del indel en el conjunto estándar de lecturas alineadas.

Se pueden usar las bases de datos de mutaciones del cáncer para abordar estos problemas y mejorar el rendimiento. Para reducir el descubrimiento de indels positivos falsos (mejorar la especificidad), se pueden examinar las regiones alrededor de los indels comúnmente esperados en busca de alineaciones problemáticas debido al contexto de la secuencia y abordarse de forma similar a las sustituciones anteriores. Para mejorar la sensibilidad de la detección de indels, se pueden usar varios enfoques diferentes para usar la información de los indels esperados en el cáncer. Por ejemplo, se pueden simular los indels esperados que contengan lecturas cortas e intentar la alineación. Se pueden estudiar las alineaciones y se pueden ajustar los parámetros de alineación de las regiones de indels problemáticas, por ejemplo, reduciendo las penalizaciones por apertura/extensión de huecos o alineando lecturas parciales (por ejemplo, la primera o segunda mitad de una lectura).

De forma alternativa, se puede intentar la alineación inicial no solo con el genoma de referencia normal, sino también con versiones alternativas del genoma que contengan cada una de las mutaciones indel del cáncer conocidas o probables. En este enfoque, las lecturas de indels que inicialmente no se pudieron alinear o que se alinearon incorrectamente se colocan con éxito en la versión alternativa (mutada) del genoma.

De esta manera, la alineación de indel (y, por lo tanto, la llamada) se puede optimizar para los genes/sitios de cáncer esperados. Como se usa en el presente documento, un algoritmo de alineación de secuencias incorpora un procedimiento o enfoque computacional usado para identificar desde qué punto en el genoma es más probable que se haya originado una secuencia de lectura (por ejemplo, una secuencia de lectura corta, por ejemplo, de la secuenciación de nueva generación) evaluando la similitud entre la secuencia de lectura y una secuencia de referencia. Se puede aplicar una variedad de algoritmos al problema de alineación de secuencias. Algunos algoritmos son relativamente lentos, pero permiten una especificidad relativamente alta. Estos incluyen, por ejemplo, algoritmos basados en programación dinámica. La programación dinámica es un procedimiento para

resolver problemas complejos dividiéndolos en etapas más simples. Otros enfoques son relativamente más eficaces, pero típicamente no son tan exhaustivos. Estos incluyen, por ejemplo, algoritmos heurísticos y procedimientos probabilísticos diseñados para la búsqueda en bases de datos a gran escala.

Los parámetros de alineación se usan en algoritmos de alineación para ajustar el rendimiento de un algoritmo, por ejemplo, para producir una alineación global o local óptima entre una secuencia de lectura y una secuencia de referencia. Los parámetros de alineación pueden dar ponderaciones para emparejamientos, emparejamientos erróneos e indels. Por ejemplo, las ponderaciones más bajas permiten alineaciones con más emparejamientos erróneos e indels.

El contexto de secuencia, por ejemplo, la presencia de secuencias repetitivas (por ejemplo, repeticiones en tándem, repeticiones intercaladas), regiones de baja complejidad, indels, pseudogenes o parálogos, puede afectar a la especificidad de alineación (por ejemplo, causar alineación incorrecta). Como se usa en el presente documento, alineación incorrecta se refiere a la colocación de pares de bases de la lectura corta en localizaciones incorrectas en el genoma.

La sensibilidad de alineación se puede incrementar cuando se selecciona un algoritmo de alineación o se ajusta un parámetro de alineación en base al tipo de tumor, por ejemplo, un tipo de tumor que tiende a tener una mutación o tipo de mutación particular.

La sensibilidad de alineación se puede incrementar cuando se selecciona un algoritmo de alineación o se ajusta un parámetro de alineación en base a un tipo de gen particular (por ejemplo, oncogén, gen supresor de tumores). Las mutaciones en diferentes tipos de genes asociados al cáncer pueden tener un impacto diferente según el fenotipo del cáncer. Por ejemplo, los alelos de oncogenes mutantes son típicamente dominantes. Los alelos supresores de tumores mutantes son típicamente recesivos, lo que significa que, en la mayoría de los casos, ambos alelos de un gen supresor de tumores se deben ver afectados antes de que se manifieste un efecto.

La sensibilidad de alineación se puede ajustar (por ejemplo, incrementar) cuando se selecciona un algoritmo de alineación o se ajusta un parámetro de alineación en base al tipo de mutación (por ejemplo, polimorfismo mononucleotídico, indel (inserción o delección), inversión, translocación, repetición en tándem).

La sensibilidad de alineación se puede ajustar (por ejemplo, incrementar) cuando se selecciona un algoritmo de alineación o se ajusta un parámetro de alineación en base al sitio de mutación (por ejemplo, un punto caliente de mutación). Un punto caliente de mutación se refiere a un sitio en el genoma donde las mutaciones se producen hasta 100 veces más frecuentemente que la tasa de mutación normal.

La sensibilidad/especificidad de alineación se puede ajustar (por ejemplo, incrementar) cuando se selecciona un algoritmo de alineación o se ajusta un parámetro de alineación en base al tipo de muestra (por ejemplo, una muestra FFIP).

Se pueden seleccionar algoritmos de alineación para ajustar (por ejemplo, incrementar) la sensibilidad/especificidad de alineación, en base al tipo de muestra (por ejemplo, una muestra FFIP, una muestra de sangre o una muestra de aspirado de médula ósea).

La optimización de la alineación se describe en la técnica, por ejemplo, como se establece en la publicación de solicitud de patente internacional n.º WO 2012/092426.

Llamada de mutaciones

La llamada de bases se refiere a la salida sin procesar de un dispositivo de secuenciación. La llamada de mutaciones se refiere al procedimiento de selección de un valor de nucleótido, por ejemplo, A, G, T o C, para una posición nucleotídica que se está secuenciando. Típicamente, las lecturas de secuenciación (o llamada de bases) para una posición proporcionarán más de un valor, por ejemplo, algunas lecturas darán una T y algunas darán una G. La llamada de mutaciones es el procedimiento de asignación de un valor de nucleótido, por ejemplo, uno de esos valores a la secuencia. Aunque se denomina llamada de "mutaciones", se puede aplicar para asignar un valor de nucleótido a cualquier posición nucleotídica, por ejemplo, posiciones correspondientes a alelos mutantes, alelos naturales, alelos que no se han caracterizado como mutantes o naturales, o a posiciones no caracterizadas por su variabilidad. Los procedimientos para llamada de mutaciones pueden incluir uno o más de los siguientes: hacer llamadas independientes basadas en la información en cada posición en la secuencia de referencia (por ejemplo, examinar las lecturas de secuencia; examinar las llamadas de bases y las puntuaciones de calidad; calcular la probabilidad de las bases observadas y las puntuaciones de calidad dado un genotipo potencial; y asignar genotipos (por ejemplo, usando la regla de Bayes)); retirar los positivos falsos (por ejemplo, usando umbrales de profundidad para rechazar SNP con una profundidad de lectura mucho menor o mayor de lo esperado; realineación local para retirar los positivos falsos debidos a pequeños indels); y realizar un análisis basado en el desequilibrio de acoplamiento (DA)/imputación para refinar las llamadas.

Las ecuaciones para calcular la probabilidad del genotipo asociada con un genotipo y posición específicos se describen, por ejemplo, en Li H. y Durbin R. *Bioinformatics*, 2010; 26(5): 589-95. La expectativa previa de una mutación particular en cierto tipo de cáncer se puede usar al evaluar muestras de ese tipo de cáncer. Dicha probabilidad se puede derivar de bases de datos públicas de mutaciones de cáncer, por ejemplo, Catalogue of Somatic Mutation in Cancer (COSMIC), HGMD (Human Gene Mutation Database), The SNP Consortium, Breast Cancer Mutation Data Base (BIC) y Breast Cancer Gene Database (BCGD).

Se describen ejemplos de análisis basados en DA/imputación, por ejemplo, en Browning B.L. y Yu Z. *Am. J. Hum. Genet.* 2009, 85(6):847-61. Se describen ejemplos de procedimientos de llamada de SNP de baja cobertura, por ejemplo, en Li Y. *et al.*, *Annu. Rev. Genomics Hum. Genet.* 2009, 10:387-406.

Después de la alineación, se puede realizar la detección de sustituciones usando un procedimiento de llamada, por ejemplo, el procedimiento bayesiano de llamada de mutaciones; que se aplica a cada base en cada uno de los intervalos subgenómicos, por ejemplo, exones del gen que se va a evaluar, donde se observa la presencia de alelos alternativos. Este procedimiento comparará la probabilidad de observar los datos de lectura en presencia de una mutación con la probabilidad de observar los datos de lectura solo en presencia de un error de llamada de bases. Las mutaciones se pueden llamar si esta comparación soporta suficientemente fuerte la presencia de una mutación.

Se han desarrollado procedimientos que abordan desviaciones limitadas de frecuencias del 50 % o 100 % para el análisis del ADN del cáncer (por ejemplo, SNVMix -*Bioinformatics*. 15 de marzo de 2010; 26(6): 730-736). Sin embargo, el procedimiento divulgado en el presente documento permite considerar la posibilidad de la presencia de un alelo mutante en cualquier lugar entre el 1 % y el 100 % del ADN de la muestra, y especialmente a niveles inferiores al 50 %. Este enfoque es en particular importante para la detección de mutaciones en muestras FFIP de baja pureza de ADN tumoral natural (multiclonal).

Una ventaja de un enfoque bayesiano de detección de mutaciones es que la comparación de la probabilidad de la presencia de una mutación con la probabilidad de error de llamada de bases solo se puede ponderar mediante una expectativa previa de la presencia de una mutación en el sitio. Si se observan algunas lecturas de un alelo alternativo en un sitio mutado con frecuencia para el tipo de cáncer dado, entonces la presencia de una mutación se puede llamar con confianza incluso si la cantidad de pruebas de mutación no cumple con los umbrales normales. Esta flexibilidad se puede usar para incrementar la sensibilidad de detección para mutaciones aún más raras/muestras de pureza aún menor, o para hacer que la prueba sea más robusta a las disminuciones en la cobertura de lectura. La probabilidad de que un par de bases aleatorio en el genoma esté mutado en el cáncer es de $\sim 1e-6$. La probabilidad de mutaciones específicas en muchos sitios en un panel de genoma de cáncer multigénico típico puede ser órdenes de magnitud mayor. Estas probabilidades se pueden derivar de bases de datos públicas de mutaciones de cáncer (por ejemplo, COSMIC). La llamada de indels es un procedimiento de búsqueda de bases en los datos de secuenciación que difieren de la secuencia de referencia por inserción o delección, que típicamente incluye una puntuación de confianza asociada o una métrica de prueba estadística.

Los procedimientos de llamada de indels pueden incluir las etapas de identificar un indel candidato, calcular la probabilidad de genotipo a través de la realineación local y realizar una inferencia y llamada de genotipo basada en DA. Típicamente, se usa un enfoque bayesiano para obtener indels candidatos potenciales, y a continuación se someten a prueba estos candidatos conjuntamente con la secuencia de referencia en un marco bayesiano.

Los algoritmos para generar indels candidatos se describen, por ejemplo, en McKenna A. *et al.*, *Genome Res.* 2010; 20(9): 1297-303; Ye K. *et al.*, *Bioinformatics*, 2009; 25(21):2865-71; Lunter G. y Goodson M. *Genome Res.* 2010, publicación en línea previa a la impresión; Li H. *et al.*, *Bioinformatics* 2009, *Bioinformatics* 25(16):2078-9.

Los procedimientos para generar llamadas de indels y probabilidades de genotipo a nivel individual incluyen, por ejemplo, el algoritmo Dindel (Albers C.A. *et al.*, *Genome Res.* 2011;21(6):961-73). Por ejemplo, se puede usar el algoritmo EM bayesiano para analizar las lecturas, realizar llamadas de indels iniciales y generar probabilidades de genotipo para cada indel candidato, seguido de la imputación de genotipos usando, por ejemplo, QCALL (Le S.Q. y Durbin R. *Genome Res.* 2011;21(6):952-60). Los parámetros, tales como las expectativas previas de observar el indel, se pueden ajustar (por ejemplo, incrementar o disminuir), en base al tamaño o la localización de los indels.

La optimización de la llamada de mutaciones se describe en la técnica, por ejemplo, como se establece en la publicación de solicitud de patente internacional n.º WO 2012/092426.

Otras implementaciones

En implementaciones de un procedimiento descrito en el presente documento, se usa una etapa o parámetro en el procedimiento para modificar una etapa o parámetro posterior en el procedimiento.

En una implementación, se usa una característica de la muestra tumoral para modificar una etapa o parámetro

posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección; diseño o selección de cebos; condiciones de hibridación; secuenciación; mapeo de lecturas; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

5 En una implementación, se usa una característica de un ácido nucleico de tumor aislado, o control, para modificar una etapa o parámetro posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección; diseño o selección de cebos; condiciones de hibridación; secuenciación; mapeo de lecturas; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

10 En una implementación, se usa una característica de una colección para modificar una etapa o parámetro posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección posterior; diseño o selección de cebos; condiciones de hibridación; secuenciación; mapeo de lecturas; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

15 En una implementación, se usa una característica de una captura de colección para modificar una etapa o parámetro posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección posterior; diseño o selección de cebos; condiciones de hibridación; secuenciación; mapeo de lecturas; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

20 En una implementación, se usa una característica del procedimiento de secuenciación para modificar una etapa o parámetro posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección posterior; diseño o selección de cebos; determinación posterior de condiciones de hibridación; secuenciación posterior; mapeo de lecturas; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

25 En una implementación, se usa una característica del conjunto de lecturas mapeadas para modificar una etapa o parámetro posterior en uno o más o todos de: aislamiento de ácido nucleico de dicha muestra; construcción de colección posterior; diseño o selección de cebos; determinación posterior de condiciones de hibridación; secuenciación posterior; mapeo de lecturas posterior; selección de un procedimiento de llamada de mutaciones; llamada de mutaciones o anotación de mutaciones.

30 En una implementación, el procedimiento comprende adquirir un valor para una característica de muestra tumoral, por ejemplo, adquirir un valor: para la proporción de células tumorales en dicha muestra, para la celularidad de dicha muestra tumoral; o de una imagen de la muestra tumoral.

35 En implementaciones, el procedimiento incluye, en respuesta a dicho valor adquirido para una característica de muestra tumoral, seleccionar un parámetro para: aislamiento de ácido nucleico de una muestra tumoral, construcción de colección; diseño o selección de cebos; hibridación de cebos/miembros de colección; secuenciación; o llamada de mutaciones.

40 En una implementación, un procedimiento comprende además adquirir un valor para la cantidad de tejido tumoral presente en dicha muestra tumoral, comparar dicho valor adquirido con un criterio de referencia y, si se cumple dicho criterio de referencia, aceptar dicha muestra tumoral, por ejemplo, aceptar dicha muestra tumoral si dicha muestra tumoral contiene más de un 30, 40 o 50 % de células tumorales.

45 En una implementación, un procedimiento comprende además adquirir una submuestra enriquecida en células tumorales, por ejemplo, mediante macrodissección de tejido tumoral de dicha muestra tumoral, de una muestra tumoral que no cumple con el criterio de referencia.

50 En una implementación, un procedimiento comprende además determinar si un control primario, por ejemplo, una muestra de sangre, está disponible y, de ser así, aislar un ácido nucleico de control (por ejemplo, ADN) de dicho control primario.

55 En una implementación, un procedimiento comprende además determinar si hay TCN presente en dicha muestra tumoral (por ejemplo, cuando no está disponible una muestra de control primario).

60 En una implementación, un procedimiento comprende además adquirir una submuestra enriquecida en células no tumorales, por ejemplo, mediante macrodissección de tejido no tumoral de dicho TCN en una muestra tumoral no acompañada por un control primario.

En una implementación, un procedimiento comprende además determinar que no está disponible un control primario ni TCN y marcar dicha muestra tumoral para su análisis sin un control emparejado.

65 En una implementación, un procedimiento comprende además aislar ácido nucleico de dicha muestra tumoral para proporcionar una muestra de ácido nucleico tumoral aislada.

En una implementación, un procedimiento comprende además aislar un ácido nucleico de un control para proporcionar una muestra de ácido nucleico de control aislada.

5 En una implementación, un procedimiento comprende además rechazar una muestra sin ácido nucleico detectable.

En una implementación, un procedimiento comprende además adquirir un valor para el rendimiento de ácido nucleico en dicha muestra de ácido nucleico aislada y comparar el valor adquirido con un criterio de referencia, por ejemplo, en el que si dicho valor adquirido es menor que dicho criterio de referencia, entonces se amplifica dicha muestra de ácido nucleico aislada antes de la construcción de la colección.

