

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 769 796**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12N 15/11** (2006.01)

**C12Q 1/68** (2008.01)

**C07H 21/04** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.07.2013 PCT/US2013/049402**

87 Fecha y número de publicación internacional: **09.01.2014 WO14008447**

96 Fecha de presentación y número de la solicitud europea: **03.07.2013 E 13737778 (4)**

97 Fecha y número de publicación de la concesión europea: **04.09.2019 EP 2872629**

54 Título: **Oligonucleótidos de bloqueo aumentados en  $T_m$  y señuelos para un enriquecimiento de diana mejorado y una selección fuera de diana reducida**

30 Prioridad:

**03.07.2012 US 201261667919 P**  
**21.12.2012 US 201261745435 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**29.06.2020**

73 Titular/es:

**INTEGRATED DNA TECHNOLOGIES INC. (50.0%)**  
**1710 Commercial Park**  
**Coralville, IA 52241, US y**  
**FOUNDATION MEDICINE, INC. (50.0%)**

72 Inventor/es:

**BEHLKE, MARK AARON;**  
**HAVENS, JOHN ROBERT;**  
**ROSE, SCOTT DANIEL;**  
**JAROSZ, MIRNA;**  
**ZWIRKO, ZACHARY;**  
**LIPSON, DORON y**  
**JUHN, FRANK SOO**

74 Agente/Representante:

**CURELL SUÑOL, S.L.P.**

ES 2 769 796 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

Oligonucleótidos de bloqueo aumentados en  $T_m$  y señuelos para un enriquecimiento de diana mejorado y una selección fuera de diana reducida.

### Campo de la invención

La presente invención se refiere a composiciones de oligonucleótidos modificados y a su utilización en métodos para la selección y secuenciación de ácidos nucleicos. En particular, la invención se refiere a oligonucleótidos aumentados en  $T_m$  como bloqueantes y señuelos ("baits"), así como a otros reactivos para un enriquecimiento mejorado de la diana y una selección fuera de diana ("off-target") reducida. Las composiciones y reactivos de oligonucleótidos encuentran aplicaciones robustas en la preparación de moldes de ácidos nucleicos para aplicaciones de secuenciación de nueva generación.

### Antecedentes de la invención

La hibridación de ácidos nucleicos presenta un papel significativo en aplicaciones biotecnológicas referidas a la identificación, selección y secuenciación de ácidos nucleicos. Las aplicaciones de secuenciación con ácidos nucleicos genómicos como materiales diana exige la selección de dianas de ácidos nucleicos de interés a partir de una mezcla altamente compleja. La calidad de los esfuerzos de secuenciación depende de la eficiencia del procedimiento de selección que, a su vez, se basa en el grado de enriquecimiento de las dianas de ácidos nucleicos respecto a las secuencias no diana.

Se ha utilizado una diversidad de métodos para enriquecer en las secuencias deseadas a partir de un reservorio complejo de ácidos nucleicos, tales como ADN genómico o ADNc. Entre estos métodos se incluyen la reacción en cadena de la polimerasa (PCR), las sondas de inversión molecular (MIP) o captura de secuencias mediante formación de híbridos ("captura de híbridos"; ver, por ejemplo, Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. y Turner, D.J., "Target-enrichment strategies for next-generation sequencing", Nat. Methods, "Target-enrichment strategies for next-generation sequencing," Nat. Methods 7:111-118, 2010). La captura de híbridos ofrece ventajas respecto a otros métodos en el aspecto de que este método requiere menos procedimientos de amplificación o manipulación enzimática del ácido nucleico diana que los demás métodos. En consecuencia, el método de captura de híbridos introduce menos errores en la biblioteca de secuenciación final. Por este motivo, el método de captura de híbridos es un método preferente para el enriquecimiento en secuencias deseadas a partir de un reservorio complejo de ácidos nucleicos y resulta ideal para preparar moldes en aplicaciones de secuenciación de nueva generación (NGS).

Las aplicaciones de NGS habitualmente implican romper aleatoriamente ADN genómico o ADNc largo en tamaños de fragmento más pequeños con una distribución de tamaños de 100 a 3.000 pb de longitud, según la plataforma de NGS utilizada. Los extremos del ADN se tratan enzimáticamente para facilitar la ligación y se ligan adaptadores universales de ADN a los extremos para proporcionar los moldes de NGS resultantes. Las secuencias adaptadoras terminales proporcionan un sitio universal para la hibridación de cebadores, de manera que puede conseguirse la expansión clonal de las dianas de ADN deseadas e introducirse en los procedimientos automatizados de secuenciación utilizados en las aplicaciones de NGS. El método de captura de híbridos está destinado a reducir la complejidad del reservorio de fragmentos aleatorios de ADN de, por ejemplo,  $3 \times 10^9$  bases (el genoma humano) a subgrupos mucho más pequeños, de  $10^3$  a  $10^8$  bases que están enriquecidos en las secuencias específicas de interés. La eficiencia de este procedimiento se relaciona directamente con la calidad de la captura y enriquecimiento conseguidos para secuencias de ADN deseadas a partir del reservorio complejo inicial.

Las aplicaciones de NGS típicamente utilizan el método de captura de híbridos de enriquecimiento de la manera siguiente. Un reservorio preparado de moldes de NGS se desnaturaliza por calor y se mezcla con un reservorio de oligonucleótidos sonda de captura ("señuelos"). Los señuelos se diseñan para hibridarse con las regiones de interés dentro del genoma diana y habitualmente presentan 60 a 200 bases de longitud y además se modifican para contener un ligando que permita la posterior captura de estas sondas. Un método de captura común incorpora un grupo (o grupos) de biotina en los señuelos. Tras completarse la hibridación para formar los híbridos de molde de ADN:señuelo, se lleva a cabo la captura con un componente que presenta afinidad para únicamente el señuelo. Por ejemplo, puede utilizarse estreptavidina-perlas magnéticas para unirse a la fracción de biotina de los señuelos biotinilados que se hibridan con las dianas de ADN deseadas a partir del reservorio de moldes de NGS. El lavado elimina los ácidos nucleicos no unidos, reduciendo la complejidad del material retenido. A continuación, se eluyó el material retenido respecto de las perlas magnéticas y se introdujo en procedimientos automatizados de secuenciación.

Aunque la hibridación del ADN con los señuelos puede ser exquisitamente específicas, se mantienen secuencias no deseadas en el reservorio enriquecido tras completar el método de captura de híbridos. La fracción más grande de dichas secuencias no deseadas se debe a sucesos no deseados de hibridación entre moldes de NGS sin complementariedad con los señuelos y los moldes de NGS que sí la tienen. Dos tipos de hibridaciones no deseadas que aparecen en el método de captura de híbridos incluyen las secuencias siguientes: (1) miembros de ADN

altamente repetitivos que se encuentran en ADN genómico endógeno, y (2) las secuencias adaptadoras terminales que se introducen en cada uno de los moldes de NGS del reservorio.

Hodges et al. enseñan la utilización de selección de híbridos de intervalos genómicos discretos en micromatrices de diseño personalizado para la secuenciación en paralelo masiva (Hodges E. et al., Nature 4:960-974, 2009). Hodges et al. identificaron una fuente de hibridación cruzada derivada de la utilización de secuencias adaptadoras complementarias comunes presentes en los fragmentos hibridados que pueden unirse entre sí al desnaturalizar el ADN antes de la hibridación. Hodges et al. complementaron la mezcla de hibridación con un exceso molar de cuatro "oligos de bloqueo" diferentes que complementaban cada cadena de la secuencia adaptadora, compitiendo para la unión de los adaptadores.

Los miembros repetitivos de ADN endógeno, tales como una secuencia Alu o una secuencia LINE, presentes en un fragmento de ADN en el reservorio complejo, pueden hibridarse con otros miembros similares presentes en otro fragmento de ADN no relacionado. Estos fragmentos, que pueden derivarse originariamente de localizaciones muy diferentes dentro del genoma, se unen durante el procedimiento de hibridación del método de captura de híbridos. En el caso de que uno de dichos fragmentos de ADN represente un fragmento deseado que contiene un sitio de unión para un señuelo, el fragmento no deseado resultará capturado junto con el fragmento deseado. Esta clase de moldes de NGS no deseados puede reducirse mediante la adición de un exceso de miembros repetidos a la reacción de hibridación. Más comúnmente, se añade ADN C<sub>0</sub>t-1 humano a la reacción de hibridación, que se une a Alu, LINE y otros sitios repetidos en la diana y bloquea la capacidad de los moldes de NGS de interactuar entre sí de esta manera.