En una implementación, un procedimiento comprende además adquirir un valor para el tamaño de los fragmentos de ácido nucleico en dicha muestra de ácido nucleico aislada y comparar el valor adquirido con un criterio de referencia, por ejemplo, un tamaño, por ejemplo, tamaño promedio, de al menos 300, 600 o 900 pb. En respuesta a esta determinación, se puede ajustar o seleccionar un parámetro descrito en el presente documento.

En una implementación, un procedimiento comprende además adquirir una colección en la que el tamaño de dichos fragmentos de ácido nucleico es menor o igual que un valor de referencia, dicha colección se prepara sin una etapa de fragmentación entre el aislamiento del ADN y la preparación de la colección.

En una implementación, un procedimiento comprende además adquirir fragmentos de ácido nucleico y, si el tamaño de dichos fragmentos de ácido nucleico es igual o mayor que un valor de referencia, se fragmentan y a continuación se preparan en una colección.

25 En una implementación, un procedimiento comprende además marcar cada uno de una pluralidad de miembros de la colección, por ejemplo, mediante la adición de una secuencia de ácido nucleico distinta identificable (un código de barras) a cada uno de una pluralidad de miembros.

En una implementación, un procedimiento comprende además unir un cebador a cada uno de una pluralidad de miembros de la colección.

En una implementación, un procedimiento comprende además proporcionar una pluralidad de cebos y seleccionar una pluralidad de cebos, respondiendo dicha selección a: 1) una característica del paciente, por ejemplo, edad, estadio del tumor, tratamiento previo o resistencia; 2) tipo de tumor; 3) una característica de la muestra tumoral; 4) una característica de una muestra de control; 5) presencia o tipo de control; 6) una característica de la muestra de ácido nucleico tumoral (o de control) aislada; 7) una característica de la colección; 8) una mutación que se sabe que está asociada con el tipo de tumor en la muestra tumoral; 9) una mutación que no se sabe que esté asociada con el tipo de tumor en la muestra tumoral; 10) la capacidad de secuenciar (o hibridar o recuperar) una secuencia preseleccionada o de identificar una mutación preseleccionada, por ejemplo, la dificultad asociada con secuenciar una región con alto contenido de GC o un reordenamiento; u 11) los genes que se están secuenciando.

En una implementación, un procedimiento comprende además responder, por ejemplo, a una determinación de un número bajo de células tumorales en dicha muestra tumoral, seleccionando un cebo, o pluralidad de cebos, proporcionando una captura relativamente altamente eficiente de los miembros de un primer gen en comparación con los miembros de un segundo gen, por ejemplo, en el que una mutación en el primer gen está asociada al fenotipo tumoral para el tipo de tumor de la muestra tumoral.

En una implementación, un procedimiento comprende además adquirir un valor para la característica de captura de la colección, por ejemplo, la concentración o representación de ácido nucleico, y comparar el valor adquirido con un criterio de referencia para la concentración, o para la representación, de ácido nucleico.

En una implementación, un procedimiento comprende además seleccionar una colección con un valor para una característica de colección que no cumple con el criterio de referencia para reprocesado.

55 En una implementación, un procedimiento comprende además seleccionar una colección con un valor para una característica de colección que cumple con el criterio de referencia para la cuantificación de la colección.

En una implementación, un procedimiento comprende además proporcionar una asociación de un tipo de tumor, un gen y una alteración genética (un TGA) para un sujeto.

En una implementación, un procedimiento comprende además proporcionar una base de datos preseleccionada que tiene una pluralidad de elementos, en la que cada elemento comprende un TGA.

En una implementación, un procedimiento comprende además caracterizar un TGA de un sujeto que comprende: determinar si dicho TGA está presente en una base de datos preseleccionada, por ejemplo, una base de datos de

TGA validados;

asociar información para el TGA de la base de datos predeterminada con dicho TGA (anotación) de dicho sujeto; y

5 opcionalmente, determinar si un segundo o posterior TGA para dicho sujeto está presente en dicha base de datos preseleccionada y, de ser así, asociar información para el segundo o posterior TGA de la base de datos predeterminada con dicho segundo TGA presente en dicho paciente.

10 En una implementación, un procedimiento comprende además memorizar la presencia o ausencia de un TGA, y opcionalmente una anotación asociada, de un sujeto para elaborar un informe.

En una implementación, un procedimiento comprende además transmitir dicho informe a una parte receptora.

15 En una implementación, un procedimiento comprende además caracterizar un TGA de un sujeto que comprende: determinar si dicho TGA está presente en una base de datos preseleccionada, por ejemplo, una base de datos de TGA validados;

20 determinar si un TGA que no está en dicha base de datos preseleccionada tiene un G o A clínicamente relevante conocido y, de ser así, proporcionar una entrada para dicho TGA en dicha base de datos preseleccionada.

En una implementación, un procedimiento comprende además memorizar la presencia o ausencia de una mutación encontrada en el ADN de la muestra tumoral de un sujeto para elaborar un informe.

25 En una implementación, un procedimiento comprende además memorizar la presencia o ausencia de un TGA, y opcionalmente una anotación asociada, de un sujeto para elaborar un informe.

En una implementación, un procedimiento comprende además transmitir dicho informe a una parte receptora.

30 Se proporciona una representación de diagrama de flujo de una implementación de un procedimiento para el análisis multigénico de una muestra tumoral en la FIG. 1.

Ejemplificación

35 Esta invención se ilustra adicionalmente mediante los siguientes ejemplos que no se deben interpretar como limitantes.

EJEMPLO 1: aislamiento de ácidos nucleicos de una muestra tumoral

40 Los ácidos nucleicos (ADN y ARN) se aislaron de varios tipos de muestras, por ejemplo, muestras fijadas con formol e incluidas en parafina (FFIP), de sangre y de aspirado de médula ósea (AMO).

45 Por ejemplo, para aislar el ADN genómico de una muestra FFIP, se mezclaron secciones de 3 x 20 µm cortadas de un bloque de parafina con 400 µl de tampón FTL mediante agitación en vórtex y se incubaron a 90 °C durante 15 minutos en un tubo de centrifuga de 1,5 ml. Un intervalo de 88-92 °C fue aceptable para la incubación. A continuación, se incubó la muestra con 20 µl de proteinasa K a 55 °C durante 6 horas y 10 µl de RNasa (1 mg/ml) a temperatura ambiente durante 5 minutos. A continuación, se añadieron 460 µl de tampón BL y 500 µl de etanol absoluto a la muestra. La solución de muestra resultante se mantuvo a temperatura ambiente hasta su uso

50 posterior.

Para preparar la columna para la unión al ADN, se añadieron 100 µl de tampón de equilibrado a una columna MicroElute y se centrifugó la columna a 10 000 x g durante 30 segundos. Se transfirieron 700 µl de la solución de muestra descrita anteriormente a la columna MicroElute y se centrifugó la columna a 10 000 x g durante 1 minuto.

55 Se repitió la etapa de centrifugación si el fluido no pasó completamente a través de la columna MicroElute. La solución de muestra restante se aplicó a la columna MicroElute de la misma manera como se describe anteriormente. A continuación, se trató la columna MicroElute con 500 µl de tampón HB y se centrifugó a 10 000 x g durante 1 minuto. A continuación, se añadieron 700 µl de tampón de lavado de ADN diluido con etanol a la columna MicroElute y se centrifugó la columna a 10 000 x g durante 1 minuto. Se lavó de nuevo la columna MicroElute

60 usando 700 µl de tampón de lavado de ADN diluido con etanol, se centrifugó a 10 000 x g durante 1 minuto y se centrifugó a >13 000 x g durante 3 minutos para secar la columna. Se colocó la columna MicroElute en un tubo de centrifuga estándar de 1,5 ml con la parte superior retirada. Se añadieron 50-75 µl de tampón de elución precalentado a 70 °C a la columna y se incubó a temperatura ambiente durante 3 minutos. Se centrifugó la columna en un tubo de recogida a >13 000 x g durante 1 minuto. Se añadieron otros 50-75 µl de tampón de elución

65 precalentado a 70 °C a la columna MicroElute y se incubó a temperatura ambiente durante 3 minutos. Se centrifugó de nuevo la columna en un tubo de recogida a >13 000 x g durante 1 minuto. Se transfirió la solución completa a

un tubo de centrifuga nuevo de 1,5 ml y se almacenó a -20 °C.

El tampón FTL, la proteinasa K, el tampón BL, el tampón de equilibrado, la columna MicroElute, el tampón HB, el tampón de lavado de ADN y el tampón de elución se proporcionaron en el kit E.Z.N.A.™ FFPE DNA (OMEGA bio-
 5 tek, Norcross, GA; Cat. Nos. D3399-00, D3399-01 y D3399-02).

Se divulgan procedimientos adicionales para aislar ácidos nucleicos (por ejemplo, ADN) de tejidos fijados con formaldehído o paraformaldehído e incluidos en parafina (FFIP), por ejemplo, en Cronin M. *et al.*, (2004) *Am J Pathol.* 164(1):35-42; Masuda N. *et al.*, (1999) *Nucleic Acids Res.* 27(22):4436-4443; Specht K. *et al.*, (2001) *Am J Pathol.* 158(2):419-429, el protocolo RecoverAll™ Total Nucleic Acid Isolation de Ambion (Ambion, n.º de cat. AM1975, septiembre de 2008), el manual técnico del kit Maxwell® 16 FFPE Plus LEV DNA Purification (Promega literatura n.º TM349, febrero de 2011), el manual del kit QIAamp® DNA FFPE Tissue (Qiagen, n.º de cat. 37625, octubre de 2007). El kit RecoverAll™ Total Nucleic Acid Isolation usa xileno a temperaturas elevadas para solubilizar muestras incluidas en parafina y un filtro de fibra de vidrio para capturar ácidos nucleicos. El kit Maxwell®
 10 16 FFPE Plus LEV DNA Purification se usa con el instrumento Maxwell® 16 para la purificación de ADN genómico de secciones de 1 a 10 µm de tejido FFIP. El ADN se purifica usando partículas paramagnéticas (PMP) recubiertas de sílice y se eluye en un volumen de elución bajo. El kit QIAamp® DNA FFPE Tissue usa la tecnología de micro-ADN QIAamp® para la purificación de ADN genómico y mitocondrial.

Procedimientos ejemplares adicionales para aislar ácidos nucleicos (por ejemplo, ARN y ADN) de muestras FFIP, de sangre y de aspirados de médula ósea se describen a continuación. Se extrajo ARN de tubos de sangre de ARN PAXgene® fresca o congelada usando el kit Maxwell® 16 LEV simplyRNA Blood para Maxwell CSC® (Promega, n.º de cat. A1S1311). El ARN y el ADN de los aspirados de médula ósea se extrajeron usando kit Maxwell® 16 LEV simplyRNA Blood para Maxwell CSC® (Promega, n.º de cat. A1S1311) y el kit Maxwell® 16
 25 Blood DNA Purification para Maxwell® CSC (Promega, n.º de cat. AS1010), respectivamente. El ARN y el ADN de las muestras FFIP se coextrajeron usando el kit Maxwell® 16 CSC RNA FFPE (Promega, n.º de cat. AS1360) y el kit Maxwell 16 FFPE Tissue LEV DNA (Promega, n.º de cat. AS1130), respectivamente. La desparafinación de las muestras FFIP se realizó usando la solución híbrida CitriSolv™ (Fisher Scientific, n.º de cat. 04-355-121). El ADN y el ARN de líneas celulares se purificaron usando el kit Maxwell® 16 LEV simplyRNA Blood para Maxwell CSC® (Promega, n.º de cat. A1S1311) y el kit Maxwell® 16 Blood RNA Purification para Maxwell® CSC (Promega, n.º de cat. AS1010), respectivamente.

EJEMPLO 2: cuantificación de ácidos nucleicos aislados

El ADN aislado se cuantificó usando el reactivo Quant-iT™ PicoGreen® dsDNA (Life Technologies, n.º de cat. P7581). El ARN aislado se cuantificó usando el reactivo Quant-iT™ RiboGreen® RNA (Life Technologies, n.º de cat. R11491).

EJEMPLO 3: corte del ADN

El instrumento Covaris™ E210 con refrigerador circulante se ajustó a 4 °C. Se llenó el tanque de agua del instrumento con agua destilada/desionizada hasta el nivel "6" en la línea de llenado. Se inició el software SonoLab™ y se permitió que el sistema ejecutara la secuencia de búsqueda cuando se solicitó. Se desgasificó el agua en el tanque del instrumento durante al menos 45 minutos antes de cortar las muestras.

Para preparar las muestras de ADN genómico para el corte, las muestras se cuantificaron primero usando un ensayo PicoGreen® (Invitrogen) en un lector de microplacas (Spectramax M2, Molecular Devices). En base a la concentración, se usaron 120 µl de ADN de entrada deseado (2 ng/µl) con TE bajo (Tris 10 mM, EDTA 0,2 mM, pH 8,0) para el experimento. Se pipetearon lentamente las muestras individuales de 100 µl en los Covaris MicroTUBE (Covaris, n.º de cat. 520045) a través de los septa en la tapa del tubo. Los Covaris MicroTUBE se colocaron a continuación en la gradilla de tubos de la serie E de Covaris. Para el corte de 200 pb, los ajustes fueron como sigue: ciclo de trabajo del 10 %, intensidad de 5, 200 ciclos/ráfaga, tiempo de 180 segundos y modo de barrido de frecuencia. Después del corte, se centrifugaron brevemente los Covaris MicroTUBE usando un adaptador apropiado en una minicentrífuga, y se transfirieron las muestras cortadas a tubos de microcentrifuga limpios de 1,5 ml. Cada muestra de ADN cortado se purificó usando una columna QIAGEN MinElute®. En resumen, se añadió tampón QIAGEN PBI 5x a la muestra en un tubo de microcentrifuga de 1,5 ml (por ejemplo, se añadieron 500 µl de tampón PBI a 100 µl de muestra). Cada muestra se agitó en vórtex, se centrifugó brevemente y se transfirió a una columna de centrifugación MinElute. Se centrifugó la columna de centrifugación MinElute a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. Se añadieron 750 µl de tampón QIAGEN PE a la columna, se centrifugó a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. La columna de centrifugación se centrifugó de nuevo a 13 000 rpm durante 1 minuto y se transfirió a un tubo de microcentrifuga limpio de 1,5 ml. Se secó la columna al aire durante 2-3 minutos. Para la primera elución, se añadieron 18 µl de tampón de elución QIAGEN a cada columna, se incubaron durante 2-3 minutos y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Para la segunda elución, se añadieron 15 µl de tampón de elución QIAGEN, se incubaron durante 1 minuto y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Se recogió el eluyente y se descartó la columna de centrifugación.

Típicamente, se usan 200 ng para el corte del ADN, pero la cantidad de ADN puede variar de 20 a 200 ng o más.

EJEMPLO 4: alternativa al corte del ADN

Este ejemplo describe un procedimiento alternativo para el corte del ADN del ejemplo 3.

Un ADN genómico bicatenario se desnaturaliza primero a ADN monocatenario y, a continuación, se mezcla con cebadores, ADN polimerasa (por ejemplo, ADN polimerasa Exo-), dNTP y una pequeña cantidad de ddNTP. La secuencia cebadora puede ser un hexámero aleatorio o un hexámero aleatorio marcado con una secuencia de adaptador en el extremo 5'. Los procedimientos para usar la amplificación de hexámeros aleatorios marcados para clonar y secuenciar cantidades diminutas de ADN se describen, por ejemplo, en Wong K.K. *et al.*, *Nucleic Acids Res.* 1996; 24(19):3778-83. La reacción se incubó en las condiciones que permiten la hibridación cebador-molde y la síntesis de ADN. La síntesis de ADN finalizará cuando se incorpore un ddNTP en la primera cadena recién sintetizada. La longitud de la primera cadena de ADN sintetizada se puede controlar mediante la proporción entre dNTP y ddNTP. Por ejemplo, la proporción molar entre dNTP y ddNTP es al menos aproximadamente 1000:1, aproximadamente 5000:1 o aproximadamente 10000:1. Después de la síntesis de la primera cadena, los fragmentos cortos (tales como cebadores y la primera cadena de ADN sintetizada con longitud corta y ddNTP) se pueden retirar mediante selección por tamaño (por ejemplo, usando una columna de centrifugación de selección por tamaño). La primera cadena de ADN resultante se mezcla con cebadores (por ejemplo, hexámeros aleatorios o hexámeros aleatorios marcados con una secuencia de adaptador), ADN polimerasa (por ejemplo, ADN polimerasa Exo+) y dNTP. Se puede usar una ADN polimerasa Exo+ para retirar el 3'-ddNTP terminal de la primera cadena de ADN o incluso para generar extremos romos sobre el segundo sitio de cebado. A continuación, la reacción se incubó en las condiciones que permiten la hibridación cebador-molde y la síntesis de ADN. Después de la síntesis de la segunda cadena, los fragmentos de ADN bicatenarios resultantes se pueden purificar y usar directamente en la construcción de la colección. De forma alternativa, los fragmentos de ADN bicatenario se pueden amplificar por PCR usando cebadores que contienen secuencias de adaptador si estas secuencias de adaptador se han incluido en los cebadores para la síntesis de la primera y segunda cadena. Los cebadores para la amplificación por PCR también pueden incluir las secuencias completas y/o secuencias de código de barras.

EJEMPLO 5: síntesis de ADNc

Antes de la construcción de la colección, el ARN aislado se convirtió en ADNc bicatenario, que sirvió como entrada para producir colecciones de secuenciación ligadas a adaptador.

El ADNc bicatenario se creó en un procedimiento de dos etapas. En primer lugar, el ARN total se hibridó con una mezcla de cebadores de ADN de hexámero aleatorio (N6) y oligo-dT (Integrated DNA Technologies (IDT)). Los cebadores hibridados se extendieron mediante una retrotranscriptasa (RT) (Promega, n.º de cat. M3683) que usó la cadena de ARN como molde y generó un ADN monocatenario complementario al ARN de entrada. La primera cadena de ADN generada durante la síntesis de la primera cadena (SPC) permaneció hibridada con el molde de ARN tras completarse la reacción. A continuación, se produjo la síntesis de la segunda cadena (SSC) cuando se añadió una mezcla que contenía una ADN polimerasa, ADN ligasa y RNasa H al híbrido ADN-ARN (New England BioLabs, ref. de cat. E6111L). La RNasa H cortó el ARN unido al ADN, creando por tanto cebadores de ARN que se usaron por la ADN polimerasa para cebar la síntesis de la segunda cadena de ADN usando la primera cadena de ADN como molde. El producto de la síntesis de la segunda cadena fue un ADN bicatenario compatible con la construcción de la colección.

El protocolo de síntesis de ADNc se diseñó para adaptarse al procesamiento tanto de entradas de alta calidad (por ejemplo, ARN purificado de entradas congeladas frescas recogidas y almacenadas apropiadamente (por ejemplo, sedimentos celulares), tubos de ARN en sangre PAXgene, aspirados de médula ósea en RNAlater) como de baja calidad (por ejemplo, ARN purificado de muestras FFIP) usando un flujo de trabajo semiautomatizado con diferencias mínimas en la configuración y ejecución del ensayo en base a la calidad del ARN de entrada. La calidad del ARN se analizó en un bioanalizador. Los picos claros de ARN ribosómico (ARNr) indicaron que el ARNm era en gran parte de longitud completa.

La longitud típica del fragmento de ADNc de las entradas de ARN de baja calidad fue de entre 100 y 500 pb y no requirió procesamiento adicional (por ejemplo, fragmentación) antes de comenzar la construcción de la colección. La longitud típica del fragmento de ADNc de las entradas de ARN de alta calidad excedía en general los 500 pb (hasta 2-3 kpb) y el ADNc se fragmentó antes de usarse como entrada para la construcción de la colección. La fragmentación se logró mediante sonicación usando el mismo protocolo que se usó para el ADN genómico (Ejemplo 3).

Procedimientos ejemplares adicionales para sintetizar ADNc a partir de ARN total aislado de muestras FFIP se describen como sigue. Por ejemplo, se generó ADNc a partir de ARN total extraído de una única sección de tejido FFIP de 5-10 µm usando el kit Roche High Pure y se retrotranscribió a ADNc con cebadores de hexámeros aleatorios mediante el sistema de síntesis de primera cadena SuperScript® III (Invitrogen). El ADNc bicatenario se

preparó con el módulo de síntesis de segunda cadena de ARNm NEBNext® (New England Biolabs) y se usó como entrada para la construcción de la colección, la captura de híbridos y la secuenciación como para las muestras de ADN FFIP. El análisis de los niveles de expresión se realizó con una combinación de herramientas de análisis.

5 EJEMPLO 6: preparación de la colección

Reacción de reparación de extremos

Se descongelaron los reactivos de reparación de extremos (NEB, n.º E6050L) y se preparó una mezcla maestra de reparación de extremos sobre hielo. Para preparar 70 µl de mezcla maestra por muestra, se mezclaron 55 µl de agua libre de nucleasas con 10 µl de tampón de reacción de reparación de extremos 10x y 5 µl de mezcla de enzimas de reparación de extremos. A continuación, se añadieron 70 µl de mezcla maestra a 30 µl de cada muestra de ADN cortado en una placa de PCR de 96 pocillos sobre hielo. Se incubó la reacción en un termociclador a 20 °C durante 30 minutos. Cada muestra se purificó usando una columna QIAGEN MinElute®. En resumen, se añadió tampón QIAGEN PBI 5x a la muestra (por ejemplo, se añadieron 500 µl de tampón PBI a 100 µl de muestra) en un tubo de microcentrífuga de 1,5 ml. Cada muestra se agitó en vórtex, se centrifugó brevemente y se transfirió a una columna de centrifugación MinElute. Se centrifugó la columna de centrifugación MinElute a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. Se añadieron 750 µl de tampón QIAGEN PE a la columna, se centrifugó a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. La columna de centrifugación se centrifugó de nuevo a 13 000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga limpio de 1,5 ml. Se secó la columna al aire durante 2-3 minutos. Para la primera elución, se añadieron 22 µl de tampón de elución QIAGEN (Tris 10 mM, pH 8,5) a cada columna, se incubaron durante 2-3 minutos y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Para la segunda elución, se añadieron 22 µl de tampón de elución QIAGEN, se incubaron durante 1 minuto y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Se recogió el eluyente y se descartó la columna de centrifugación.

Adición de base 3'-A

Se descongelaron los reactivos de adición de base A (NEB, n.º E6053L) sobre hielo y se preparó una mezcla maestra de adición de base A sobre hielo. Para preparar 10 µl de mezcla maestra por muestra, se mezclaron 2 µl de agua libre de nucleasas con 5 µl de tampón de reacción de adición de colas dA 10x y 3 µl de fragmento Klenow (3'→5' exo-). Se añadieron 10 µl de mezcla maestra a 40 µl de cada muestra de ADN de extremos reparados purificado en una placa de PCR de 96 pocillos sobre hielo. Se incubó la reacción en un termociclador a 37 °C durante 30 min. Cada muestra se purificó usando una columna QIAGEN MinElute®. En resumen, se añadió tampón QIAGEN PBI 5x a la muestra (por ejemplo, se añadieron 250 µl de tampón PBI a 50 µl de muestra) en un tubo de microcentrífuga de 1,5 ml. Cada muestra se agitó en vórtex, se centrifugó brevemente y se transfirió a una columna de centrifugación MinElute. Se centrifugó la columna de centrifugación MinElute a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. Se añadieron 750 µl de tampón QIAGEN PE a la columna, se centrifugó a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. La columna de centrifugación se centrifugó de nuevo a 13 000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga limpio de 1,5 ml. Se secó la columna al aire durante 2-3 minutos. Para la primera elución, se añadieron 13 µl de tampón de elución QIAGEN (Tris 10 mM, pH 8,5) a cada columna, se incubaron durante 2-3 minutos y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Para la segunda elución, se añadieron 13 µl de tampón de elución QIAGEN, se incubaron durante 1 minuto y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Se recogió el eluyente y se descartó la columna de centrifugación.

Ligación de adaptadores múltiples

Se descongelaron los reactivos de ligación (NEB, n.º E6056L) y se preparó una mezcla maestra de ligación sobre hielo. Para preparar 36 µl de mezcla maestra por muestra, se añadieron 12 µl de tampón de reacción de ligación rápida 5x a 3,3 µl de adaptador Illumina Multiplex (15 µM, incluido en Illumina n.º de cat. PE-400-1001) (se usaron 3,3 µl de adaptador/1 µg de ADN de entrada inicial). Por ejemplo, para una muestra de 500 ng de ADN de entrada, los adaptadores se diluyeron primero en agua (2 µl de adaptadores más 2 µl de H₂O), a continuación se añadieron 3,3 µl de esta mezcla de adaptadores diluida, 15,7 µl de agua libre de nucleasas y 5 µl de ADN ligasa Quick T4 a la reacción de ligación. Para >1 µg de material de partida, se usaron >3,3 µl de adaptadores. Por tanto, se añadió menos agua para mantener el volumen total de mezcla de adaptadores diluida y agua libre de nucleasas en 19 µl.

Se añadieron 36 µl de mezcla maestra y 24 µl de cada muestra de ADN con cola de dA a los pocillos de una placa de PCR de 96 pocillos sobre hielo. Se incubó la reacción en un termociclador a 25 °C durante 30 min. Cada muestra se purificó usando una columna QIAGEN MinElute®. En resumen, se añadió tampón QIAGEN PBI 5x a la muestra (por ejemplo, se añadieron 300 µl de tampón PBI a 60 µl de muestra) en un tubo de microcentrífuga de 1,5 ml. Cada muestra se agitó en vórtex, se centrifugó brevemente y se transfirió a una columna de centrifugación MinElute. Se centrifugó la columna de centrifugación MinElute a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. Se añadieron 750 µl de tampón QIAGEN PE a la columna, se centrifugó a 13 000 rpm durante 1 minuto y se descartó el flujo continuo. La columna de centrifugación se centrifugó de nuevo a 13 000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga limpio de 1,5 ml. Se secó la columna al aire durante 2-3 minutos. Para

la primera elución, se añadieron 20 µl de tampón de elución QIAGEN (Tris 10 mM, pH 8,5) a cada columna, se incubaron durante 2-3 minutos y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Para la segunda elución, se añadieron 20 µl de tampón de elución QIAGEN, se incubaron durante 1 minuto y, a continuación, se centrifugaron a 13 000 rpm durante 1 minuto. Se recogió el eluyente y se descartó la columna de centrifugación.

5

Enriquecimiento por PCR

Se descongelaron los reactivos de PCR y se preparó una mezcla maestra de PCR sobre hielo. Para 62 µl de mezcla maestra por muestra, se usaron 50 µl de mezcla maestra 2X Phusion High Fidelity con tampón HF (Finnzyme, NEB n.º de cat. F-531S), 8 µl de agua libre de nucleasas, 2 µl de Illumina Primer 1.0 (25 µM) y 2 µl de Illumina Primer 2.0 (0,5 µM). A continuación, se mezclaron 62 µl de la mezcla maestra con 2 µl de Illumina Index Primer (25 µM, incluido en Illumina n.º de cat. PE-400-1001) con el código de barras apropiado y 36 µl de muestra de ADN ligada en una placa de PCR de 96 pocillos. Se incubó la reacción en un termociclador como sigue:

1 ciclo	98 °C 30 s
18 ciclos	98 °C 10 s
	65 °C 30 s
	72 °C 30 s
1 ciclo	72 °C 5 min
	4 °C mantener

15

Cada reacción de PCR se sometió a selección por tamaño con 1,8 volúmenes de microesferas AMPureXP (Agencourt; Beckman Coulter Genomics, n.º de cat. A6388). En resumen, se añadieron 1,8x microesferas AMPureXP a la muestra (por ejemplo, se añadieron 180 µl de microesferas a 100 µl de muestra) en un tubo de microcentrífuga de 1,5 ml, se agitaron en vórtex y se incubaron durante 5 minutos con mezcla de rotación de extremo a extremo. Se colocaron los tubos sobre un soporte magnético hasta que la solución se aclaró (2 minutos). Se descartó el sobrenadante sin perturbar las microesferas capturadas por el imán. Se añadieron 600 µl de etanol al 70 % recién preparado a las microesferas, se incubaron durante 1 min seguido de la retirada del etanol. Se añadió una segunda alícuota de 600 µl de etanol al 70 % recién preparado a las microesferas, se incubaron durante 1 min y se retiró el etanol. Se volvieron a colocar los tubos sobre el soporte magnético durante 1-2 minutos para recapturar las microesferas. Se retiró cualquier resto de etanol y se secaron las microesferas al aire a temperatura ambiente durante 5-10 minutos. Se añadieron 30 µl de tampón de elución QIAGEN a las microesferas, se agitaron en vórtex y se incubaron durante 2 minutos. Se volvieron a colocar los tubos sobre el soporte magnético hasta que la solución se aclaró (2 minutos). Se transfirió el sobrenadante a un tubo nuevo de 1,5 ml y se desecharon las microesferas. Se cuantificaron las muestras de ADN eluidas usando un ensayo Q-PCR. Estas cuantificaciones permitirán el agrupamiento equimolar para garantizar una representación equitativa de cada colección dentro de una selección de captura de híbridos agrupada.

20

25

30

EJEMPLO 7: selección de híbridos

Agrupar colecciones de muestras indexadas

Se prepararon agrupamientos (hasta 12 unidades) de colecciones que se habían indexado, purificado y cuantificado mediante Q-PCR sobre hielo. Se prepararon agrupamientos equimolares en tubos de microcentrífuga de 1,5 ml para garantizar que todas las muestras estuvieran representadas de forma equitativa en el procedimiento de selección de híbridos. La entrada total de ADN para cada uno de estos agrupamientos puede variar de 2000 ng a 500 ng. Típicamente, el ADN de entrada total es de 2000 ng. Por tanto, si se agrupan doce muestras, se pueden agrupar 166,67 ng de cada una para lograr un total de 2000 ng. El volumen final de un agrupamiento de colección de 2000 ng debería ser de 4 µl. Debido a las concentraciones variables de las colecciones indexadas, se puede preparar un agrupamiento con un volumen mayor, pero a continuación se debe secar el agrupamiento mediante centrífuga Speedvac (usando poco calor) y reconstituirse en 4 µl de agua libre de nucleasas.

45

Cuanto mayor sea el rendimiento en la construcción de una colección, mayor será la complejidad de la colección.

Hibridar las colecciones de ADN agrupadas con cebos de ARN biotinilado

50

En este experimento se usó el kit Agilent SureSelect Target Enrichment Paired End (n.º G3360A-J). Se descongelaron el tampón de hibridación n.º 3, el SureSelect Block n.º 1, el SureSelect Block n.º 2, el Paired End Primer Block 1.0, el Index Primer Block 1-12, RNase Block y el cebo de ARN biotinilado sobre hielo. Se prepararon

las siguientes mezclas maestras.

- a. Mezcla tampón de hibridación (13 µl por reacción):
 - 5 i. Tampón de hibridación n.º 1 (Agilent) - 25 µl
 - ii. Tampón de hibridación n.º 2 (Agilent) - 1 µl
 - 10 iii. Tampón de hibridación n.º 3 (Agilent) - 10 µl
 - iv. Tampón de hibridación n.º 4 (Agilent) - 13 µl
- b. Mezcla de bloqueo (8 µl por reacción):
 - 15 i. SureSelect Block n.º 1 (Agilent) - 2,5 µl
 - ii. SureSelect Block n.º 2 (Agilent) - 2,5 µl
 - 20 iii. Paired End primer Block 1.0 (IDT, resuspendido a 200 uM con H₂O) - 1,5 µl
 - iv. Index Primer Block 1-12 (IDT, resuspendido a 200 uM con H₂O) - 1,5 µl
- c. Dilución del RNase Block
 - 25 i. Para cebos de ARN biotinilados personalizados con territorio <3 Mb: se diluyó 1 µl de RNase Block (Agilent) en 9 µl de agua.
 - ii. Para cebos personalizados con un territorio de cebo >3 Mb: se diluyó 1 µl de RNase Block en 3 µl de agua (todavía 0,5 µl de RNase Block por 7 µl de reacción de captura)
 - 30 d. Mezcla de cebos: (7 µl por reacción)
 - i. Cebos de ARN - 2 µl (para cebos que tienen un territorio de cebo >3 Mb, se usaron 5 µl de cebo)
 - 35 ii. RNase Block diluido - 5 µl (para cebos que tienen un territorio de cebo >3 Mb, se usaron 2 µl de RNase Block diluido como se indica anteriormente)

Después de que se prepararan la mezcla de tampón de hibridación, la mezcla de bloqueo y la(s) mezcla(s) de cebo, se agitó en vórtex la mezcla de tampón de hibridación, se centrifugó y se calentó a 65 °C en el bloque térmico. Se mezclaron 4 µl de cada colección de muestras agrupadas para la selección de híbridos con 8 µl de la mezcla de bloqueo en una placa de PCR de 96 pocillos. Se incubó la reacción en un termociclador a 95 °C durante 5 minutos y, a continuación, se mantuvo a 65 °C. Cuando las colecciones de muestras agrupadas/mezcla de bloqueo se habían incubado a 95 °C durante 5 minutos y, a continuación, a 65 °C durante 2,5 minutos, se colocó la mezcla de cebos (= mezcla de cebo/bloqueo RNase) en el bloque térmico a 65 °C durante 2,5 minutos. Se centrifugaron rápidamente los tubos que contenían el tampón de hibridación y, a continuación, se devolvieron inmediatamente al bloque térmico a 65 °C. Se pipetearon 13 µl de la mezcla de tampón de hibridación calentada en cada muestra de mezcla de colección/bloqueo mientras la placa de 96 pocillos permanecía en el termociclador a 65 °C. Después de que la mezcla de cebos se incubara durante 2,5 minutos a 65 °C, se añadieron 7 µl de la mezcla de cebos a cada muestra de mezcla colección/bloqueo/tampón de hibridación mientras la placa de 96 pocillos permanecía en el termociclador a 65 °C. Se incubó la reacción (el volumen total era de 32 µl) a 65 °C durante 24 horas en un termociclador.