Una clase más problemática de moldes de NGS no deseados que se recuperan durante la captura de híbridos surge de interacciones entre secuencias adaptadoras terminales que se construyen en cada uno de los moldes de NGS del reservorio. Debido a que el reservorio de moldes de NGS típicamente contiene secuencias adaptadoras terminales idénticas en cada fragmento de ADN, las secuencias adaptadoras se encuentran presentes a una concentración eficaz muy elevada en la solución de hibridación. En consecuencia, pueden hibridarse moldes de NGS no relacionados entre sí por sus extremos, dando lugar de esta manera a una "cadena conectada en serie" de fragmentos de ADN de otro modo no relacionados que se unen entre sí. Por lo que, si uno de dichos fragmentos unidos contiene un sitio de unión para un señuelo, se captura la cadena conectada en serie entera. De esta manera, la captura de un solo fragmento deseado puede producir un gran número de fragmentos no deseados, reduciendo la eficiencia global de enriquecimiento para el fragmento deseado. Esta clase de suceso de captura no deseado puede reducirse mediante la adición de un exceso de secuencias adaptadoras de cadena sencilla a la reacción de hibridación. Sin embargo, la capacidad de reducir eficazmente los denominados sucesos de captura de cadena conectada en serie con un exceso de secuencias adaptadoras está limitada por una eficiencia de aproximadamente 50% a 60% para la captura del fragmento deseado.

A pesar de la utilización de ADN C<sub>0</sub>t-1 y oligonucleótidos adaptadores de bloqueo en la reacción de hibridación, queda una cantidad significativa de fragmentos de ADN no deseados contaminantes en el reservorio de secuenciación después de la etapa de captura de híbridos, en gran parte debido a que los métodos de bloqueo no presentan un éxito completo. De esta manera, existe una necesidad de mejora de la eficiencia de captura y de reducción de la contaminación por secuencias no deseadas, de manera que puedan dedicarse los recursos a la secuenciación de una fracción más grande de dianas de interés y se generen menos dianas que no resultan de interés.

De esta manera, las interacciones de ácidos nucleicos fuera de diana pueden limitar la eficiencia de la selección de los ácidos nucleicos diana mediante hibridación (por ejemplo, hibridación en solución) con una sonda de captura, por ejemplo un oligonucleótido de señuelo. La selección de secuencias fuera de diana puede resultar, por ejemplo, en uno o más de: rendimientos reducidos de captura de hibridación y/o la captura de híbridos artefactos que, a su vez, conduce a ineficiencias en etapas posteriores, por ejemplo, en la secuenciación.

La selección fuera de diana típicamente se incrementa al reducirse las condiciones de astringencia de la selección de híbridos, por ejemplo al seleccionar un dúplex de diana:captura que presenta una temperatura de fusión de los ácidos nucleicos más baja (por ejemplo, los dúplex de ADN:ADN comparado con los dúplex de ARN:ADN). De esta manera, la captura de una secuencia fuera de diana puede resultar más problemática en las hibridaciones de ADN:ADN.

Típicamente, entre los miembros de una biblioteca se incluyen un inserto de biblioteca, con frecuencia una secuencia de un gen de interés, por ejemplo, un segmento para la secuenciación. Si un miembro está en la diana, el inserto de biblioteca forma un dúplex con la sonda de captura. Típicamente, los miembros de la biblioteca también incluyen una o más secuencias no diana. Estas típicamente no son partes de un gen de interés sino, por el contrario, secuencias adaptadoras, cebadores de amplificación o etiquetas, o etiquetas de código de barras. La secuencia no diana del miembro de biblioteca hibridado con sonda de captura puede, mediante formación de dúplex con otras secuencias en la mezcla de reacción, conducir a la selección de secuencias no deseadas, por ejemplo miembros de una biblioteca de secuencias fuera de diana. Aunque sin respaldo teórico, la concatenación

de miembros de una biblioteca de secuencias fuera de diana y secuencias fuera de diana puede resultar en la selección de secuencias fuera de diana.

En la presente memoria se dan a conocer métodos y composiciones para minimizar la selección de ácidos nucleicos fuera de diana, por ejemplo que minimizan la selección de miembros de la biblioteca que no forman un dúplex con la sonda de captura. Se dan a conocer en la presente memoria métodos y composiciones que reducen la selección de secuencias no diana, por ejemplo la selección mediada por adaptadores.

Se conocen en la técnica oligonucleótidos con modificaciones específicas que fomentan una temperatura de fusión ( $T_m$ ) incrementada al hibridarse con un ácido nucleico. Petersen et al. enseñan ácidos nucleicos bloqueados (ANB) que presentan una termoestabilidad incrementada al hibridarse con ADN o ARN, según se refleja en dúplex de ANB:ADN y ANB:ARN que presentan temperaturas de fusión ( $T_m$ ) incrementadas (Petersen M. et al., Trends in Biotechnology 21:74-81, 2003). Fiandaca et al. enseñan la utilización de sondas bloqueantes de APN que potencian la especificidad de la unión de sondas en los ensayos, en los que la estabilidad de los dúplex de APN/ácido nucleico es mayor que la estabilidad de híbridos compuestos de ácidos nucleicos naturales (Fiandaca M.J. et al., Peptide Nucleic Acids: Protocols and Applications. Horizon Scientific Press (GB), 129-141, 1999). Dako et al. enseñan que las sondas bloqueantes basadas en ácido péptido nucleico (APN) que son complementarias a una o más secuencias no diana en un método de supresión de la unión de la secuencia no diana a una sonda de captura (Dako A.S. et al., documento n° WO 98/24933 A1 (11/06/1998)). Chen enseña un ligante de surco menor (MGB) y oligonucleótidos que presentan un grupo MGB. Dichos oligonucleótidos-MGB presentan temperaturas de fusión ( $T_m$ ) más elevadas y una especificidad incrementada para los dúplex hibridados (Chen, C., documento n° WO 2011/139920 A2 (2011-11-10)). Frebourg et al. enseñan etiquetas de cebador de amplificación en las que una parte de la etiqueta del cebador puede conducir a un incremento de la temperatura de fusión ( $T_m$ ) al hibridarse con un segmento de ensayo (Frebourg T. et al., documento n° US 2005/244830 A1 (03/11/2005)). Krainer et al. enseñan un método de hibridación de oligonucleótidos en el que por lo menos un oligonucleótido no ADN competitivo (por ejemplo, ANB o APN) se utiliza para cada sonda sobre un soporte (Krainer S. et al., documento n° WO 2011/000836 A1 (06/01/2011)).

#### Breve resumen de la invención

En un aspecto, la invención se refiere a un método de selección de un ácido nucleico molde deseado a partir de una población de ácidos nucleicos molde. El método incluye dos etapas. La primera etapa es poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  para formar una mezcla. La segunda etapa incluye aislar el ácido nucleico molde deseado a partir de la mezcla. En un primer aspecto, el método proporciona, como parte de la etapa de contacto, la subetapa de incubar la mezcla a una temperatura aproximadamente igual al valor de la  $T_m$  aumentada óptima del oligonucleótido aumentado en  $T_m$ . En una primera forma de realización preferida de este aspecto, el oligonucleótido aumentado en  $T_m$  incluye una pluralidad de grupos que aumentan la  $T_m$ . En este aspecto, la pluralidad de grupos que aumentan la  $T_m$  comprende entre aproximadamente 2 y aproximadamente 25 grupos que aumentan la  $T_m$ . Algunas formas de realización adicionales proporcionan que la pluralidad de grupos que aumentan la  $T_m$  comprenda grupos de ácido nucleico bloqueado o grupos de ácidos nucleicos bicíclicos. Entre los aspectos preferentes de dichas formas de realización se incluyen miembros de los grupos de ácidos nucleicos bloqueados o de los grupos de ácidos nucleicos bicíclicos que presentan nucleobases seleccionadas del grupo que consiste en citosina, adenina y timina. En un segundo aspecto, el método incluye como el oligonucleótido aumentado en  $T_m$  por lo menos un miembro seleccionado del grupo que consiste en las SEC ID n° 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, 32, 34 y 36. En un tercer aspecto, el método proporciona un oligonucleótido aumentado en  $T_m$  que incluye un bloqueante. En una primera forma de realización preferida de dicho aspecto, el bloqueante presenta una complementariedad sustancial de las secuencias respecto a por lo menos una secuencia en un extremo de cada miembro de la población de ácidos nucleicos molde. En todavía otra forma de realización preferida, el bloqueante incluye además un dominio de código de barras que presenta una pluralidad de nucleótidos. En algunas formas de realización, la pluralidad de nucleótidos incluye entre aproximadamente 5 y aproximadamente 12 nucleótidos dispuestos de manera sustancialmente contigua. En otras formas de realización, el dominio de código de barras incluye nucleótidos que presentan como nucleobases por lo menos un miembro seleccionado del grupo de adenina, timina, citosina, guanina o una base universal, tal como inosina, 3-nitropirrol, 5-nitroindol y combinaciones de los mismos.