Preparación de las microesferas magnéticas

- 55 El tampón de lavado SureSelect n.º 2 se precalentó a 65 °C en el bloque térmico. Se agitaron en vórtex las microesferas Dynal MyOne Streptavidin T1 (Invitrogen) y se resuspendieron. Se lavaron las microesferas añadiendo 200 µl de tampón de unión SureSelect por 50 µl de microesferas Dynal (por ejemplo, se usaron 1200 µl de tampón de unión SureSelect para preparar 300 µl de microesferas Dynal). Se agitaron en vórtex las microesferas durante 5 segundos y se centrifugaron brevemente. Se colocaron las microesferas sobre un soporte magnético durante aproximadamente 15 segundos o hasta que se capturaron todas las microesferas. Se retiró el sobrenadante y se desechó. Se repitió el lavado con tampón de unión SureSelect dos veces más para un total de tres lavados. Después del lavado, se resuspendieron las microesferas en 200 µl de tampón de unión SureSelect por 50 µl de microesferas Dynal (por ejemplo, se usaron 1200 µl de tampón de unión SureSelect para preparar 300 µl de microesferas Dynal). Se agitaron en vórtex las microesferas resuspendidas y se centrifugaron brevemente. Se dividieron las microesferas resuspendidas en alícuotas de 200 µl y se añadieron a tubos de microcentrífuga individuales de 1,5 ml.
- 60
- 65

Selección del ADN capturado en híbridos

Después de 24 horas de incubación, cada muestra hibridada de la placa de PCR en el termociclador a 65 °C se pipeteó rápidamente en un tubo que contenía 200 µl de microesferas preparadas a temperatura ambiente. Se agitaron en vórtex las mezclas de muestra y microesferas durante 5 segundos y se incubaron en un rotador a temperatura ambiente durante 30 minutos, para garantizar una mezcla adecuada. A continuación, se centrifugaron rápidamente los tubos. Se capturaron las microesferas con un imán (durante 2 minutos) y se retiró y se descartó el sobrenadante. Se resuspendieron las microesferas en 500 µl de tampón de lavado SureSelect n.º 1, para un lavado de baja rigurosidad. Se agitaron en vórtex las muestras durante 5 segundos y se incubaron durante 15 minutos a temperatura ambiente fuera del imán. Se agitaron en vórtex las muestras durante 5 segundos cada 3-5 minutos. Se centrifugaron rápidamente los tubos. A continuación, se capturaron las microesferas en un soporte magnético durante 2 minutos y se retiró y se descartó el sobrenadante. Para un lavado de alta rigurosidad para retirar el material no diana, se lavaron las microesferas con tampón de lavado SureSelect n.º 2 precalentado a 65 °C. En resumen, se resuspendieron las microesferas en 500 µl de tampón de lavado SureSelect n.º 2 precalentado y se mezclaron en un agitador vórtex durante 5 segundos para resuspender las microesferas. Se centrifugaron brevemente las microesferas en una centrífuga y se incubaron a 65 °C durante 10 min en un bloque térmico con mezclado por agitación en vórtex ocasional durante 5 segundos a temperatura ambiente. A continuación, se centrifugaron brevemente las microesferas en una centrífuga y se capturaron con un imán durante 2 minutos. Se repitió el lavado 2 veces más con tampón de lavado SureSelect n.º 2 precalentado a 65 °C para un total de tres lavados. A continuación, se retiró por completo el tampón de lavado y se añadieron 50 µl de tampón de elución SureSelect a las microesferas, seguido de agitación en vórtex durante 5 segundos para mezclar las microesferas. Se incubaron las muestras durante 10 minutos a temperatura ambiente con mezclado con agitación en vórtex ocasional durante 5 segundos. Se centrifugaron brevemente las microesferas en una centrífuga y se capturaron con un soporte magnético. Se pipeteó el sobrenadante que contenía el ADN capturado a un tubo nuevo de microcentrífuga de 1,5 ml. Se añadieron 50 µl de tampón de neutralización SureSelect al ADN capturado. Se agitaron en vórtex las muestras durante 5 segundos, se centrifugaron brevemente en una centrífuga y se purificaron usando 1,8 volúmenes de microesferas AMPureXP. Se eluyó el ADN en 40 µl de agua libre de nucleasas.

Enriquecimiento por PCR del ADN capturado

Se descongelaron los reactivos de PCR y se preparó una mezcla maestra de PCR sobre hielo. Para 60 µl de mezcla maestra por muestra, se mezclaron 50 µl de mezcla maestra 2X Phusion High Fidelity con tampón HF (NEB, n.º F-531S) con 8 µl de agua libre de nucleasas, 1 µl de QPCR Primer1.1 (100 µM en H₂O) y 1 µl de QPCR Primer2.1 (100 µM en H₂O). Las secuencias de cebadores para Q-PCR son:

QPCR Primer1.1 (purificada por HPLC a partir de IDT):

5' AATGATACGGCGACCACCGAGAT3' (SEQ ID NO: 2)

QPCR Primer2.1 (purificada por HPLC a partir de IDT):

5' CAAGCAGAAGACGGCATACGA3' (SEQ ID NO: 3)

Se añadieron 60 µl de mezcla maestra a 40 µl de cada muestra de ADN capturado purificado en una placa de PCR de 96 pocillos. Se incubó la reacción en un termociclador como sigue:

1 ciclo	98 °C	30 s
12 ciclos	98 °C	10 s
	65 °C	30 s
	72 °C	30 s
1 ciclo	72 °C	5 min
	4 °C	Mantener

Cada 100 µl de reacción de PCR se purificó con 1,8 volúmenes de microesferas AMPureXP y se eluyeron en 35 µl de tampón de elución (Tris 10 mM, pH 8.5). Se cuantificaron las muestras de ADN seleccionado/capturado en híbridos usando un ensayo Q-PCR. El ensayo Q-PCR detectó los adaptadores terminales y las lecturas indicaron

cuánto de cada muestra se debía cargar en una cubeta de lectura de secuenciación para obtener la densidad de agrupación apropiada.

EJEMPLO 8: procedimientos de secuenciación

Lo siguiente ejemplifica determinadas implementaciones de los procedimientos y condiciones experimentales usados para identificar las alteraciones de acuerdo con los ejemplos. Se puede realizar un cribado de translocación adicional usando, por ejemplo, un análisis qRT-PCR de ADNc preparado a partir de una muestra tumoral preseleccionada.

Tanto la secuenciación de ADN (ADN-seq) como la secuenciación de ARN (ARN-seq) se realizaron para identificar alteraciones genéticas usando ácidos nucleicos aislados de muestras FFIP, de sangre y de aspirado de médula ósea. Por ejemplo, se realizó secuenciación masiva en paralelo de ADN en colecciones basadas en ligación de adaptadores capturadas por hibridación usando ADN aislado de tejido fijado e incluido en parafina de archivo. Se usó una combinación de herramientas de análisis para analizar los datos y asignar las llamadas de alteración del ADN. Se realizó un cribado de translocación adicional usando el análisis qRT-PCR de ADNc preparado a partir de tumores congelados o bien la evaluación IHC de muestras FFIP de archivo. Se realizó una secuenciación masiva en paralelo de ADNc para confirmar la expresión de ambas translocaciones novedosas usando ARN aislado de tejido FFIP. Se secuenció el ADN genómico de referencia normal emparejado procedente de la sangre para el paciente índice con CPNM para confirmar el origen somático del reordenamiento.

Secuenciación de ADN genómico

La secuenciación de 2574 exones de 145 genes de cáncer se realizó usando ADN de muestras tumorales fijadas con formol e incluidas en parafina (FFIP) de archivo; 24 de pacientes con CPNM.

Las colecciones de secuenciación se construyeron mediante el procedimiento de ligación de adaptadores usando ADN genómico seguido de selección de hibridación con sondas de captura de hibridación de ARN optimizadas (kit personalizado Agilent SureSelect). La secuenciación en el instrumento HiSeq2000 (Illumina) se realizó usando 36x36 lecturas emparejadas a una profundidad promedio de 253X. El procesamiento de datos y las asignaciones de mutaciones para sustituciones de bases, indels, alteraciones en el número de copias y reordenamientos genómicos se realizaron usando una combinación de herramientas optimizadas para la llamada de mutaciones en tejido tumoral.

EJEMPLO 9: genes y variantes seleccionados ejemplares para análisis múltiple

Los genes, variantes y tipos de cáncer ejemplares seleccionados para el análisis múltiple se describen en las **tablas 1-9**.

EJEMPLO 10: un enfoque bayesiano para la detección sensible de alteraciones genómicas somáticas a partir de la secuenciación de nueva generación de muestras clínicas de cáncer

El enfoque bayesiano descrito en el presente documento se implementó en los siguientes ejemplos.

La utilidad de este enfoque se ilustra mediante cálculos de potencia que describen el impacto de los antecedentes basados en datos en la detección de sustituciones en el intervalo inferior de frecuencias de mutación relevantes en el entorno clínico. Como se muestra en la **FIG. 2**, los valores de expectativa previa (por ejemplo, 1e-6 o 10 % previo) y frecuencia de mutación (por ejemplo, 1 %, 5 % o 15 % de mutación) corresponden a los valores descritos en (i) y (ii) de "A Bayesian Approach for Sensitive Detection of Somatic Genomic Alterations from Next-generation Sequencing of Clinical Cancer Specimens", respectivamente. La **FIG. 2** muestra que la incorporación de expectativas previas puede mejorar la potencia de detección de mutaciones más raras, por ejemplo, reduciendo la profundidad de cobertura requerida en los sitios mutados, o incrementando la potencia estimada (sensibilidad) para detectar mutaciones.

En la publicación de solicitud de patente internacional n.º WO2012/092426 se divulgan ejemplos adicionales que muestran las aplicaciones del enfoque bayesiano a diversas muestras (por ejemplo, una muestra multiclonal de baja pureza construida, muestras de tumores de pulmón y colon y muestras de cáncer de mama) y para detectar mutaciones infrecuentes.

EJEMPLO 11: selección de alto rendimiento de dianas basada en solución usando sondas de captura de oligonucleótidos sintetizadas individualmente

La disponibilidad de técnicas de selección de dianas genómicas basada en solución ha permitido el rápido desarrollo de aplicaciones de secuenciación dirigidas, algunas de las cuales han dado lugar a la introducción de pruebas clínicas de secuenciación. Los reactivos de captura de hibridación comercializados se basan en oligonucleótidos sintetizados en matriz, que se convierten en sondas de ADN o ARN biotinilado ("cebos"). Sin

embargo, los procedimientos para generar estos agrupamientos complejos de sondas se enfrentan a desafíos de rendimiento, por ejemplo, la captura de dianas con alto contenido de GC.

En el presente documento se describe un enfoque alternativo que usa oligonucleótidos 5'-biotinilados sintetizados individualmente ("oligo-cebos") para capturar una región diana de ~130 kb que representa 57 genes relacionados con el cáncer clínicamente relevantes y aprovechables. Las colecciones de secuenciación indexadas seleccionadas usando estos oligo-cebos con un procedimiento de hibridación de 24 horas proporcionaron un enriquecimiento de diana de 5000 veces. 50M 49x49 lecturas de extremos emparejados generaron una cobertura de diana promedio de 2100x con una desviación estándar de 568x (27 %). Todas las dianas se cubrieron con éxito, con un 99,95 % de las bases diana cubiertas a >500x. Además, la cobertura de diana no tenía prácticamente ningún sesgo de GC. Las dianas con contenido de GC >70 % promediaron una cobertura de 1975x, y las dianas con contenido de GC <35 % promediaron una cobertura de 1996x.

El alto rendimiento se mantuvo usando tiempos de hibridación incluso más cortos: el 99,3 % de las bases diana se cubrieron a >500x después de una hibridación de 2,5 horas.

Uso de SSPE (esperma de salmón, Pe)/solución de Denhardt dio mejores resultados que los tampones de hibridación/lavado que contienen TEACI, TMACI y/o sulfato de dextrano.

Los oligo-cebos se pueden añadir a agrupamientos de cebos derivados de matrices para incrementar la cobertura de regiones que de otro modo serían difíciles de capturar (por ejemplo, alto % de GC), o para añadir rápidamente nuevo contenido génico. Este enfoque ofrece un procedimiento altamente eficaz y escalable para desarrollar pruebas clínicas de secuenciación dirigidas de alto rendimiento.

Los procedimientos ejemplares para optimizar los cebos de captura y las condiciones experimentales para evaluar los conjuntos de cebos se divulgan en la publicación de solicitud de patente internacional n.º WO2012/092426.

EJEMPLO 12: secuenciación rutinaria y ultraprofunda para el análisis de tumores sensibles usando bajas entradas de ADN derivado de tejido fijado con formol

La amplia adopción de tecnologías de secuenciación de ADN de alto rendimiento ha facilitado rápidos avances en la genómica del cáncer. Sin embargo, el procedimiento habitual en el diagnóstico genómico del cáncer todavía implica la realización de pruebas centradas en genes individuales y mutaciones específicas. A medida que crece el número de mutaciones clínicamente aprovechables, este paradigma de una única mutación por prueba se vuelve inviable, en particular cuando las muestras de tejido son limitadas, como es en general el caso con las biopsias. Para abordar la necesidad clínica de un análisis genómico exhaustivo de muestras tumorales, los inventores han desarrollado una prueba clínica que proporciona datos de secuencias masivas en paralelo para más de 200 genes relacionados con el cáncer. Además, esta prueba demostró ser clínicamente relevante, produciendo datos de secuenciación ultraprofunda a partir de muestras de tejido fijadas con formol e incluidas en parafina (FFIP) con entradas de ADN de tan solo 50 ng, y de muestras almacenadas durante hasta once años.

Para evaluar el rendimiento de esta prueba en una amplia variedad de muestras, se aisló ADN de 96 muestras FFIP de un conjunto de bloques almacenados, que incluía 12 pares de tejido tumoral/normal de cada uno de tejidos de mama, colon, pulmón y riñón distribuidos uniformemente para cada tejido con los siguientes periodos de almacenamiento: 1, 3, 5, 7, 9 y 11 años. Se usaron 200 ng y/o 50 ng de ADN de entrada para construir colecciones de secuenciación indexadas, que a continuación se enriquecieron para más de 200 genes relacionados con el cáncer usando un procedimiento de captura de híbridos basado en solución y se secuenciaron en una plataforma Illumina HiSeq™ 2000.

Para las 76 muestras que proporcionaron al menos 200 ng de ADN para la construcción de la colección, la cobertura de secuenciación promedió 1000x después de la retirada de los duplicados de PCR, con >95 % de las muestras proporcionando una mediana de cobertura de >350x. Para las muestras en las que se usaron 50 ng para la construcción de colecciones, la cobertura promedió 450x. El rendimiento de la secuenciación fue consecuente en todos los tipos de tejido de muestra y periodos de almacenamiento. Dicha secuenciación ultraprofunda permite la detección con alta confianza de mutaciones presentes a frecuencias de tan solo 5-10 %.

EJEMPLO 13: análisis del genoma tumoral usando células tumorales circulantes

Las células tumorales circulantes (CTC) proporcionan una oportunidad única para tomar muestras de neoplasias malignas humanas de una manera en serie mínimamente invasiva. El uso de CTC para la caracterización molecular de los genomas del cáncer presenta dos desafíos clave. En primer lugar, las CTC se deben aislar eficazmente de la sangre, donde pueden estar en un número 10^7 veces menor que las células no tumorales. En segundo lugar, el número limitado de genomas tumorales presentes en una muestra de CTC se debe capturar de forma accesible mientras se minimiza la pérdida de material y la introducción de sesgo.