En un tercer aspecto, el método presenta como la etapa de contacto el objetivo de resultar en una inhibición sustancial de la formación de complejo entre el ácido nucleico molde deseado y ácidos nucleicos molde no deseados. En un cuarto aspecto, el método incluye como etapa de aislamiento de los ácidos nucleicos molde deseados, dos etapas adicionales. La primera etapa es la formación de un complejo híbrido entre el ácido nucleico deseado y un segundo oligonucleótido. La segunda etapa es separar el complejo híbrido a partir de la mezcla. Respecto a dicho cuarto aspecto, el segundo oligonucleótido incluye un señuelo. En determinadas formas de realización, el señuelo comprende una secuencia que presenta una complementariedad sustancial de la secuencia respecto a una secuencia dentro del ácido nucleico molde deseado. En otras formas de realización, el señuelo comprende una pluralidad de grupos que aumentan la  $T_m$ . En todavía otras formas de realización, el señuelo incluye una modificación covalente que permite la selección del complejo híbrido. Como parte de estas últimas

formas de realización, la modificación covalente es un grupo biotinilado. Todavía otras formas de realización permiten que el complejo híbrido se ponga en contacto con un soporte sólido inmovilizado con avidina o estreptavidina.

En otro aspecto, la invención se refiere a un método para llevar a cabo la secuenciación masiva en paralelo. El método incluye cuatro etapas. La primera etapa es de preparación de una población de biblioteca de ácidos nucleicos molde. La segunda etapa es de puesta en contacto de la población de biblioteca de ácidos nucleicos molde con por lo menos un oligonucleótido aumentado en  $T_m$  como bloqueante, una pluralidad de oligonucleótidos a modo de señuelos y ADN  $C_{ot}$ -1, a fin de formar una mezcla. La tercera etapa es de aislamiento de una pluralidad de ácidos nucleicos molde deseados a partir de la mezcla. La cuarta etapa es de secuenciación de la pluralidad de ácidos nucleicos molde deseados. El miembro o miembros de la pluralidad de oligonucleótidos a modo de señuelos presenta una complementariedad sustancial respecto a una secuencia dentro de por lo menos un miembro de la pluralidad de ácidos nucleicos molde deseados. En un primer aspecto, el método incluye miembros de la población de biblioteca de ácidos nucleicos molde, incluyendo cada uno por lo menos una secuencia adaptadora terminal idéntica con un intervalo de tamaño de entre aproximadamente 15 nucleótidos y aproximadamente 75 nucleótidos. En un segundo aspecto, el método incluye un bloqueante con complementariedad sustancial de la secuencia respecto a por lo menos una secuencia adaptadora terminal idéntica de la población de biblioteca de ácidos nucleicos molde. En un tercer aspecto, el método incluye como secuencia o secuencias adaptadoras terminales idénticas, un dominio de código de barras. En un cuarto aspecto, el método proporciona un bloqueante con complementariedad sustancial de la secuencia respecto a por lo menos una secuencia adaptadora terminal idéntica. En un quinto aspecto, el método incluye como etapa de contacto, la etapa de incubar la mezcla a una temperatura aproximadamente igual al valor de la  $T_m$  aumentada óptima del oligonucleótido u oligonucleótidos aumentados en  $T_m$ . En un sexto aspecto, el método proporciona que el oligonucleótido u oligonucleótidos aumentados en  $T_m$  como bloqueantes incluya por lo menos un miembro seleccionado del grupo que consiste en las SEC ID nº 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, 32, 34 y 36. En un séptimo aspecto, el método proporciona que la etapa de aislar una pluralidad de ácidos nucleicos molde deseados a partir de la mezcla incluya dos etapas. La primera etapa es de formación de una pluralidad de complejos híbridos entre la pluralidad de ácidos nucleicos molde deseados y una pluralidad de oligonucleótidos a modo de señuelos. La segunda etapa es separar la pluralidad de complejos híbridos a partir de la mezcla. En un octavo aspecto, el método proporciona como pluralidad de nucleótidos de señuelo, una pluralidad de grupos que aumentan la  $T_m$ . En una forma de realización de dicho aspecto, cada señuelo incluye una modificación covalente para permitir la selección del complejo híbrido que incluye el señuelo. En una forma de realización adicional de dicho aspecto, la modificación covalente es un grupo biotinilado. Como otra forma de realización de dicho aspecto, la pluralidad de complejos híbridos se pone en contacto con un soporte sólido inmovilizado con avidina o estreptavidina.

En una forma de realización, el método comprende además proporcionar miembros seleccionados de la biblioteca (en ocasiones denominados en la presente memoria "captura de biblioteca").

En una forma de realización, el método comprende además separar los miembros seleccionados de la biblioteca respecto de la sonda de captura.

En una forma de realización, el método comprende además secuenciar la inserción de un miembro seleccionado de la biblioteca, por ejemplo secuenciando las inserciones de por lo menos 2, 5, 10, 15, 20, 30 o 50 alteraciones de genes o ácidos nucleicos, por ejemplo alteraciones de genes o ácidos nucleicos indicadas en la presente memoria.

En una forma de realización, el valor para un parámetro relacionado con la interacción de unión puede ser un valor de afinidad, tasa de asociación, inversa de la tasa de disociación o temperatura de fusión del ácido nucleico (por ejemplo,  $T_m$ , la temperatura a la que la mitad de las cadenas de ADN se encuentran en el estado de doble hélice y la mitad se encuentra en el estado de enrollado aleatorio).

En una forma de realización, el método comprende la utilización de un primer oligonucleótido bloqueante que forma un dúplex con una primera secuencia de ácidos nucleicos no diana, por ejemplo una primera secuencia adaptadora y, opcionalmente, un segundo oligonucleótido bloqueante que forma un dúplex con una segunda secuencia de ácidos nucleicos no diana, por ejemplo una segunda secuencia adaptadora. Un grupo de bloqueantes oligonucleótidos comprende una pluralidad de diferentes bloqueantes oligonucleótidos.

En una forma de realización, el oligonucleótido bloqueante inhibe la formación de un dúplex entre una secuencia en la reacción y la secuencia no diana de un miembro de biblioteca que está formando dúplex con la sonda de captura (por ejemplo, el oligonucleótido bloqueante inhibe la formación de cadenas concatenadas de miembros de biblioteca).

En una forma de realización, un miembro de biblioteca comprende una inserción, por ejemplo un intervalo subgenómico, y una secuencia no diana, por ejemplo una secuencia común a una pluralidad de miembros de biblioteca. En una forma de realización, las inserciones son secuencias subgenómicas, por ejemplo de ácidos

nucleicos procedentes de una muestra tumoral, y la secuencia no diana es una secuencia no presente en el genoma o una secuencia no presente en las secuencias subgenómicas, por ejemplo una etiqueta de amplificación o etiqueta de código de barras.