Los análisis genéticos de CTC previos han usado PCR específica de alelos; estos procedimientos permiten la

detección de números de copias muy bajos de mutaciones específicas en un fondo de secuencias naturales $\geq 10^4$ veces superior. Si bien aborda los dos desafíos de abundancia de CTC y eficacia de captura, este enfoque se limita intrínsecamente a caracterizaciones limitadas de variantes seleccionadas y preespecificadas. Para llevar el análisis molecular de CTC a la era genómica, los inventores han acoplado un sistema de captura microfluidica de células raras que permite recuperar las CTC con un fondo de solo cientos, en lugar de decenas de miles, de glóbulos blancos, con una plataforma de nueva generación que permite una resecuenciación profunda de más de 200 genes asociados al cáncer a partir de una única muestra de CTC.

Usando mezclas complejas de hasta diez líneas de células cancerosas, permite la detección de mutaciones sensibles (~ 94 % para abundancia de alelos ≥ 10 %) a partir de tan solo 100 células totales, al tiempo que conserva en gran medida las frecuencias de alelos ($R^2 \sim 0,90$). Al recapturar las células cultivadas añadidas a la sangre completa, se obtuvieron perfiles de mutación multigénica de muestras que contenían tan solo diez células cancerosas. Este nivel de sensibilidad hace que la mayoría de las muestras clínicas de CTC estén al alcance del análisis NGS. En una serie de muestras de sangre de pacientes con cáncer de mama se investigó la posible heterogeneidad de las CTC comparando la frecuencia de células positivas para Her2Neu con la abundancia relativa de ADN positivo para mutaciones somáticas.

EJEMPLO 14: detección de mutaciones, translocaciones y cambios en la expresión génica asociados al cáncer a través de la integración de la secuenciación profunda de ADN y ARN dirigida de muestras tumorales FFIP

La aplicación generalizada de la terapia personalizada al cáncer requiere una caracterización exhaustiva, sensible y oportuna de las diversas anomalías presentes en el genoma y el transcriptoma de un tumor. El ARN y el ADN de la mayoría de las muestras clínicas de cáncer, comúnmente almacenados como bloques fijados con formol e incluidos en parafina (FFIP), son de mala calidad y han sido difíciles de usar para el análisis molecular. Los ensayos emergentes de secuenciación de ADN de nueva generación funcionan bien con el ADN dañado y son suficientemente sensibles para detectar muchos tipos de anomalías genómicas. Actualmente, no existe un protocolo de secuenciación de ARN comparable para el análisis exhaustivo del transcriptoma de muestras tumorales FFIP.

Resultados:

Se desarrolló un procedimiento de secuenciación y análisis de ARN dirigido compatible con FFIP para la detección sensible de mutaciones, reordenamientos y cambios de expresión en más de 200 genes asociados al cáncer. Los protocolos se validaron en ARN de línea celular y se usaron para estudiar más de 50 tumores FFIP de cáncer de pulmón no microcítico (CPNM). Se detectaron mutaciones y fusiones génicas conocidas (por ejemplo, BCR-ABL1) en líneas celulares. La reproducibilidad técnica en el análisis de expresión digital superó $R^2 = 0,99$ y $>0,9$ para líneas celulares y ARN FFIP, respectivamente. Como se esperaba en los genomas del cáncer, el ARN-seq proporcionó pruebas de anomalías en el genoma, incluyendo mutaciones puntuales y reordenamientos novedosos que involucran oncogenes conocidos. Se reveló una expresión diferencial altamente significativa de oncogenes que incluyen EGFR, FGFR3, CDH5, KIT y RET, que varía de 2,5 a 70 veces en diferentes tumores. La combinación de datos de secuenciación de ARN y ADN en muestras FFIP idénticas corroboró las consecuencias funcionales de las alteraciones genómicas; los ejemplos incluyeron la expresión de alelos TP53 mutados y la expresión reducida de STK11 en un tumor que presentaba pérdida de heterocigosidad a nivel de ADN. Se prevé que la aplicación de tecnologías de secuenciación de nueva generación al ARN FFIP y la integración con los procedimientos de secuenciación de ADN existentes amplíen la comprensión de la biología del cáncer clínicamente relevante y mejoren la atención al paciente.

Procedimientos:

El ARN se extrae de secciones de tejido FFIP, típicamente 1 o 2 rizos de 10 μm , usando el kit High Pure Paraffin de Roche de acuerdo con las instrucciones del fabricante. El ARN extraído se almacena a -80 °C. El rendimiento y la calidad del ARN se evalúan mediante RiboGreen (Invitrogen) y un Bioanalyzer RNA Pico Chip (Agilent), respectivamente, de acuerdo con las instrucciones del fabricante. Los rendimientos típicos están entre 500 ng y 2 μg con una puntuación RIN de menos de 4.

La primera cadena de ADN complementario (ADNc) se produce a partir de entre 100 y 600 ng de ARN FFIP en una reacción de 20 μl usando SuperScript III (Invitrogen) de acuerdo con el protocolo del fabricante, con 550 pmoles de hexámero aleatorio como cebador. La síntesis de la segunda cadena, para generar un ADNc completamente bicatenario, se realiza inmediatamente después de la síntesis de la primera cadena mediante la adición de 60 μl de una mezcla maestra de NEBNext Second Strand Synthesis Module (New England Biolabs) y la incubación durante 150 minutos a 16 °C de acuerdo con el protocolo del fabricante. La calidad y el rendimiento del ADNc bicatenario se pueden evaluar usando PicoGreen (Invitrogen) y un Bioanalyzer High Sensitivity Chip (Agilent), respectivamente. En general, el rendimiento de la síntesis de ADNc completo se usa como entrada para el protocolo de construcción de la colección FMI estándar.

La construcción de una colección de secuenciación compatible con extremos emparejados y la posterior selección de híbridos y secuenciación del ADNc generado a partir del ARN FFIP se realiza usando protocolos similares a los empleados para ADN FFIP descrito en el presente documento, pero comenzando directamente en la etapa de reparación de extremos, ya que la naturaleza altamente fragmentada del ARN FFIP elimina la necesidad de corte.

El análisis de los datos de secuenciación del ARN FFIP se puede realizar usando procedimientos conocidos en la técnica. Por ejemplo, el análisis de los datos de secuenciación del ARN FFIP se puede realizar mediante el mapeo de todos los pares de lecturas para una secuencia del genoma de referencia (hg19) y/o un transcriptoma de referencia (todas las secuencias de transcritos conocidos, por ejemplo, RefSeq). Las lecturas mapeadas se usan a continuación para identificar una fusión génica, mutaciones en las secuencias génicas, un corte y empalme alternativo y para cuantificar la expresión génica como se describe en la literatura, por ejemplo, por Berger *et al.* (2010) *Genome Res.* 20(4):413-27 (PMID 20179022) y Garber *et al.* (2011) *Nat Methods.* 8(6):469-77 (PMID 21623353). Como lo demuestran Levin *et al.* (2009) *Genome Biol.* 10(10):R115 (PMID 19835606), se puede emplear ARN-seq dirigida para mejorar la detección de mutaciones y el descubrimiento de fusiones en un conjunto seleccionado de genes, y conserva la información cuantitativa para el análisis de expresión.

EJEMPLO 15: llamada de mutaciones sensible y exacta mediante secuenciación ultraprofunda de muestras tumorales clínicas

El rápido avance en la comprensión de la genómica del cáncer y el creciente número de terapias dirigidas disponibles proporcionan oportunidades cada vez mayores para un tratamiento eficaz del cáncer basado en un análisis tumoral exhaustivo. Aunque se ha logrado un progreso significativo en los enfoques experimentales y computacionales para analizar los genomas tumorales mediante la secuenciación de nueva generación en el entorno de investigación, la extensión de estas técnicas al entorno clínico plantea desafíos adicionales significativos. Una de las claves es la limitada pureza y la heterogeneidad de las muestras clínicas, junto con el requisito de proporcionar una alta sensibilidad y exactitud para una amplia gama de mutaciones potencialmente aprovechables clínicamente.

Para abordar este desafío, los inventores han desarrollado una prueba clínica que puede generar datos de secuenciación ultraprofunda (>700x) para más de 200 genes relacionados con el cáncer a partir de muestras tumorales FFIP de rutina, y herramientas computacionales que pueden explotar esta profundidad para proporcionar altos niveles de sensibilidad y exactitud para diferentes tipos de mutaciones presentes en fracciones bajas. El pipeline analítico de los inventores detecta variantes cortas en los datos de secuencias mapeados que representan frecuencias de mutación conocidas, y combina la detección de puntos de ruptura y el ensamblaje local para identificar inserciones y deleciones más grandes, que a menudo se pierden en procedimientos alternativos. Además, se identifican alteraciones en el número de copias y reordenamientos que involucran genes clave del cáncer.

Para validar el rendimiento analítico de los procedimientos recientemente desarrollados, los inventores diseñaron e implementaron un extenso estudio de mezclas de muestras como modelo para acontecimientos raros en ADN heterogéneo, incluyendo 20 líneas celulares HapMap normales y 28 líneas de células cancerosas caracterizadas individualmente. Los inventores informan de un 100 % de sensibilidad para las sustituciones y >90 % de sensibilidad para indels de 1-50 pb de longitud presentes en >10 % de una mezcla, ambos con PPV >99 %. La aplicación de la prueba a una cohorte de 227 muestras de melanoma, cáncer de próstata, mama, colon y pulmón reveló 427 mutaciones oncoiniciadoras somáticas conocidas y probables, el 40 % de las cuales estaban presentes en fracciones de muestra por debajo del 20 % y el 18 % por debajo del 10 %, lo que subraya la importancia de la llamada de mutaciones sensibles.

Los procedimientos para la detección de mutaciones de cáncer en los márgenes quirúrgicos se describen en la publicación de solicitud de patente internacional n.º WO 2012/092426.

EJEMPLO 16: análisis exhaustivo de secuencias de inmunoglobulina usando secuenciación de nueva generación basada en captura de híbridos en neoplasias hematológicas de linfocitos B

La secuenciación de los genes que codifican las inmunoglobulinas es crítica para detectar poblaciones de células clonales, así como para determinar el pronóstico y las decisiones terapéuticas en pacientes con neoplasias malignas linfáticas, incluyendo las leucemias de linfocitos B, los linfomas y el mieloma múltiple. Los ensayos actuales para identificar la secuencia de inmunoglobulina reordenada en neoplasias malignas de linfocitos B se basan en general en la amplificación basada en PCR específica de secuencia de regiones de inmunoglobulina (IG) conservadas. Este ejemplo describe un enfoque novedoso basado en la captura de híbridos para secuenciar las cadenas de inmunoglobulina que permite la identificación de los dominios variables de la cadena pesada y ligera, las secuencias de la región determinante de complementariedad (CDR) y el estado de hipermutación somática (HMS).

Las secuencias de inmunoglobulina (IG) de línea germinal que incluyen segmentos variables (V), de diversidad (D), de acoplamiento (J) y de clase abarcan más de 1 Mb en el cromosoma 14 para la cadena pesada, más de

344 kb en el cromosoma 22 para la cadena ligera lambda y más de 477 kb en el cromosoma 2 para la cadena ligera kappa. Durante la recombinación VDJ durante el desarrollo normal de linfocitos B, se crea una única secuencia V-D-J (~320 bases). Se sometieron a prueba dos estrategias principales. La primera implicaba cebar solo los segmentos D, J y de clase. Se demostró que esta estrategia tiene éxito en ARN. La segunda estrategia implicaba cebar de forma adicional o aislada los segmentos V. Esta estrategia permitió la identificación de la secuencia de IGHV en ARN y ADN.

Procedimientos:

El ARN y el ADN se extrajeron con éxito de 60 muestras, incluyendo 7 líneas celulares de linfoma de células del manto (LCM) y 53 aspirados clínicos de médula ósea de leucemia linfocítica crónica (LLC). Se capturaron colecciones de secuenciación de ADN y ADNc ligadas a adaptador mediante hibridación en solución usando conjuntos de cebos personalizados dirigidos a los segmentos variables, de acoplamiento y de clase de inmunoglobulina. Todas las colecciones capturadas se secuenciaron a alta profundidad (Illumina HiSeq) en un laboratorio con certificación CLIA (Foundation Medicine).

Las lecturas de secuenciación de extremos emparejados están a menudo en la misma escala de longitud (~50 bases) que las regiones repetitivas del dominio variable de inmunoglobulina. Las regiones estructurales y CDR de los dominios V individuales tienen homología conocida entre sí. Por tanto, se podría mapear potencialmente una lectura de 50 pares de bases en una región estructural para varios segmentos variables diferentes. Además, la secuencia de referencia Hg19 usada en la alineación no contiene todas las secuencias de segmentos variables de línea germinal conocidas. Por ejemplo, existen cuatro alelos IGHV3-21*01, IGHV3-21*02, IGHV3-21*03, IGHV3-21*04 pero Hg19 contiene IGHV3-21*01, es decir, el primer alelo. El procedimiento de análisis descrito a continuación se diseñó para resolver estos problemas.

El análisis para llamar al clon y la HMS se realizó usando un programa descrito a continuación. El programa incluye, por ejemplo, las siguientes etapas: (1) seleccionar lecturas químéricas para los segmentos de clase y de acoplamiento de IG; (2) filtrar esas lecturas para las lecturas que van a los segmentos IGHV, IGKV o IGLV; (3) construir la secuencia consenso del dominio V para cada segmento V candidato en la etapa (2) (por ejemplo, tomar cada lectura de 50 pares de bases de las lecturas mapeadas para construir una secuencia de ~300 pares de bases de longitud); y (4) ejecutar la secuencia construida en la etapa (3) a través del programa IgBlast (esto identifica el dominio V y da un cálculo del % de HMS (www.ncbi.nlm.nih.gov/igblast/)). IgBlast es una herramienta estándar de la industria bien conocida y validada para la alineación de la cadena IG secuenciada con el VDJ de emparejamiento de línea germinal más cercano en la base de datos de secuencias IGMT, permitiendo el cálculo del % de HMS y la identificación de la secuencia CDR3.

Se realizó una revisión manual de cada candidato para confirmar la llamada automatizada descrita anteriormente. La revisión manual (que también se podría automatizar) incluye, por ejemplo, las siguientes etapas: (1) clasificar la salida de IgBlast para la longitud de alineación (por ejemplo, sin considerar ninguna alineación por debajo de 200 bases de longitud; y (2) verificar, por ejemplo, que las secuencias CDR3 son casi idénticas para un clon dado, los SNP son casi idénticos para una agrupación de lecturas dada, las lecturas mapeadas son en gran medida únicas y las lecturas mapean toda la región FR1 a CDR3 sin huecos.

Como es probable que una célula tumoral que proviene de un linfocito B maduro tenga segmentos variables y de acoplamiento reordenados, este análisis dio una señal fuerte incluso si la pureza del tumor era relativamente baja (por ejemplo, aproximadamente un 20 % de pureza).

Resultados:

El enfoque basado en la captura se validó usando 7 líneas celulares de LCM y 53 muestras de LLC analizadas usando un ensayo comercial basado en PCR con certificación CLIA (Invivoscribe). Las secuencias de inmunoglobulina derivadas de las 7 líneas celulares de LCM fueron 100 % concordantes con la identificación del dominio variable de la cadena pesada y ligera publicada y el porcentaje de HMS. La comparación con 53 muestras previamente sometidas a pruebas clínicas mostró una concordancia del 98 % (52/53) para la identificación de la presencia de una población clonal y una concordancia del 100 % (39/39) para la identificación del dominio IGHV. La comparación con 50 muestras de LLC previamente analizadas para HMS dio como resultado una concordancia global del 96 % (48/50) con una concordancia del 94 % (31/33) para no HMS y una concordancia del 100 % (17/17) para la presencia de HMS. Además, se identificaron clones secundarios en 11 muestras.

Se usaron muestras de sangre sana y muestras de sangre enriquecida como controles negativos. Estas muestras deben tener reordenamientos VDJ pero no poblaciones monoclonales. En las 3 muestras de sangre completa normal no se identificaron clones tumorales. La población de linfocitos B enriquecida tenía llamadas de agrupaciones IGHV. Todos se consideraron policlonales y no se habrían denominado clones tumorales. Existió un 100 % de concordancia con las 6 muestras de clones tumorales negativos.