- 5 En una forma de realización, los miembros de biblioteca o miembros seleccionados de biblioteca incluyen intervalos subgenómicos de por lo menos 2, 5, 10, 15, 20, 30 o 50 alteraciones de genes o ácidos nucleicos, por ejemplo alteraciones de genes o ácidos nucleicos indicadas en la presente memoria.
- 10 En una forma de realización, una pluralidad de miembros de biblioteca, o miembros seleccionados de biblioteca, por ejemplo en X (en donde X es igual a 2, 5, 10, 20, 50, 100, 200 o más) miembros de biblioteca, o miembros seleccionados de biblioteca, presenta una primera secuencia no diana en el extremo 5' de la inserción y una segunda secuencia no diana en el extremo 3' de la inserción.
- 15 En una forma de realización, la secuencia no diana incluye una secuencia no diana que está presente en una pluralidad de secuencias no diana, por ejemplo una secuencia para la amplificación y una secuencia no diana que es única, por ejemplo un código de barras. Típicamente, algunos, más sustancialmente la totalidad o la totalidad de los miembros de la biblioteca, incluirá una secuencia no diana común. En formas de realización, la biblioteca o los miembros seleccionados de biblioteca, comprende por lo menos X miembros (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que presentan una secuencia no diana común.
- 20 En una forma de realización, el oligonucleótido bloqueante forma un dúplex con una secuencia de ácidos nucleicos no diana de por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más), en donde el dúplex presenta una  $T_m$  que es superior a la  $T_m$  de un dúplex formado por una secuencia de ácidos nucleicos no diana con un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana.
- 25 En una forma de realización, la temperatura de fusión más elevada del ácido nucleico del dúplex oligonucleótido bloqueante modificado es aproximadamente 5°C a aproximadamente 25°C más elevada que la de un dúplex oligonucleótido bloqueante no modificado, o superior (por ejemplo, 2°C, 5°C, 10°C, 15°C, 20°C, 25°C o más). En una forma de realización, la  $T_m$  del dúplex entre el oligonucleótido bloqueante y la secuencia de ácidos nucleicos no diana del miembro de biblioteca es más elevada que la  $T_m$  de un dúplex de la secuencia de ácidos nucleicos no diana y su complemento exacto.
- 30 En otras formas de realización, el oligonucleótido bloqueante presenta una tasa de asociación a una secuencia de ácidos nucleicos no diana de por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que es superior a la tasa de asociación de la secuencia de ácidos nucleicos no diana a un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana.
- 35 En una forma de realización, la tasa de asociación más elevada es entre aproximadamente 2 veces y más de 10 veces más elevada que la de la secuencia de ácidos nucleicos no diana con el ácido nucleico de fondo (por ejemplo, 2, 4, 6, 8 o 10 veces, o más, más elevada).
- 40 En todavía otras formas de realización, el oligonucleótido bloqueante presenta una tasa de disociación respecto a la secuencia de ácidos nucleicos no diana de por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que es inferior a la tasa de disociación de la secuencia de ácidos nucleicos no diana respecto a un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana.
- 45 En una forma de realización, la tasa de disociación es entre aproximadamente 2 veces y más de 10 veces más elevada que la de la secuencia de ácidos nucleicos no diana con el ácido nucleico de fondo (por ejemplo, 2, 4, 6, 8 o 10 veces, o más más elevada).
- 50 En una forma de realización, la longitud del oligonucleótido bloqueante resulta en un incremento de la interacción de unión del oligonucleótido bloqueante de la secuencia de ácidos nucleicos no diana del miembro de biblioteca (por ejemplo, la secuencia adaptadora), respecto al ácido nucleico de fondo.
- 55 En una forma de realización, el dúplex formado entre el oligonucleótido bloqueante y la secuencia de ácidos nucleicos no diana de por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) es más largo que el dúplex formado entre la secuencia no diana y su complemento, por ejemplo entre las cadenas de Watson y Crick de un adaptador de doble cadena. En formas de realización, el dúplex formado entre un oligonucleótido bloqueante y la secuencia de ácidos nucleicos no diana es por lo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 nucleótidos más largo que el dúplex formado entre la secuencia no diana y su complemento, por ejemplo entre las cadenas de Watson y Crick de un adaptador de doble cadena.
- 60 En una forma de realización, el oligo bloqueante comprende uno o más nucleótidos no naturales. En formas de realización, un dúplex formado entre el oligonucleótido bloqueante que presenta nucleótidos no naturales y la secuencia de ácidos nucleicos no diana de por lo menos X miembros de biblioteca (en el que X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) presenta un valor de un parámetro relacionado con la interacción de unión (por ejemplo, la afinidad, la tasa de asociación, la inversa de la tasa de disociación o la  $T_m$ ) que es superior al valor de la secuencia de ácidos nucleicos no diana respecto a un ácido nucleico de fondo, por ejemplo otras secuencias de ácidos nucleicos no diana complementarias. Entre los oligonucleótidos no naturales ejemplificativos se incluyen
- 65

nucleótidos de ADN o ARN modificado. Entre los nucleótidos modificados ejemplificativos (por ejemplo, nucleótidos de ARN o ADN modificado) se incluyen, aunque sin limitarse a ellos, un ácido nucleico bloqueado (ANB), en el que la fracción de ribosa de un nucleótido de ANB se modifica con un puente adicional que conecta el oxígeno 2' y el carbono 4'; ácido péptido nucleico (APN), por ejemplo un APN compuesto de unidades repetidas de N-(2-aminoetil)-glicina unidas mediante enlaces peptídicas; un oligonucleótido de ADN o ARN modificado para capturar regiones de bajo contenido en GC; un ácido nucleico bicíclico (ANBi), un oligonucleótido entrecruzado, una 5-metil-desoxicitidina modificada y 2,6-diaminopurina. Se conocen en la técnica otros nucleótidos de ADN y ARN modificados.

En una forma de realización, el oligonucleótido bloqueante es de ARN o comprende ARN y la secuencia de ácidos nucleicos no diana, por ejemplo, un adaptador, es de ADN o comprende ADN. En una forma de realización, la secuencia de ácidos nucleicos no diana es una secuencia común a una pluralidad de miembros de biblioteca, por ejemplo, por lo menos X miembros de biblioteca (en donde X es igual a 2, 5, 10, 20, 50, 100 o 200), por ejemplo una secuencia que puede utilizarse para la amplificación, por ejemplo la amplificación por PCR de puente.

En una forma de realización, la secuencia de ácidos nucleicos no diana es una secuencia que puede utilizarse para la amplificación, por ejemplo la amplificación por PCR de puente, y el ácido nucleico de fondo es una segunda secuencia no diana.

En una forma de realización, la sonda de captura es de ADN (por ejemplo, no de ARN). En formas de realización, la sonda de captura incluye uno o más oligonucleótidos de ADN (por ejemplo, un oligonucleótido de ADN natural o no natural).

En una forma de realización, la sonda de captura es de ARN. En formas de realización, la sonda de captura incluye uno o más oligonucleótidos de ARN (por ejemplo, un oligonucleótido de ARN natural o no natural).

En una forma de realización, un oligonucleótido bloqueante presenta 20-80, 30-80, 40-80, 50-80, 70-80, 30-75, 30-65, 30-55, 30-45, 40-70, 40-60, 40-50, 50-60, 50-70, 60-70 nucleótidos de longitud. En una forma de realización, la inserción de biblioteca es de 50-10,000, 50-1,000, 50-500, 50-200, 77-150, o 100-150 nucleótidos de longitud, tal como se indica en otro sitio de la presente memoria.

En otro aspecto, la invención proporciona una preparación que comprende una pluralidad de oligonucleótidos bloqueantes, por ejemplo tal como se indica en la presente memoria. En una forma de realización, la preparación comprende además uno o ambos de: una pluralidad de miembros de biblioteca, por ejemplo tal como se indica en la presente memoria, y una sonda de captura, por ejemplo tal como se indica en la presente memoria.

En otro aspecto, la invención proporciona un kit que comprende una pluralidad de oligonucleótidos bloqueantes, por ejemplo tal como se indica en la presente memoria. En una forma de realización, el kit comprende además uno o ambos de: una pluralidad de miembros de biblioteca, por ejemplo tal como se indica en la presente memoria, y una sonda de captura, por ejemplo tal como se indica en la presente memoria. En formas de realización, los componentes se proporcionan en recipientes separados, por ejemplo, el oligonucleótido bloqueante se proporciona en n recipiente y otro componente, por ejemplo un amortiguador o una pluralidad de miembros de biblioteca, por ejemplo, tal como se indica en la presente memoria, o una sonda de captura, por ejemplo, tal como se indica en la presente memoria, se proporciona en uno o más recipientes diferentes.

En otro aspecto, la invención proporciona un método para reducir la selección de ácidos nucleicos fuera de diana descrito en la presente memoria, en combinación con otro método descrito en la presente memoria, por ejemplo un método de secuenciación descrito en la presente memoria, un método de alineación descrito en la presente memoria, un método de llamada de mutaciones descrito en la presente memoria o un método que utiliza un señuelo descrito en la presente memoria.

La selección de fuera de diana también puede minimizarse mediante la utilización de secuencias no diana que sean suficientemente cortas para que un dúplex de secuencias no diana sea menos estable que un dúplex de secuencia de inserto de un miembro de biblioteca y la sonda de captura. De esta manera, en otro aspecto, la invención proporciona un método para reducir la selección de ácidos nucleicos fuera de diana, por ejemplo en fase sólida o la hibridación en solución. El método puede utilizarse en la selección de miembros de biblioteca para la posterior secuenciación.

El método comprende:

- (a) opcionalmente, adquirir una biblioteca que comprende una pluralidad de miembros diana, por ejemplo miembros de ácido nucleico diana (por ejemplo, ADN o ARN), en la que uno o más de los miembros diana comprende una secuencia de inserto (por ejemplo, un segmento de un gen de interés) y una secuencia de ácidos nucleicos no diana (por ejemplo, una secuencia adaptadora), y

(b) poner en contacto la biblioteca con una sonda de captura, por ejemplo un grupo de señuelos o una pluralidad de grupos de señuelos,

en el que las secuencias no diana son suficientemente cortas para que el valor para un parámetro relacionado con la interacción de unión entre la secuencia de inserto y la sonda de captura sea más elevada que el valor para la secuencia de ácidos nucleicos no diana y su complemento, minimizando de esta manera la selección de secuencias fuera de diana.

En una forma de realización, el método comprende además proporcionar miembros seleccionados de biblioteca (en ocasiones denominados en la presente memoria "captura de biblioteca").

En una forma de realización, el método comprende además separar los miembros seleccionados de biblioteca respecto de la sonda de captura.

En una forma de realización, el método comprende además secuenciar la inserción de un miembro seleccionado de biblioteca, por ejemplo secuenciando las inserciones de por lo menos 2, 5, 10, 15, 20, 30 o 50 alteraciones de genes o ácidos nucleicos, por ejemplo alteraciones de genes o ácidos nucleicos indicadas en la presente memoria.

En una forma de realización, el valor para un parámetro relacionado con la interacción de unión puede ser un valor de afinidad, tasa de asociación, inversa de la tasa de disociación o temperatura de fusión del ácido nucleico (por ejemplo,  $T_m$ , la temperatura a la que la mitad de las cadenas de ADN se encuentran en el estado de doble hélice y la mitad se encuentra en el estado de enrollado aleatorio).

En una forma de realización, un miembro de biblioteca comprende una inserción, por ejemplo un intervalo subgenómico, y una secuencia no diana, por ejemplo una secuencia común a una pluralidad de miembros de biblioteca. En una forma de realización, las inserciones son secuencias subgenómicas, por ejemplo de ácidos nucleicos procedentes de una muestra tumoral, y la secuencia no diana es una secuencia no natural o una secuencia no presente en las secuencias subgenómicas, por ejemplo una etiqueta de amplificación o etiqueta de código de barras.

En una forma de realización, los miembros de biblioteca o miembros seleccionados de biblioteca incluyen intervalos subgenómicos de por lo menos 2, 5, 10, 15, 20, 30 o 50 alteraciones de genes o ácidos nucleicos, por ejemplo alteraciones de genes o ácidos nucleicos indicadas en la presente memoria.

En una forma de realización, una pluralidad de miembros de biblioteca, o miembros seleccionados de biblioteca, por ejemplo en X (en donde X es igual a 2, 5, 10, 20, 50, 100, 200 o más) miembros de biblioteca, o miembros seleccionados de biblioteca, presenta una primera secuencia no diana en el extremo 5' de la inserción y una segunda secuencia no diana en el extremo 3' de la inserción.

En una forma de realización, la secuencia no diana incluye una secuencia no diana que está presente en una pluralidad de secuencias no diana, por ejemplo una secuencia para la amplificación y una secuencia no diana que es única, por ejemplo un código de barras. Típicamente, algunos, más sustancialmente la totalidad o la totalidad de los miembros de la biblioteca, incluirá una secuencia no diana común. En formas de realización, la biblioteca o los miembros seleccionados de biblioteca, comprende por lo menos X miembros (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que presentan una secuencia no diana común.

En una forma de realización, la secuencia de inserción forma un dúplex con la sonda de captura para por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más), en donde el dúplex presenta una  $T_m$  que es superior a la  $T_m$  de un dúplex formado por una secuencia de ácidos nucleicos no diana con un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana. En una forma de realización, la temperatura de fusión de ácidos nucleicos más elevada del dúplex de secuencia de inserción/sonda de captura es de entre aproximadamente 5°C y 25°C o superior (por ejemplo, 5°C, 10°C, 15°C, 20°C, 25°C o superior). En una forma de realización, la  $T_m$  del dúplex entre la secuencia de inserción/sonda de captura es más elevada que la  $T_m$  de un dúplex de la secuencia de ácidos nucleicos no diana y su complemento exacto.

En otras formas de realización, la secuencia de inserción presenta una tasa de asociación a la sonda para por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que es superior a la tasa de asociación de la secuencia de ácidos nucleicos no diana a un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana. En una forma de realización, la tasa de asociación más elevada es entre aproximadamente 2 veces y más de 10 veces más elevada que la de la secuencia de ácidos nucleicos no diana con el ácido nucleico de fondo (por ejemplo, 2, 4, 6, 8 o 10 veces, o más, más elevada).

En todavía otras formas de realización, la secuencia de inserción presenta una tasa de disociación respecto a la sonda de captura para por lo menos X miembros de biblioteca (en donde X es igual a 1, 2, 5, 10, 20, 50, 100, 200 o más) que es inferior a la tasa de disociación de la secuencia de ácidos nucleicos no diana respecto a un ácido nucleico de fondo, por ejemplo el complemento de la secuencia no diana. En una forma de realización, la tasa de

disociación es entre aproximadamente 2 veces y más de 10 veces más elevada que la de la secuencia de ácidos nucleicos no diana con el ácido nucleico de fondo (por ejemplo, 2, 4, 6, 8 o 10 veces, o más más elevada).

5 En una forma de realización, la secuencia de ácidos nucleicos no diana es una secuencia que puede utilizarse para la amplificación, por ejemplo la amplificación por PCR de puente, y el ácido nucleico de fondo es una segunda secuencia no diana.

10 En una forma de realización, la sonda de captura es de ADN (por ejemplo, no de ARN). En una forma de realización, la sonda de captura es de ARN.

En una forma de realización, la inserción de biblioteca es de 50-10,000, 50-1,000, 50-500, 50-200, 77-150 o 100-150 nucleótidos de longitud, tal como se indica en otro sitio de la presente memoria.

15 En una forma de realización, el método comprende además la utilización de un oligonucleótido bloqueante, tal como se indica en la presente memoria.

Se describen en la presente memoria características y formas de realización de la invención adicionales.

20 En una forma de realización, el método comprende, además:

(c) adquirir una lectura para un intervalo subgenómico de un miembro tumoral de dicha biblioteca o captura de biblioteca, por ejemplo mediante secuenciación, por ejemplo con un método de secuenciación de nueva generación,

25 (d) alinear dicha lectura, y

(e) asignar un valor de nucleótido (por ejemplo, llamando una mutación, por ejemplo, mediante un método bayesiano) de dicha lectura para una posición nucleótida preseleccionada, por ejemplo para una posición nucleótida preseleccionada en cada uno de una pluralidad de intervalos subgenómicos, por ejemplo cada uno de una pluralidad de genes,

analizando de esta manera dicha muestra.

En una forma de realización:

35 (i) cada una de las X posiciones nucleótidas se analiza bajo un conjunto único de condiciones para una o una combinación de etapas (b), (c), (d) o (e) (en la que único se refiere a diferente de otros X-1 conjuntos de condiciones y en el que X es por lo menos 2, 5, 10, 20, 30, 40, 50, 100, 200, 300 o 500). Por ejemplo, un primer grupo de condiciones, por ejemplo, un grupo de condiciones indicado en la presente memoria, se utiliza para una primera posición nucleótida, por ejemplo en un primer intervalo subgenómico o gen, y un segundo conjunto de condiciones, por ejemplo un segundo conjunto de condiciones indicado en la presente memoria, se utiliza para una segunda posición nucleótida, por ejemplo en un segundo intervalo subgenómico o gen,