Conclusiones:

Este estudio demostró que se puede usar la secuenciación de ADN y ARN dirigida basada en la captura de híbridos para caracterizar exhaustivamente la secuencia de inmunoglobulina de poblaciones de linfocitos B tumorales clonales. Esta capacidad permite la cuantificación de HMS y la identificación del dominio variable, la secuencia CDR3 y la restricción de clase. La integración de esta metodología con enfoques de análisis genómico exhaustivo amplía la utilidad clínica de dichos ensayos en pacientes con neoplasias malignas hematológicas y puede proporcionar información importante sobre la inmunooncología y la respuesta de los pacientes a las inmunoterapias, incluyendo los pacientes con tumores sólidos.

EJEMPLO 17: análisis multigénico para neoplasias malignas hematológicas usando secuenciación de ADN y secuenciación de ARN dirigidas

Este ejemplo describe la detección altamente sensible y específica de diversas clases de alteraciones genómicas en condiciones clínicamente relevantes mediante la combinación de secuenciación de ADN (ADN-seq) y secuenciación de ARN (ARN-seq) dirigidas en el ADN genómico y el ARN total extraídos de una serie de muestras de neoplasias malignas hematológicas. Las alteraciones genómicas sometidas a prueba incluyen, por ejemplo, sustituciones de bases (Sub), inserciones y deleciones (Indel), alteraciones en el número de copias (ANC) focales y reordenamientos genómicos. Se examinaron muestras de neoplasias malignas hematológicas típicas, por ejemplo, FFIP, sangre y aspirados de médula ósea.

La prueba descrita en este ejemplo incluye un conjunto de cebos de ADN-seq e incorpora un componente de ARN-seq dirigida. El componente de ARN-seq dirigida utiliza ARN convertido en ADN complementario (ADNc) en lugar de ADNg como entrada. En comparación con la ADN-seq sola, un ensayo integrado que combina ADN-seq y ARN-seq ofrece una sensibilidad mejorada para la detección de reordenamientos genómicos.

Los reordenamientos genómicos incluyen una clase heterogénea de alteraciones que son el resultado, por ejemplo, del acoplamiento de dos o más regiones diferentes del genoma humano para producir un producto no de referencia, por ejemplo, en comparación con el ensamblaje, la secuencia y la estructura del genoma humano de referencia hg19. Como resultado de la secuencia genómica reordenada, también se pueden transcribir transcritos de ARN de secuencia no de referencia. Los reordenamientos genómicos se pueden clasificar como que se producen fuera de los genes de la secuencia de referencia (RefSeq) (es decir, extragénicos) o dentro de los intrones y exones de los genes (es decir, intragénicos). Por ejemplo, los reordenamientos extragénicos, tales como los acontecimientos asociados a promotor o potenciador que implican MYC e IGH, pueden afectar a la regulación de la expresión de un transcrito de ARNm dado, o vía, pero no alteran necesariamente la secuencia de ARNm. Los reordenamientos intragénicos implican los exones y/o intrones de dos o más genes y pueden proporcionar transcritos de ARNm quiméricos compuestos de una secuencia derivada de diferentes genes y/o regiones del genoma (por ejemplo, diferentes cromosomas, diferentes hebras del mismo cromosoma). Se pueden crear fusiones génicas mediante reordenamientos genómicos que se producen en intrones específicos reordenados de forma recurrente a partir de 2 o más genes. Los intrones pueden ser difíciles de cubrir por ADN-seq a una profundidad suficiente debido a las restricciones de tamaño (>1 kb) y/o secuencia (por ejemplo, estructura de repetición). Además, es posible que no se conozcan los puntos de ruptura relevantes o que el transcrito de ARNm de fusión génica pueda resultar de un reordenamiento complejo de modo que puede que no sea posible identificar los socios de fusión relevantes mediante ADN-seq. La ARN-seq mejora la sensibilidad de detección al caracterizar directamente los productos de ARNm de las fusiones génicas, en lugar de depender de la secuenciación de intrones y la extrapolación de la estructura del ARNm a partir de los datos de secuencia de ADN.

Como tal, un conjunto de cebos de ADN-seq ejemplar incluye la cobertura exónica completa de 464 genes (**tabla 1**), la cobertura de intrones seleccionados en 31 genes reordenados de forma recurrente en el cáncer (**tabla 2**) y la cobertura exónica completa de 91 genes adicionales (**tabla 4**); y un componente de ARN-seq ejemplar usa un conjunto de cebos de ARN-seq para dirigirse a exones de 333 genes que son conocidos por estar reordenados de forma recurrente en neoplasias malignas hematológicas (**tabla 3**).

Preparación de reactivos de cebado para ARN-seq y ADN-seq

Se diseñaron secuencias de cebos dirigidas a los genes enumerados en las **tablas 1-4**. Los cebos se sintetizaron en Integrated DNA Technology (IDT) usando su química de síntesis Ultramer. Se preparó una mezcla equimolar de los subgrupos de conjuntos de cebos apropiados recibidos de IDT para crear los reactivos de cebado de ARN-seq y ADN-seq. A continuación, los reactivos de cebado se dividieron en alícuotas de un solo uso y se almacenaron a -20 °C. Se confirmó el rendimiento del conjunto de cebos de ADN-seq mediante la captura de una colección HapMap usando un protocolo de captura de híbridos automatizado estándar. Se confirmó el rendimiento del conjunto de cebos de ARN-seq mediante la captura de una colección HapMap y colecciones de ARN de referencia humano universal (UHRR) usando protocolos estándar.

Flujo de trabajo de ADN-seq

Los procedimientos de extracción, cuantificación, corte, construcción de colecciones, captura de híbridos y

secuenciación de ADN genómico se describen en los ejemplos 1-6.

Flujo de trabajo de ARN-seq

- 5 El flujo de trabajo de ARN se ejecutó típicamente en paralelo al flujo de trabajo de ADN-seq en ARN y ADN emparejados extraídos de la misma muestra. Por ejemplo, el ARN extraído se cuantificó primero con Ribogreen, un tinte fluorescente específico de ARN y, a continuación, se normalizaron hasta 300 ng de ARN de entradas de alta calidad (por ejemplo, sangre, AMO) o 500 ng de ARN de entradas de baja calidad (por ejemplo, FFIP) como entrada para el protocolo de síntesis de ADNc en dos etapas. En la primera etapa de la síntesis de ADNc, se generaron cadenas de ADN monocatenarias complementarias al ARN de entrada. En la segunda etapa de la síntesis de ADNc, el ADNc monocatenario se convirtió en ADN bicatenario. El ADNc bicatenario se usó a continuación como entrada para la construcción de la colección. Si el ADNc se sintetizó a partir de ARN de alta calidad, entonces, para garantizar que la longitud del inserto fuera compatible con la secuenciación, se cortó el ADNc antes de su uso en la construcción de la colección. Las etapas de construcción de la colección, captura de híbridos, cuantificación y secuenciación se realizan todas como para las colecciones de ADN-seq estándar, con la excepción del uso del reactivo de cebado para ARN-seq y los archivos y protocolos de análisis asociados.

Creación de la base de datos de artefactos de ADN-seq

- 20 Para minimizar la llamada de alteraciones positivas falsas con el flujo de trabajo de ADN-seq, se creó una base de datos de potenciales artefactos de secuenciación a partir de líneas celulares HapMap normales. Por ejemplo, se usaron 200 ng de ADN de 20 líneas celulares HapMap individuales como entrada para la construcción de la colección estándar. A continuación, se capturaron 2 µg de cada colección con el conjunto de cebos de ADN-seq. Después de secuenciar las colecciones capturadas y procesar todos los datos a través del pipeline de análisis estándar, se creó una base de datos de artefactos basada en cualquier llamada de variante presente en los genes. Dado que las muestras se derivaron de líneas celulares HapMap normales, cualquier alteración no SNP detectada en los genes se consideró un artefacto y se usaría para filtrar los datos de prueba.

Creación de la base de datos de artefactos de ARN-seq R2

- 30 Para minimizar la detección de reordenamientos genómicos positivos falsos con el flujo de trabajo de ARN-seq, se creó una base de datos de potenciales artefactos relacionados con el proceso, de secuenciación y de análisis en los datos de ARN-seq. Por ejemplo, se usaron 300 ng de ARN de las 8 muestras de sangre normal como entrada para el protocolo de síntesis de ADNc de alta calidad y se usaron 500 ng de ARN de las 6 muestras FFIP normales como entrada para el protocolo de síntesis de ADNc de baja calidad. Se usó el rendimiento completo de la reacción de ADNc como entrada para la construcción de la colección. A continuación, se capturaron 2 µg de cada colección con el conjunto de cebos de ARN-seq. Después de secuenciar las colecciones capturadas y procesar todos los datos a través del pipeline de análisis de ARN-seq, se creó una base de datos de artefactos de reordenamientos genómicos. Dado que las muestras de sangre y FFIP se derivaron de donantes normales, cualquier reordenamiento genómico detectado se consideró un artefacto o un acontecimiento de línea germinal, y se usaría para filtrar los datos de prueba.

Validación

- 45 El conjunto de cebos de ADN-seq logró una cobertura y exactitud altamente uniformes para la detección de sustituciones de bases, inserciones y deleciones cortas y alteraciones en el número de copias focales con muestras clínicas desidentificadas. Los resultados integrados de ADN-seq y ARN-seq lograron una alta sensibilidad y especificidad para la detección de reordenamientos genómicos/fusiones génicas en mezclas de líneas celulares. Este ensayo fue altamente reproducible y el flujo de trabajo fue ampliamente compatible con las muestras clínicas de oncología hematológica de rutina, incluyendo FFIP, sangre y aspirados de médula ósea. Los resultados también fueron altamente concordantes con los resultados de las pruebas con certificación CLIA de referencia obtenidos de 76 muestras clínicas desidentificadas.

- 55 Este ensayo logró una alta exactitud en la detección de sustituciones de bases, indels, ANC focales y reordenamientos genómicos/fusiones génicas. Los datos de ADN-seq fueron altamente concordantes (>97 %) con un ensayo de ADN-seq con certificación CLIA para la detección de sustituciones de bases, indels y ANC focales en 52 muestras clínicas desidentificadas. Los resultados integrados de ADN-seq y ARN-seq lograron una alta sensibilidad y especificidad para la detección de reordenamientos genómicos en 39 mezclas de líneas celulares. Los resultados también fueron 99 % concordantes (100 de 101 alteraciones) con las sustituciones de bases, indels y reordenamientos genómicos detectados por pruebas de referencia externas con certificación CLIA que se usan de forma rutinaria para analizar muestras clínicas.

- 65 La precisión observada con este ensayo en 5 series diferentes de 13 muestras superó la especificación de rendimiento; la precisión intraserial fue del 97 % y la precisión interserial fue del 97 %. Se evaluó la precisión en base a la comparación entre series de sustituciones de bases, indels, ANC focales y reordenamientos genómicos/fusiones génicas seleccionados, conocidos y probables para muestras que superaron la especificación

de cobertura.

- 5 La sensibilidad de este ensayo para sustituciones de bases, indels y ANC se evaluó comparando los resultados de ADN-seq en 52 muestras clínicas desidentificadas con las obtenidas previamente con un ensayo de ADN-seq con certificación CLIA. Los resultados fueron un 97,7 % concordantes (167 de 171; IC del 95 %: 94-99 %) con los resultados del ensayo estándar, demostrando que este ensayo tiene la misma alta sensibilidad para la detección de sustituciones de bases, indels y ANC que el estándar.
- 10 La sensibilidad de este ensayo para reordenamientos genómicos/fusiones génicas se determinó con un sistema modelo compuesto de 39 mezclas (proporciones de mezclas del 10 % al 50 %) de 21 líneas celulares con reordenamientos genómicos conocidos. Este ensayo logró una sensibilidad muy alta para la detección de reordenamientos genómicos, detectando 165 de 167 (98,5 %; IC del 95 %: 96-99 %) reordenamientos en todas las mezclas y 64 de 66 (97 %; IC del 95 %: 89-99) reordenamientos genómicos en la mezcla al 10 %.
- 15 La especificidad de este ensayo para sustituciones de bases, indels y ANC se evaluó comparando los resultados de ADN-seq en 52 muestras clínicas desidentificadas con las obtenidas previamente con un ensayo de ADN-seq con certificación CLIA. Los resultados fueron un 97,7 % concordantes (167 de 171; IC del 95 %: 94-99 %) con los resultados del ensayo estándar, demostrando que este ensayo tiene la misma alta especificidad para la detección de sustituciones de bases, indels y ANC que el estándar.
- 20 La especificidad de este ensayo para reordenamientos genómicos/fusiones génicas se determinó con un sistema modelo compuesto de 39 mezclas (proporciones de mezclas del 10 % al 50 %) de 21 líneas celulares con reordenamientos genómicos conocidos. Este ensayo logró una especificidad muy alta para la detección de reordenamientos genómicos, llamando 4 reordenamientos positivos falsos de 165 reordenamientos positivos verdaderos en todas las mezclas.
- 25

REIVINDICACIONES

1. Un procedimiento para evaluar un sujeto, que comprende:

(a) adquirir una primera colección de ácidos nucleicos que comprenda ADN genómico de una muestra del sujeto;

(b) poner en contacto la primera colección de ácidos nucleicos con un primer conjunto de cebos para proporcionar una primera pluralidad de miembros seleccionados que comprendan un intervalo subgenómico o una porción del mismo;

(c) amplificar cada miembro de la primera pluralidad de miembros seleccionados;

(d) adquirir una segunda colección de ácidos nucleicos que comprenda ADNc obtenido a partir de ARN de la muestra;

(e) poner en contacto la segunda colección de ácidos nucleicos con un segundo conjunto de cebos para proporcionar una segunda pluralidad de miembros seleccionados que comprendan un intervalo subgenómico expresado o una porción del mismo, en el que el intervalo subgenómico expresado incluya al menos un gen de anticuerpo, un gen de receptor de la superfamilia de las inmunoglobulinas (Ig) o un fragmento de los mismos;

(f) amplificar cada miembro de la segunda pluralidad de miembros seleccionados;

(g) adquirir la secuencia del intervalo subgenómico y el intervalo subgenómico expresado;

evaluando de este modo al sujeto.

2. El procedimiento de la reivindicación 1, en el que el intervalo subgenómico expresado es un gen del receptor de linfocitos B (BCR) o un gen del receptor de linfocitos T (TCR), o un fragmento de los mismos.

3. El procedimiento de la reivindicación 1 o 2, en el que el intervalo subgenómico es un gen de anticuerpo, un gen de receptor de la superfamilia de las inmunoglobulinas (Ig) o un fragmento de los mismos.

4. El procedimiento de una cualquiera de las reivindicaciones 1-3, en el que el intervalo subgenómico es un gen del receptor de linfocitos B (BCR) o un gen del receptor de linfocitos T (TCR), o un fragmento de los mismos.

5. El procedimiento de una cualquiera de las reivindicaciones 1-4, en el que el intervalo subgenómico comprende una secuencia de un receptor de la superfamilia de inmunoglobulinas (Ig), segmento V, un segmento D, un segmento J o una secuencia de la tabla 10.

Tabla 10
secuencia que codifica un anticuerpo;
secuencia que codifica una cadena ligera de anticuerpo;
secuencia que codifica una cadena pesada de anticuerpo;
secuencia que codifica una región variable, por ejemplo, una región variable de la cadena pesada o cadena ligera;
secuencia que codifica una región constante, por ejemplo, una región constante de la cadena pesada o cadena ligera;
secuencia diagnóstica para el tipo de cadena ligera, por ejemplo, secuencia específica kappa o lambda;
secuencia de región constante, por ejemplo, una secuencia de región constante que puede distinguir una o más clases y subclases de Ig, por ejemplo, IgG (por ejemplo, IgG1, IgG2, IgG3 o IgG4), IgM, IgD, IgA (por ejemplo, IgA1 o IgA2) o IgE;
secuencia de una CDR, por ejemplo, una CDR1, CDR2 o CDR3 de la cadena ligera o una CDR1, CDR2 o CDR3 de la cadena pesada, por ejemplo, una CDR3 de la cadena pesada;
secuencia que codifica un receptor de linfocitos T, por ejemplo, secuencia que codifica una cadena alfa, beta, delta o gamma;
secuencia que codifica una región variable de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de diversidad de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de acoplamiento de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica una región de conmutación de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;

secuencia que codifica una región constante de un anticuerpo, receptor de linfocitos B o receptor de linfocitos T;
secuencia que codifica un segmento V (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica un segmento D (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica un segmento J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica segmentos V, D y J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que codifica segmentos V y J (antes o después del reordenamiento) de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión VD de un receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión DJ de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;
secuencia que comprende un punto de unión VJ de un anticuerpo o receptor de la superfamilia de Ig, por ejemplo, un gen de inmunoglobulina, un receptor de linfocitos T o un receptor de linfocitos B;

6. El procedimiento de una cualquiera de las reivindicaciones 1-5, en el que la muestra se ha obtenido de una neoplasia maligna o premaligna hematológica.

5 7. El procedimiento de la reivindicación 6, en el que la neoplasia maligna hematológica es leucemia, leucemia linfocítica aguda (LLA), leucemia mielógena aguda (LMA), leucemia linfocítica crónica (LLC), leucemia mielógena crónica (LMC), tricoleucemia, leucemia monocítica aguda (LMoA), leucemia mielomonocítica crónica (LMMC),
10 leucemia mielomonocítica juvenil (LMMJ), leucemia linfocítica granular de células grandes, linfoma, linfoma relacionado con el SIDA, linfoma cutáneo de linfocitos T, linfoma hodgkiniano, linfoma hodgkiniano clásico o linfoma hodgkiniano con predominio linfocitario nodular, micosis fungoide, linfoma no hodgkiniano, linfoma no hodgkiniano de linfocitos B, linfoma de Burkitt, linfoma linfocítico de células pequeñas (LLC/LLP), linfoma difuso de linfocitos B grandes, linfoma folicular, linfoma inmunoblástico de células grandes, linfoma linfoblástico de precursores de linfocitos B, linfoma de células del manto, linfoma no hodgkiniano de linfocitos T, micosis fungoide, linfoma anaplásico de células grandes o linfoma linfoblástico de precursores de linfocitos B, linfoma primario del sistema nervioso central, síndrome de Sézary, macroglobulinemia de Waldenström, neoplasia mieloproliferativa crónica,
15 histiocitosis de células de Langerhans, mieloma múltiple/neoplasia de células plasmáticas, síndrome mielodisplásico o neoplasia mielodisplásica/mieloproliferativa.

20 8. El procedimiento de la reivindicación 1, en el que la muestra se ha obtenido de una célula no cancerosa o célula no maligna.

9. El procedimiento de la reivindicación 8, en el que la muestra se ha obtenido de un linfocito infiltrante del tumor.

25 10. El procedimiento de la reivindicación 1, en el que el intervalo subgenómico se selecciona de las tablas 1-4.

11. El procedimiento de una cualquiera de las reivindicaciones 1-10, que comprende proporcionar la secuencia de al menos X intervalos subgenómicos, en la que X es igual a 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 o mayor.

30 12. El procedimiento de una cualquiera de las reivindicaciones 1-9 y 11, en el que el intervalo subgenómico y el intervalo subgenómico expresado corresponden a la misma secuencia del mismo gen.

13. El procedimiento de una cualquiera de las reivindicaciones 1-12, en el que el procedimiento comprende proporcionar la abundancia relativa de un acontecimiento en el intervalo subgenómico.

35 14. El procedimiento de una cualquiera de las reivindicaciones 1-13, en el que el procedimiento comprende proporcionar la abundancia relativa de un segmento V, un punto de unión VD, un punto de unión VJ o un punto de unión DJ.