45 (ii) para cada una de X posiciones nucleótidas, respondedoras a una característica, por ejemplo una característica indicada en la presente memoria, de una alteración preseleccionada, por ejemplo una mutación, que puede ocurrir en la posición nucleótida, la posición nucleótida se analiza bajo un conjunto único de condiciones (en el que único se refiere a diferente de otros X-1 conjuntos de condiciones y en el que X es por lo menos 2, 5, 10, 20, 30, 40, 50, 100, 200, 300 o 500). Por ejemplo, respondedoras a una característica, por ejemplo una característica indicada en la presente memoria, de una alteración preseleccionada, por ejemplo una mutación, que puede ocurrir en una posición nucleótida en un primer intervalo subgenómico, la posición nucleótida se analiza bajo un primer conjunto de condiciones y responde a una característica, por ejemplo una característica indicada en la presente memoria, de una alteración preseleccionada, por ejemplo una mutación, que puede ocurrir en una posición nucleótida en un segundo intervalo subgenómico, la posición nucleótida se analiza bajo un segundo conjunto de condiciones,

(iii) en el que dicho método se lleva a cabo en una muestra, por ejemplo, una muestra tumoral conservada, bajo condiciones que permiten 95%, 98% o 99% de sensibilidad o especificidad para las posiciones nucleótidas en por lo menos 2, 5, 10, 20, 50 o 100 intervalos subgenómicos, por ejemplo genes, o

60 (iv) en el que el método comprende una o más o todas de:

a) secuenciación de un primer intervalo subgenómico, proporcionando aproximadamente 500x o una profundidad de secuenciación más elevada, por ejemplo para secuenciar una mutación presente en no más de 5% de las células de la muestra,

65

- b) secuenciación de un segundo intervalo subgenómico, proporcionando una profundidad de secuenciación de aproximadamente 200x o más elevada, por ejemplo de aproximadamente 200x a aproximadamente 500x, por ejemplo para secuenciar una mutación presente en no más de 10% de las células de la muestra,
- c) secuenciación de un tercer intervalo subgenómico para proporcionar una profundidad de secuenciación de aproximadamente 10-100x, por ejemplo, para secuenciar uno o más intervalos subgenómicos (por ejemplo, exones) seleccionados de: a) un polimorfismo de nucleótido único (SNP) farmacogenómico (PGx) que puede explicar la capacidad de un paciente de metabolizar diferentes fármacos, o b) un SNP genómico que puede utilizarse para identificar de manera única (por ejemplo, la huella génica) un paciente,
- d) secuenciación de un cuarto intervalo subgenómico, proporcionando una profundidad de secuenciación de aproximadamente 5-50x, por ejemplo para detectar un punto de rotura estructural, tal como una traslocación genómica o una mutación indel. Por ejemplo, la detección de un punto de rotura intrónico requiere una profundidad que abarque 5x50x pares de secuencia para garantizar una fiabilidad de detección elevada. Dichos conjuntos de señuelos pueden utilizarse para detectar, por ejemplo, genes de cáncer con tendencia a traslocación/mutación indel, o
- e) secuenciación de un quinto intervalo subgenómico, proporcionando una profundidad de secuenciación de aproximadamente 0,1-300x, por ejemplo para detectar cambios en el número de copia. En una forma de realización, la profundidad de secuenciación está comprendida entre aproximadamente 0,1 y 10x para detectar cambios en el número de copia. En otras formas de realización, la profundidad de secuenciación está comprendida entre aproximadamente 100 y 300x para detectar SNP/loci genómicos que se utilizan para evaluar las ganancias/pérdidas de número de copia de ADN genómico o de pérdida de heterocigosidad (PDH).

Entre los primeros y segundos conjuntos ejemplificativos de condiciones se incluyen aquellas en las que:

- se utiliza un primer conjunto de señuelos para el primer intervalo subgenómico y se utiliza un segundo conjunto de señuelos para el segundo intervalo subgenómico,
- se aplica un primer método de alineación a una lectura para el primer intervalo subgenómico y un segundo método de alineación a una lectura para el segundo intervalo subgenómico,
- se aplica un primer método de llamada de mutaciones a una posición nucleótida del primer intervalo subgenómico y un segundo método de llamada de mutaciones a una posición nucleótida del segundo intervalo subgenómico.
- En una forma de realización:
- se analiza una primera posición nucleótida con un primer conjunto de condiciones de señuelo, un primer método de alineación y un primer método de llamada de mutaciones,
- se analiza una segunda posición nucleótida con dicho primer conjunto de condiciones de señuelo, un segundo método de alineación y dicho primer método de llamada de mutaciones,
- se analiza una tercera posición nucleótida con dicho primer conjunto de condiciones de señuelo, dicho primer método de alineación y un segundo método de llamada de mutaciones,
- proporcionando tres posiciones nucleótidas, cada una analizada bajo condiciones únicas, respecto a las otras dos.

En una forma de realización, las condiciones comprenden aquellas en las que:

- se utiliza un primer conjunto de señuelos para el primer intervalo subgenómico y se utiliza un segundo conjunto de señuelos para el segundo intervalo subgenómico,
- se aplica un primer método de alineación a una lectura para el primer intervalo subgenómico y se aplica un segundo método de alineación a una lectura para el segundo intervalo subgenómico, o
- se aplica un primer método de llamada de mutaciones a una posición nucleótida del primer intervalo subgenómico y un segundo método de llamada de mutaciones a una posición nucleótida del segundo intervalo subgenómico.

Entre las características ejemplificativas se incluyen:

- (i) el gen, o tipo de gen, en el que se localiza la alteración, por ejemplo un oncogén o supresor tumoral, un gen o tipo de gen caracterizado por un gen preseleccionado o variante o tipo de variante, por ejemplo una mutación, o mediante una mutación de una frecuencia preseleccionada, u otro gen o tipo de gen indicado en la presente memoria,
- (ii) el tipo de alteración, por ejemplo, una sustitución, inserción, delección o traslocación,
- (iii) el tipo de muestra, por ejemplo, una muestra de FFPE, que se analiza para la alteración,
- (iv) secuencia en o próxima a dicha posición nucleótida de la alteración que se evalúa, por ejemplo, una secuencia que puede afectar a la tendencia esperada a la alineación incorrecta del intervalo subgenómico, por ejemplo la presencia de secuencias repetidas en o próximas a la posición nucleótida,
- (v) una expectativa previa (por ejemplo, de la literatura) de observación de una lectura que muestra la alteración, por ejemplo, mutación, por ejemplo en un tumor de tipo preseleccionado,
- (vi) la probabilidad de observar una lectura que muestra la alteración debido a error en la llamada de bases únicamente, o
- (vii) una profundidad de secuenciación preseleccionada deseada para detectar la alteración.

En una forma de realización, la característica es diferente de la identidad del nucleótido que se secuencia, es decir, la característica no es si la secuencia es a o t.

En una forma de realización, se analizaron los intervalos subgenómicos de por lo menos X genes, por ejemplo por lo menos X genes de las Tablas 1 y 1A, por ejemplo genes que presentan la anotación de prioridad 1 en las Tablas 1 y 1A se analizan bajo diferentes condiciones, y X es igual a 2, 3, 4, 5, 10, 15, 20 o 30.

En una forma de realización, el método comprende uno o más de los siguientes:

- (i) el método, por ejemplo, (b) del método anterior, comprende la utilización de un conjunto de señuelos indicado en la presente memoria,
- (ii) el método, por ejemplo, (c) del método anterior, comprende adquirir lecturas para un conjunto o grupo de intervalos subgenómicos o de un conjunto o grupo de genes indicado en la presente memoria,
- (iii) el método, por ejemplo, (d) del método anterior, comprende la utilización de una pluralidad de métodos de alineación indicados en la presente memoria,
- (iv) el método, por ejemplo, (e) del método anterior, comprende la utilización de una pluralidad de métodos para asignar un valor de nucleótido a una posición nucleótida preseleccionada, indicada en la presente memoria, o
- (v) el método comprende asignar un valor de nucleótido a un conjunto de intervalos subgenómicos indicado en la presente memoria.

En una forma de realización, el método incluye: (i) y uno, dos, tres o la totalidad de (ii)-(v). En una forma de realización, el método incluye: (ii) y uno, dos, tres o la totalidad de (i) y (iii)-(v). En una forma de realización, el método incluye: (iii) y uno, dos, tres o la totalidad de (i), (ii) (iv) y (v). En una forma de realización, el método incluye: (iv) y uno, dos, tres o la totalidad de (i)-(iii) y (v). En una forma de realización, el método incluye: (v) y uno, dos, tres o la totalidad de (i)-(iv).