Recepción de muestras y QC; Aislamiento de ADN

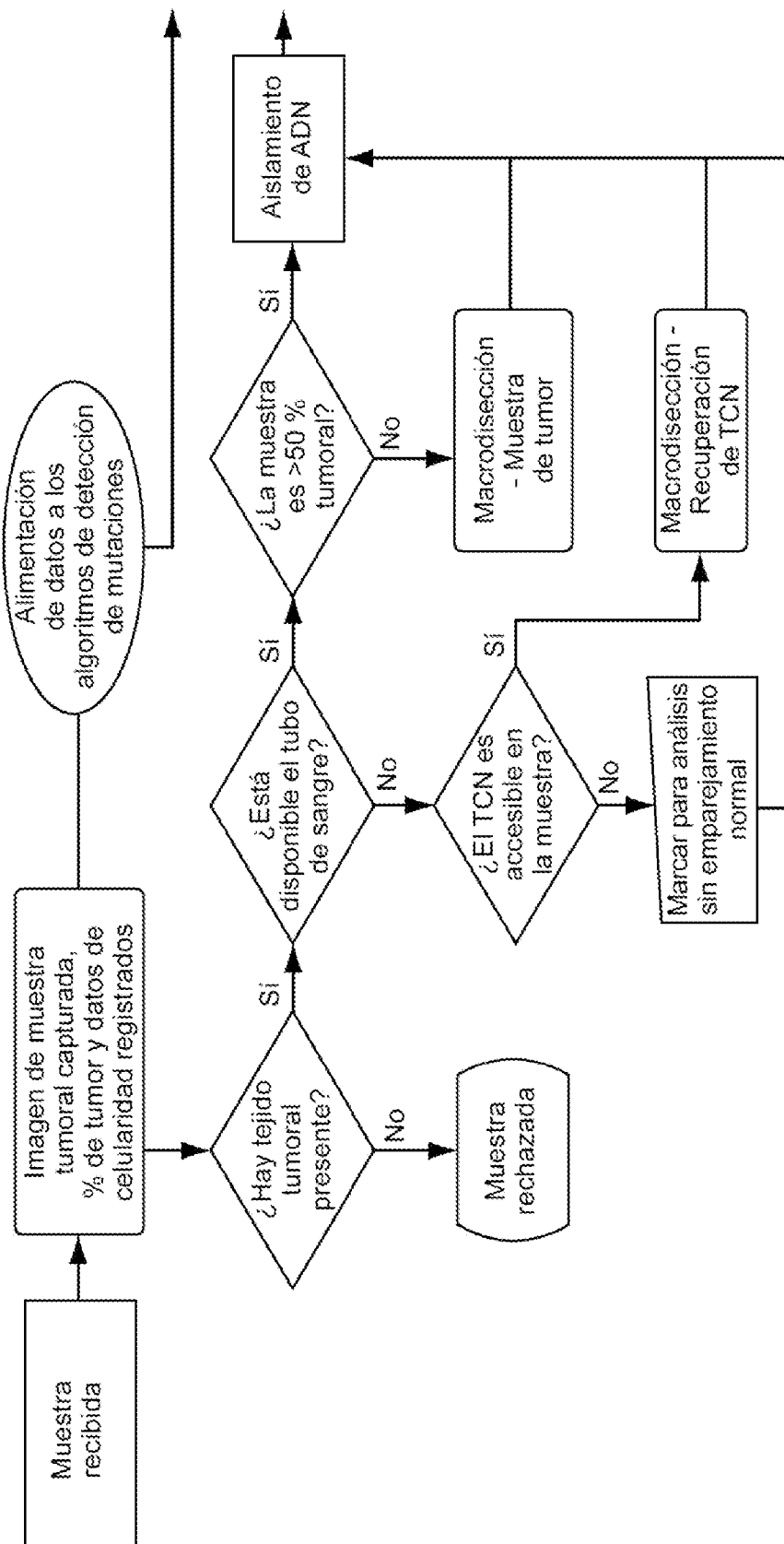


Fig. 1A

QC de ADN y generación de colección

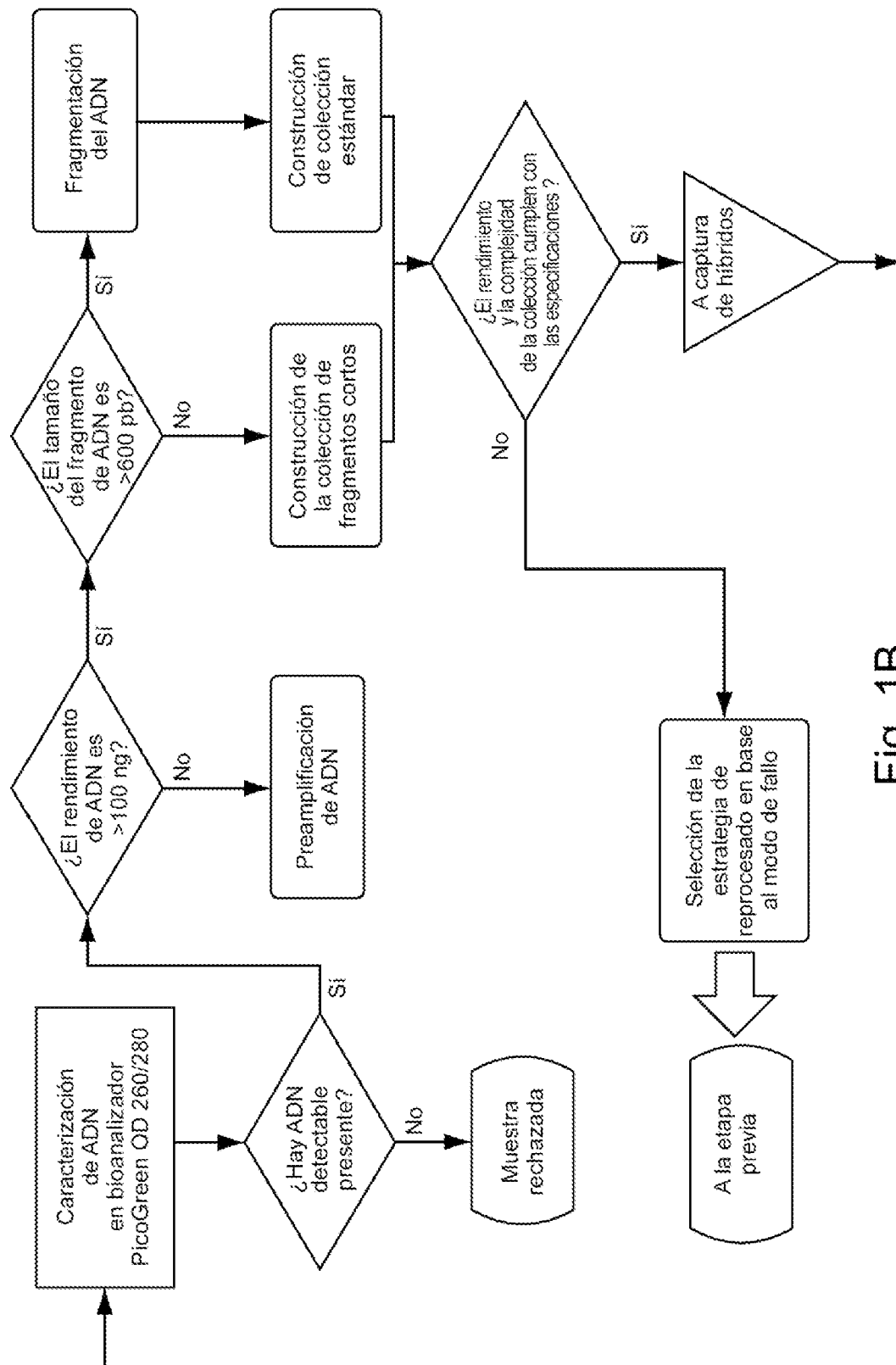


Fig. 1B

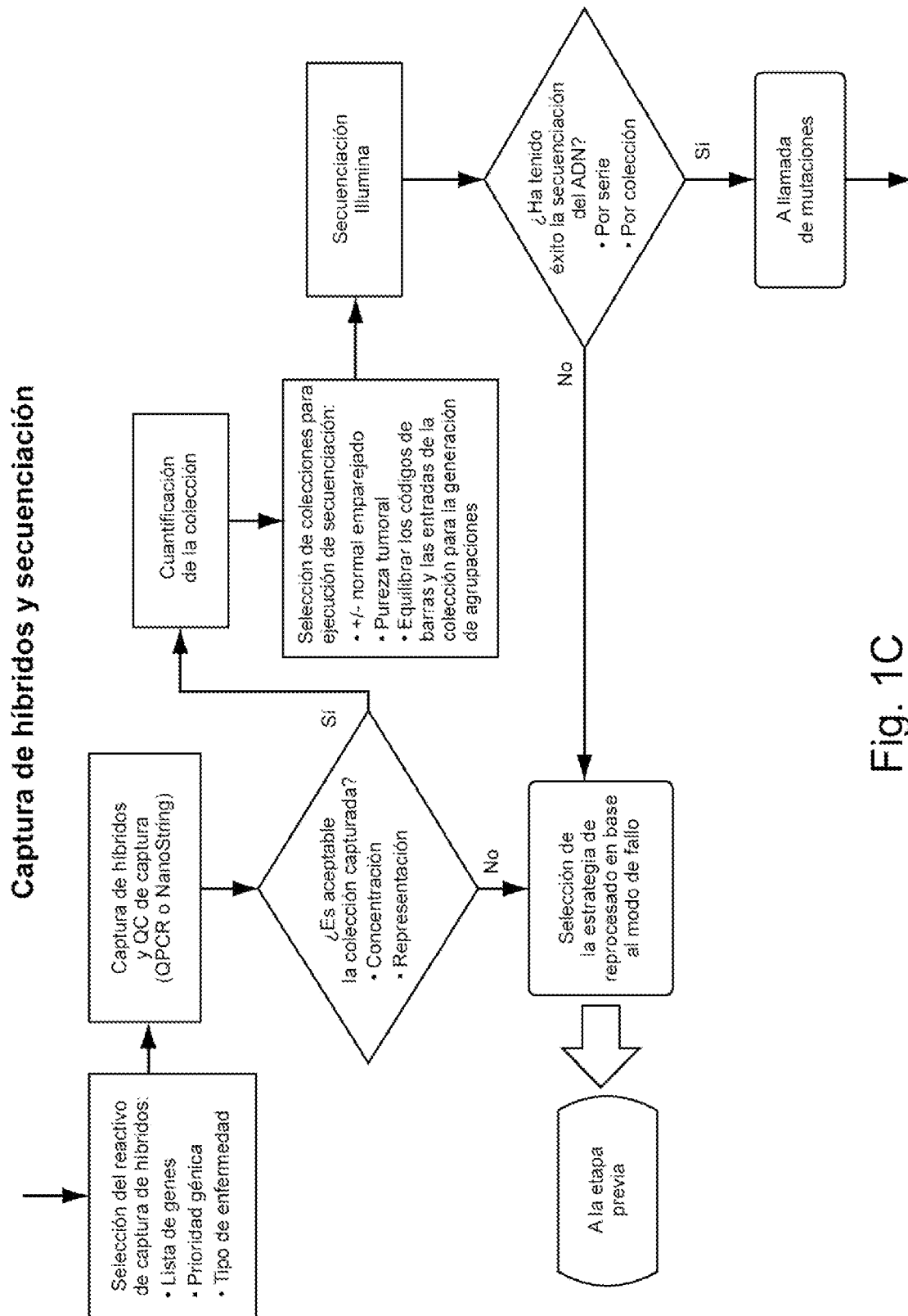


Fig. 1C

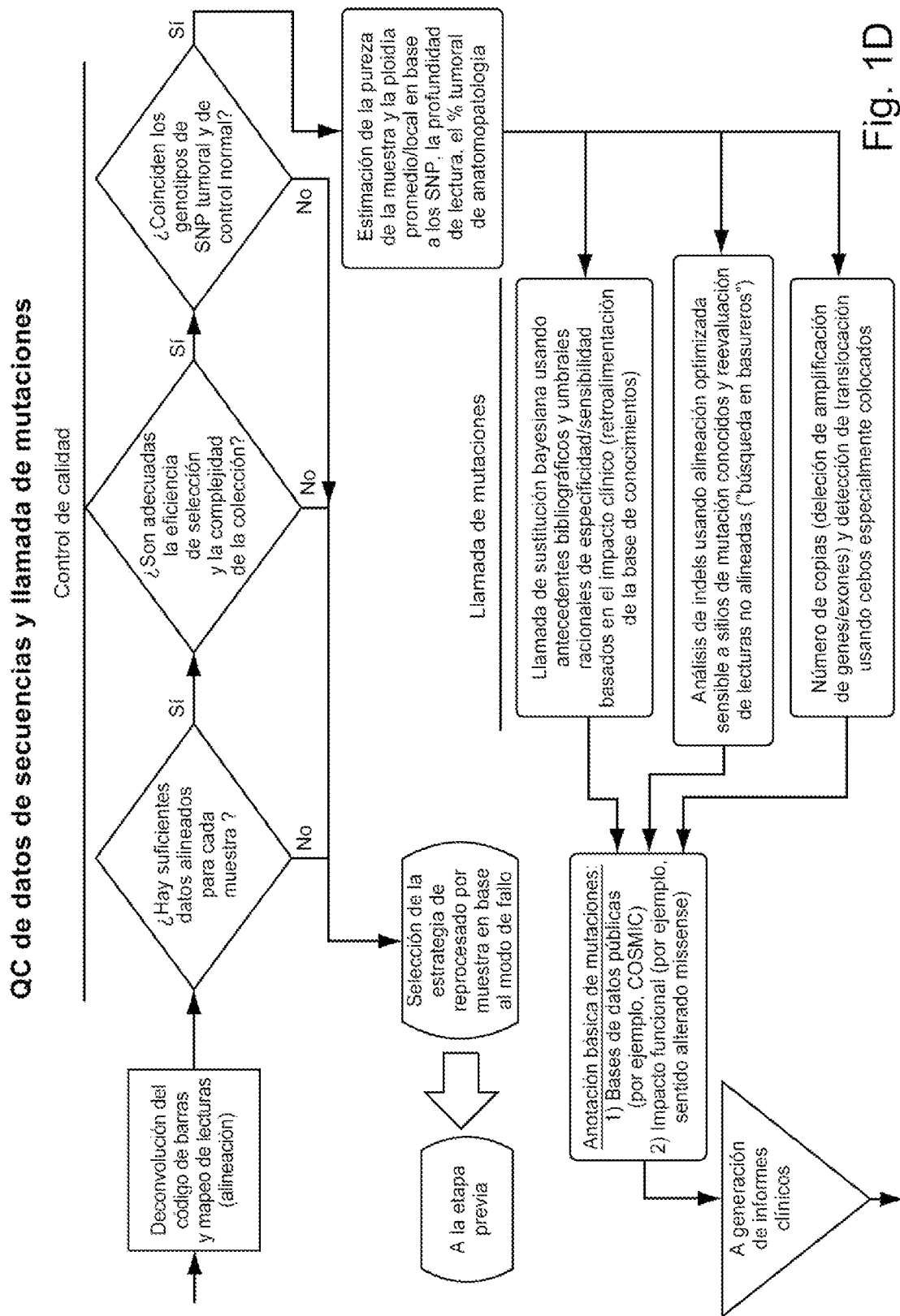


Fig. 1D

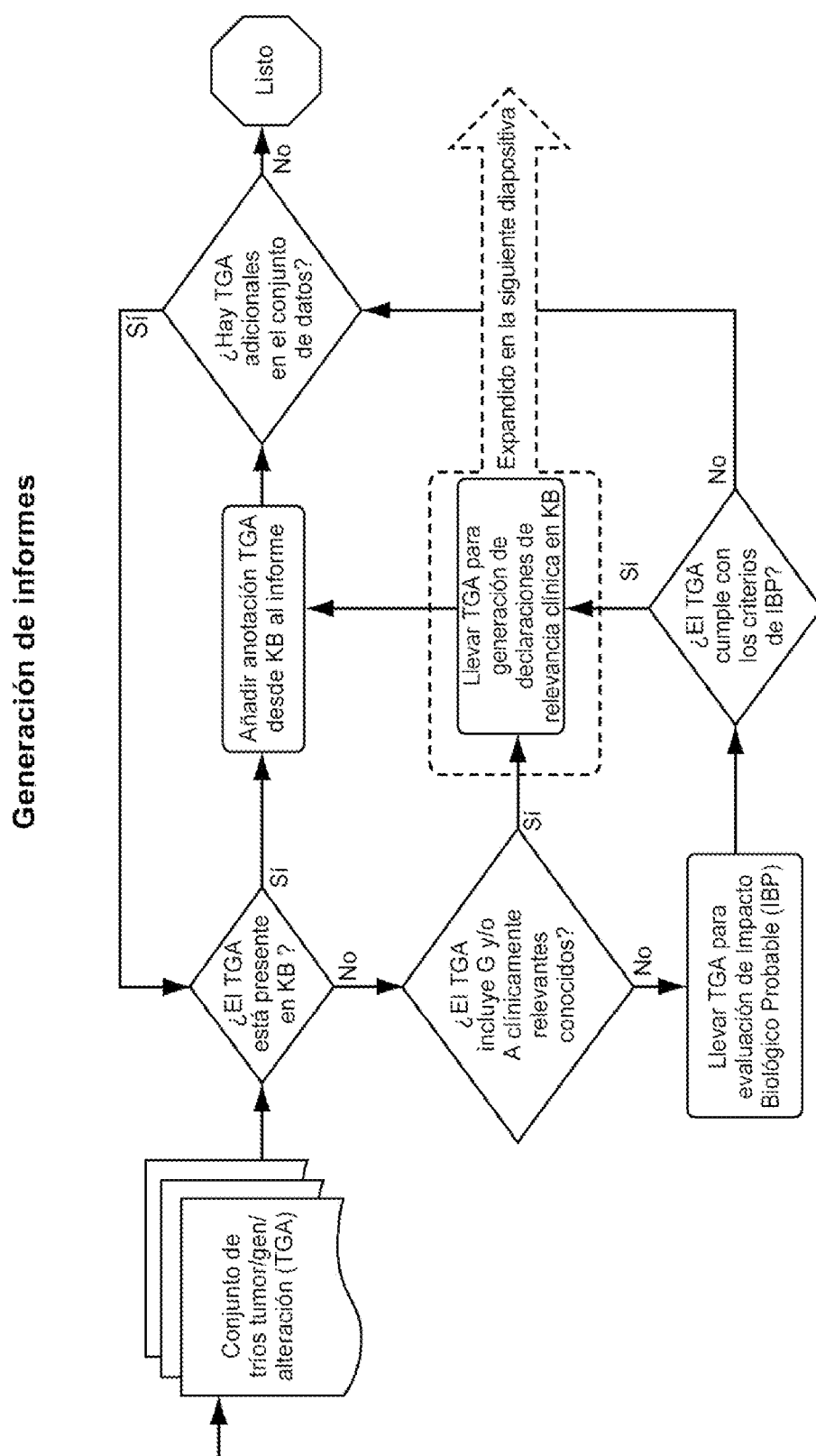


Fig. 1E

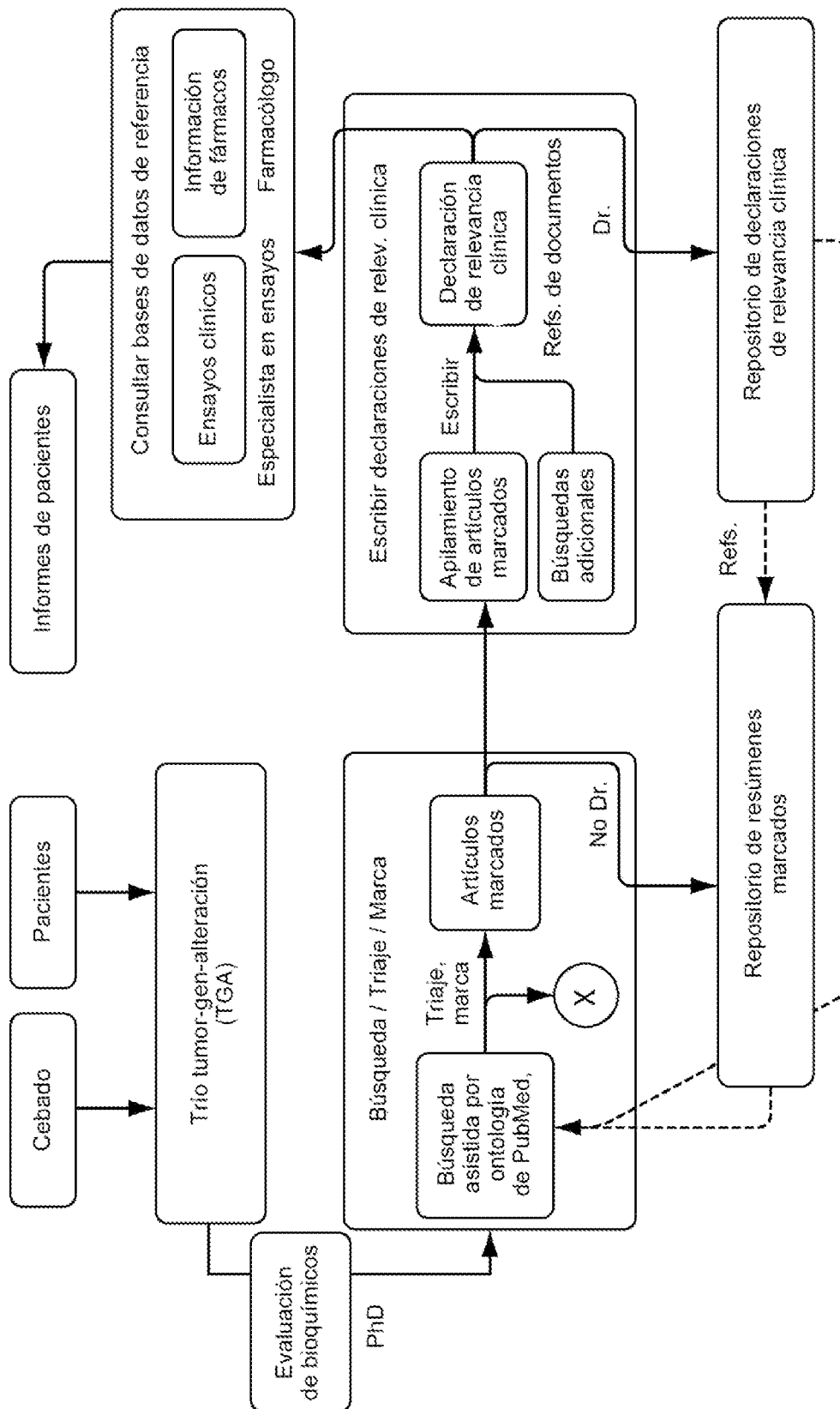


Fig. 1E

Impacto de la expectativa previa y profundidad de lectura en la detección de mutaciones

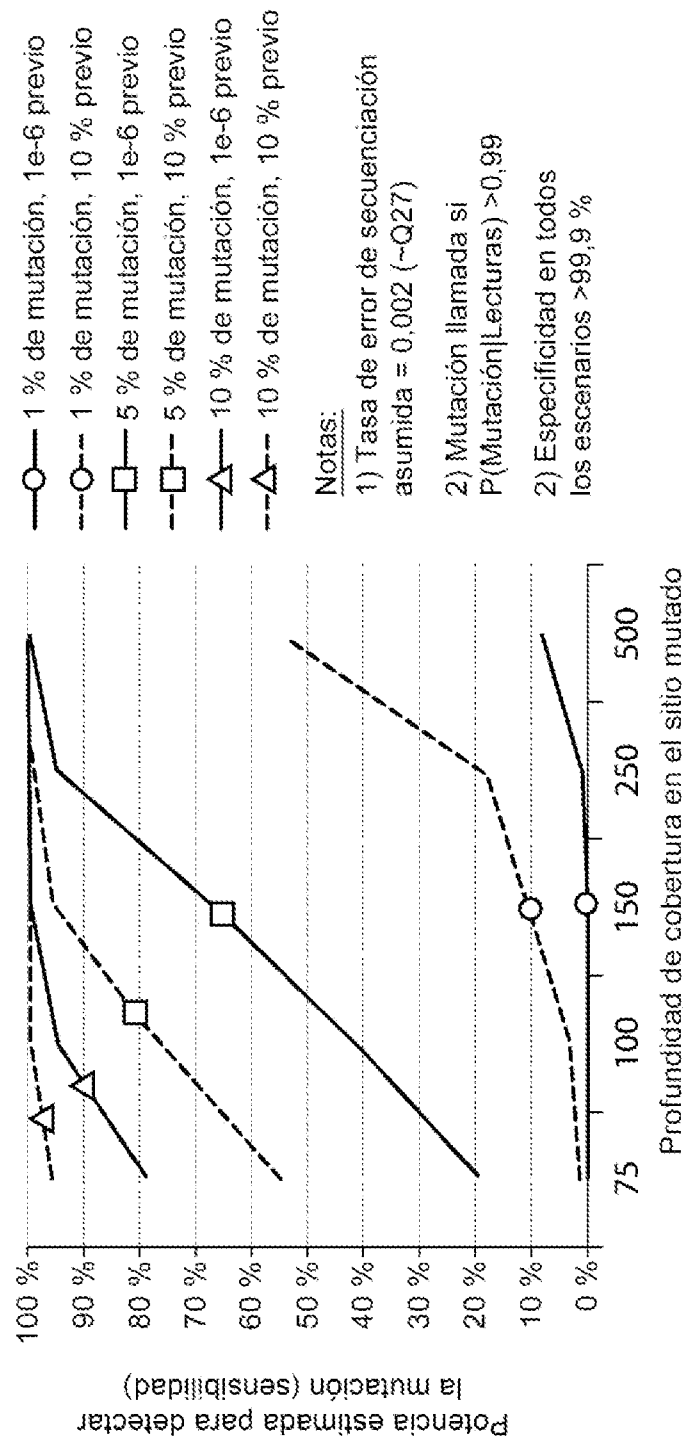


Fig. 2

REFERENCIAS CITADAS EN LA DESCRIPCION

La lista de referencias citadas por el solicitante es, únicamente, para conveniencia del lector. No forma parte del documento de patente europea. Si bien se ha tenido gran cuidado al recopilar las referencias, no pueden excluirse errores u omisiones y la OEP declina toda responsabilidad a este respecto.

Documentos citados en la descripción

- WO 2012092428 A [0002] [0146] [0147] [0190] [0208] [0218] [0297] [0303] [0319]
- WO 2013028817 A [0002]
- WO 2014130975 A [0002]
- US 20100229498 A [0133] [0157]

Literatura no patente citada en la descripción

- METZKER, M. *Nature Biotechnology Reviews*, 2010, vol. 11, 31-48 [0053] [0175]
- CRONIN M. et al. *Am J Pathol.*, 2004, vol. 164 (1), 35-42 [0106] [0263]
- MASUDA N. et al. *Nucleic Acids Res.*, 1999, vol. 27 (22), 4435-4443 [0106] [0263]
- SPECHT K. et al. *Am J Pathol.*, 2001, vol. 158 (2), 419-429 [0106] [0263]
- GNIRKE, A. et al. *Nat Biotechnol.*, 2009, vol. 27 (2), 182-189 [0133] [0157]
- ALBERT, T.J. et al. *Nat. Methods*, 2007, vol. 4 (11), 903-5 [0173]
- HODGES, E. et al. *Nat. Genet.*, 2007, vol. 39 (12), 1522-7 [0173]
- OKOU, D.T. et al. *Nat. Methods*, 2007, vol. 4 (11), 907-9 [0173]
- KAUR, H.; ARORA, A.; WENGEL, J.; MAITI, S.; ARORA, A.; WENGEL, J.; MAITI, S. Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes. *Biochemistry*, 2006, vol. 45 (23), 7347-55 [0183]
- EGHOLM, M. et al. *Nature*, 1993, vol. 365 (6446), 586-8 [0185]
- TRAPNELL C.; SÄLZBERG S.L. *Nature Biotech.*, 2009, vol. 27, 455-457 [0191]
- WARREN R. et al. *Bioinformatics*, 2007, vol. 23, 500-501 [0191]
- BUTLER J. et al. *Genome Res.*, 2008, vol. 18, 810-820 [0191]
- ZERBINO D.R.; BIRNEY E. *Genome Res.*, 2008, vol. 18, 821-829 [0191]
- LI H.; DURBIN R. *Bioinformatics*, 2010, vol. 26 (5), 569-95 [0210]
- BROWNING B.L.; YU Z. *Am. J. Hum. Genet.*, 2009, vol. 85 (5), 947-61 [0211]
- LI Y. et al. *Annu. Rev. Genomics Hum. Genet.*, 2009, vol. 10, 397-406 [0211]
- *SNV/Mix -Bioinformatics*, 15 March 2010, vol. 26 (8), 730-736 [0213]
- MCKENNA A. et al. *Genome Res.*, 2010, vol. 20 (9), 1297-303 [0216]
- YE K. et al. *Bioinformatics*, 2009, vol. 25 (21), 2865-71 [0216]
- LUNTER G.; GOODSON M. *Genome Res.*, 2010 [0216]
- LI H. et al. *Bioinformatics*, 2009, vol. 25 (16), 2078-9 [0216]
- ALBERS C.A. et al. *Genome Res.*, 2011, vol. 21 (6), 961-73 [0217]
- LE S.Q.; DURBIN R. *Genome Res.*, 2011, vol. 21 (8), 952-60 [0217]
- WONG K.K. et al. *Nucleic Acids Res.*, 1996, vol. 24 (15), 3778-83 [0270]
- BERGER et al. *Genome Res.*, 2010, vol. 20 (4), 413-27 [0315]
- GARBER et al. *Nat Methods*, 2011, vol. 8 (5), 469-77 [0315]
- LEVIN et al. *Genome Biol.*, 2009, vol. 10 (10), R115 [0315]