#### Señuelos

Los métodos indicados en la presente memoria proporcionan la selección y/o secuenciación de un gran número de genes y productos génicos de muestras, por ejemplo, muestras tumorales, de uno o más sujetos mediante la selección apropiada de señuelos, por ejemplo señuelos para la utilización en hibridación en solución, para la selección de ácidos nucleicos diana que deben secuenciarse. La eficiencia de selección para diversos intervalos subgenómicos, o clases de los mismos, se hace corresponder con conjuntos de señuelos que presentan una eficiencia de selección preseleccionada. Tal como se utiliza en la presente sección, la "eficiencia de selección" se refiere al nivel o profundidad de cobertura de secuencia ajustado según uno o más intervalos subgenómicos diana.

De esta manera, un método (por ejemplo, el miembro (b) del método recitado anteriormente) comprende poner en contacto la biblioteca con una pluralidad de señuelos para proporcionar miembros seleccionados (por ejemplo, una captura de biblioteca). En determinadas formas de realización, el método comprende poner en contacto la biblioteca con una pluralidad, por ejemplo, por lo menos dos, tres, cuatro o cinco señuelos o conjuntos de señuelos,

en la que cada señuelo o conjunto de señuelos de dicha pluralidad presenta una eficiencia preseleccionada única (respecto a los otros conjuntos de señuelos en la pluralidad) de selección. Por ejemplo, cada señuelo o conjunto de señuelos único proporciona una profundidad de secuenciación única. La expresión "conjunto de señuelos", tal como se utiliza en la presente memoria, se refiere colectivamente a un señuelo o a una pluralidad de moléculas de señuelo.

En una forma de realización, la eficiencia de selección de un primer conjunto de señuelos en la pluralidad difiere de la eficiencia de un segundo conjunto de señuelos en la pluralidad en un factor de por lo menos 2. En una forma de realización, el primer y segundo conjuntos de señuelos proporcionan una profundidad de secuenciación que difiere en un factor de por lo menos 2.

En otra forma de realización, el método comprende poner en contacto un o una pluralidad de los conjuntos de señuelos siguientes con la biblioteca:

- a) un conjunto de señuelos que selecciona suficientes miembros que comprenden un intervalo subgenómico que proporciona una profundidad de secuenciación de aproximadamente 500x o superior, por ejemplo, para secuenciar una mutación presente en no más de 5% de las células de la muestra,
- b) un conjunto de señuelos que selecciona suficientes miembros que comprenden un intervalo subgenómico para proporcionar una profundidad de secuenciación de aproximadamente 200x o superior, por ejemplo de aproximadamente 200x a aproximadamente 500x, por ejemplo para secuenciar una mutación presente en no más de 10% de las células de la muestra,
- c) un conjunto de señuelos que selecciona suficientes miembros que comprenden un intervalo subgenómico para proporcionar una profundidad de secuenciación de aproximadamente 10-100x, por ejemplo, para secuenciar uno o más intervalos subgenómicos (por ejemplo, exones) seleccionados de: a) un polimorfismo de nucleótido único (SNP) farmacogenómico (PGx) que puede explicar la capacidad de un paciente de metabolizar diferentes fármacos, o b) un SNP genómico que puede utilizarse para identificar de manera única (por ejemplo, la huella génica) un paciente,
- d) un conjunto de señuelos que selecciona suficientes miembros que comprenden un intervalo subgenómico para proporcionar una profundidad de secuenciación de aproximadamente 5-50x, por ejemplo, para detectar un punto de rotura estructural, tal como una traslocación genómica o una mutación indel. Por ejemplo, la detección de un punto de rotura intrónico requiere una profundidad que abarque 5x50x pares de secuencia para garantizar una fiabilidad de detección elevada. Dichos conjuntos de señuelos pueden utilizarse para detectar, por ejemplo, genes de cáncer con tendencia a traslocación/mutación indel, o
- e) un conjunto de señuelos que selecciona suficientes miembros que comprenden un intervalo subgenómico para proporcionar una profundidad de secuenciación de aproximadamente 0,1-300x, por ejemplo para detectar cambios en el número de copia.

En una forma de realización, la profundidad de secuenciación está comprendida entre aproximadamente 0,1 y 10x para detectar cambios en el número de copia. En otras formas de realización, la profundidad de secuenciación está comprendida entre aproximadamente 100 y 300x para detectar SNP/loci genómicos que se utilizan para evaluar las ganancias/pérdidas de número de copia de ADN genómico o de pérdida de heterocigosidad (PDH). Dichos conjuntos de señuelos pueden utilizarse para detectar, por ejemplo, genes de cáncer con tendencia a amplificación/delección. El nivel de la profundidad de secuenciación tal como se utiliza en la presente memoria (por ejemplo, el factor X de profundidad de secuenciación) se refiere al nivel de cobertura de lecturas (por ejemplo, lecturas únicas), tras la detección y eliminación de lecturas duplicadas, por ejemplo lecturas duplicadas de PCR.

En una forma de realización, el conjunto de señuelos selecciona un intervalo subgenómico que contiene una o más reorganizaciones, por ejemplo, un intrón que contiene una reorganización genómica. En dichas formas de realización, el conjunto de señuelos se diseña de manera que las secuencias repetitivas están enmascaradas para incrementar la eficiencia de selección. En aquellas formas de realización en las que la reorganización presenta una secuencia de unión conocida, pueden diseñarse conjuntos de señuelos complementarios de la secuencia de unión a fin de incrementar la eficiencia de selección.

En formas de realización, el método comprende la utilización de señuelos diseñados para capturar dos o más categorías diana diferentes, presentando cada categoría una estrategia de diseño del señuelo diferente. En formas de realización, el método de captura de híbridos y las composiciones dadas a conocer en la presente memoria capturan un subconjunto definido de secuencias diana (por ejemplo, miembros diana) y proporcionan una cobertura homogénea de la secuencia diana, minimizando simultáneamente la cobertura fuera de dicho subconjunto. En una forma de realización, las secuencias diana incluyen el exoma completo fuera del ADN genómico, o un subconjunto seleccionado del mismo. Los métodos y composiciones dados a conocer en la presente memoria proporcionan diferentes conjuntos de señuelos para conseguir diferentes profundidades y patrones de cobertura para secuencias complejas de ácidos nucleicos diana (por ejemplo, bibliotecas de ácidos nucleicos).

En una forma de realización, el método comprende proporcionar miembros seleccionados de una biblioteca de ácidos nucleicos (por ejemplo, una captura de biblioteca). El método incluye:

- 5 proporcionar una biblioteca (por ejemplo, una biblioteca de ácidos nucleicos) que comprende una pluralidad de miembros, por ejemplo, miembros ácidos nucleicos diana (por ejemplo, incluyendo una pluralidad de miembros tumorales, miembros de referencia y/o miembros PGx),
  - 10 poner en contacto la biblioteca, por ejemplo, en una reacción de base solución, con una pluralidad de señuelos (por ejemplo, señuelos oligonucleótidos) para formar una mezcla de hibridación que comprende una pluralidad de híbridos de señuelo/miembro,
  - 15 separar la pluralidad de híbridos de señuelo/miembro de dicha mezcla de hibridación, por ejemplo, mediante la puesta en contacto de dicha mezcla de hibridación con una entidad de unión que permita la separación de dicha pluralidad de híbridos de señuelo/miembro,
- proporcionando de esta manera una captura de biblioteca (por ejemplo, un subgrupo seleccionado o enriquecido de moléculas de ácidos nucleicos de la biblioteca), en la que la pluralidad de señuelos incluye de señuelos incluye dos o más de los siguientes:
- 20 a) un primer conjunto de señuelos que selecciona una diana de nivel elevado (por ejemplo, uno o más miembros tumorales que incluyen un intervalo subgenómico, tal como un gen, un exón o una base) para la que se requiere la cobertura de mayor profundidad para conseguir un nivel elevado de sensibilidad a una alteración (por ejemplo, una o más mutaciones) que aparecen a baja frecuencia, por ejemplo,
  - 25 aproximadamente 5% o menos (es decir, 5% de las células de la muestra aloja la alteración en su genoma). En una forma de realización, el primer conjunto de señuelos selecciona (por ejemplo, es complementaria) un miembro tumoral que incluye una alteración (por ejemplo, una mutación puntual) que requiere una profundidad de secuenciación de aproximadamente 500x o superior,
  - 30 b) un segundo conjunto de señuelos que selecciona una diana de nivel intermedio (por ejemplo, uno o más miembros tumorales que incluyen un intervalo subgenómico, tal como un gen, un exón o una base) para la que se requiere cobertura elevada para conseguir un nivel elevado de sensibilidad a una alteración (por ejemplo, una o más mutaciones) que aparece a una frecuencia más elevada que la diana de nivel elevado en a), por ejemplo una frecuencia de aproximadamente 10% (es decir, 10% de las células de la muestra aloja la alteración en su genoma). En una forma de realización, el segundo conjunto de señuelos selecciona (por ejemplo, es complementaria) un miembro tumoral que incluye una alteración (por ejemplo, una mutación puntual) que requiere una profundidad de secuenciación de aproximadamente 200x o superior,
  - 35 c) un tercer conjunto de señuelos que selecciona una diana de nivel bajo (por ejemplo, uno o más miembros PGx que incluyen un intervalo subgenómico, tal como un gen, un exón o una base) para la que se requiere cobertura de nivel bajo-intermedio para conseguir un nivel elevado 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 fiabilidad de detección elevada. En una forma de realización, el tercer conjunto de señuelos selecciona uno o más intervalos subgenómicos (por ejemplo,
  - 40 exones) que se seleccionan de: a) polimorfismo de nucleótido único (SNP) farmacogenómico (PGx) que puede explicar la capacidad de un paciente de metabolizar diferentes fármacos, o b) un SNP genómico que puede utilizarse para identificar de manera única (por ejemplo, la huella génica) un paciente,
  - 45 d) un cuarto conjunto de señuelos que selecciona una primera diana intrónica (por ejemplo, un miembro que incluye una secuencia de intrón) para la que se requiere una cobertura de nivel bajo-intermedio, por ejemplo para detectar un punto de rotura estructural, tal como una traslocación genómica o una mutación indel. Por ejemplo, la detección de un punto de rotura intrónico requiere una profundidad que abarque 5x50x pares de secuencia para garantizar una fiabilidad de detección elevada. Dichos cuartos conjuntos de señuelos pueden utilizarse para detectar, por ejemplo, genes de cáncer con tendencia a traslocación/mutación indel,
  - 50 e) un quinto conjunto de señuelos que selecciona una segunda diana intrónica (por ejemplo, un miembro intrónico) para la que se requiere una cobertura escasa para mejorar la capacidad de detectar los cambios del número de copia. Por ejemplo, la detección de una delección de una copia de varios exones terminales requiere una cobertura de 0,1-300x para garantizar una elevada fiabilidad de detección. En una forma de realización, la profundidad de cobertura está comprendida entre aproximadamente 0,1 y 10x para detectar los cambios en el número de copia. En otras formas de realización, la profundidad de cobertura está comprendida entre aproximadamente 100 y 300x para detectar SNP/loci genómicos que se utilizan para evaluar las ganancias/pérdidas de número de copia de ADN genómico o de pérdida de heterocigosidad (PDH). Dichos conjuntos de señuelos pueden utilizarse para detectar, por ejemplo, genes de cáncer con
  - 55 tendencia a amplificación/delección.
  - 60
  - 65

Puede utilizarse cualquier combinación de dos, tres, cuatro o más de los conjuntos de señuelos anteriormente indicados, por ejemplo una combinación del primer y segundo conjuntos de señuelos, primer y tercer conjuntos de señuelos, primer y cuarto conjuntos de señuelos, primer y quinto conjuntos de señuelos, segundo y tercer conjuntos de señuelos, segundo y cuarto conjuntos de señuelos, segundo y quinto conjuntos de señuelos, tercer y cuarto conjuntos de señuelos, tercer y quinto conjuntos de señuelos, cuarto y quinto conjuntos de señuelos; primer, segundo y tercer conjuntos de señuelos; primer, segundo y cuarto conjuntos de señuelos; primer, segundo y quinto conjuntos de señuelos; primer, segundo, tercer, cuarto conjuntos de señuelos; y de esta manera sucesivamente.

En una forma de realización, cada uno de los primer, segundo, tercer, cuarto o quinto conjuntos de señuelos, presenta una eficiencia preseleccionada de selección (por ejemplo, captura). En una forma de realización, el valor de la eficiencia de selección es igual para por lo menos dos, tres, cuatro de la totalidad de cinco señuelos según a)-e). En otras formas de realización, el valor de la eficiencia de selección es diferente para por lo menos dos, tres, cuatro de la totalidad de cinco señuelos según a)-e). En algunas formas de realización, por lo menos dos, tres, cuatro o la totalidad de cinco conjuntos de señuelos presentan un valor de eficiencia preseleccionado que difiere.

Por ejemplo, un valor para la eficiencia de selección seleccionado de uno o más de:

- (i) la primera eficiencia preseleccionada presenta un valor para la primera eficiencia de selección que es una profundidad de secuenciación de por lo menos aproximadamente 500x o superior (por ejemplo, presenta un valor de eficiencia de selección que es superior a la segunda, tercera, cuarta o quinta eficiencias preseleccionadas de selección (por ejemplo, aproximadamente 2-3 veces superior al valor de la segunda eficiencia de selección; aproximadamente 5-6 veces superior al valor de la tercera eficiencia de selección; aproximadamente 10 veces superior al valor de la cuarta eficiencia de selección; aproximadamente 50 a 5000 veces superior al valor de la quinta eficiencia de selección);
- (ii) la segunda eficiencia preseleccionada presenta un valor para la segunda eficiencia de selección que es una profundidad de secuenciación de por lo menos aproximadamente 200x o superior (por ejemplo, presenta un valor de eficiencia de selección que es superior a la tercera, cuarta o quinta eficiencias preseleccionadas de selección (por ejemplo, aproximadamente 2 veces superior al valor de la tercera eficiencia de selección; aproximadamente 4 veces superior al valor de la cuarta eficiencia de selección; aproximadamente 20 a 2000 veces superior al valor de la quinta eficiencia de selección);
- (iii) la tercera eficiencia preseleccionada presenta un valor para la tercera eficiencia de selección que es una profundidad de secuenciación de por lo menos aproximadamente 100x o superior (por ejemplo, presenta un valor de eficiencia de selección que es superior a la cuarta o quinta eficiencia preseleccionada de selección (por ejemplo, aproximadamente 2 veces superior al valor de la cuarta eficiencia de selección; aproximadamente 10 a 1000 veces superior al valor de la quinta eficiencia de selección);
- (iv) la cuarta eficiencia preseleccionada presenta un valor para la cuarta eficiencia de selección que es una profundidad de secuenciación de por lo menos aproximadamente 50x o superior (por ejemplo, presenta un valor de eficiencia de selección que es superior a la quinta eficiencia preseleccionada de selección (por ejemplo, aproximadamente 50 a 500 veces superior al valor de la quinta eficiencia de selección), o
- (v) la quinta eficiencia preseleccionada presenta un valor de la quinta eficiencia de selección que es por lo menos 10x a 0,1x la profundidad de secuenciación.

En determinadas formas de realización, el valor de la eficiencia de selección está modificado por uno o más de: representación diferencial de diferentes conjuntos de señuelos, solapamiento diferencial de subconjuntos de señuelos, parámetros de señuelo diferenciales, mezcla de diferentes conjuntos de señuelos y/o utilización de diferentes tipos de conjuntos de señuelos.

Por ejemplo, puede ajustarse una variación de la eficiencia de selección (por ejemplo, la cobertura de secuencia relativa de cada categoría de conjunto de señuelos/diana) mediante la alteración de uno o más de:

- (i) representación diferencial de diferentes conjuntos de señuelos. El diseño del conjunto de señuelos para capturar una diana dada (por ejemplo, un miembro diana) puede incluirse en un número mayor/menor de copias para potenciar/reducir las profundidades relativas de cobertura de la diana,
- (ii) solapamiento diferencial de subconjuntos de señuelos. El diseño del conjunto de cebadores para capturar una diana dada (por ejemplo, un miembro diana) puede incluir un solapamiento más largo o más corto entre señuelos vecinos para potenciar/reducir las profundidades relativas de cobertura de la diana,

(iii) parámetros diferenciales del sueño. El diseño del conjunto de cebadores para capturar una diana dada (por ejemplo, un miembro diana) puede incluir modificaciones de la secuencia/una longitud más corta a fin de reducir la eficiencia de captura y reducir las profundidades relativas de cobertura de la diana,

(iv) mezcla de diferentes conjuntos de señuelos. Los conjuntos de señuelos que están diseñados para capturar diferentes conjuntos diana pueden mezclarse en diferentes proporciones molares para potenciar/reducir las profundidades relativas de cobertura de la diana,

(v) utilizar diferentes tipos de conjuntos de señuelos oligonucleótidos. En determinadas formas de realización, el conjunto de señuelos puede incluir:

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

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

(c) uno o más señuelos 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

(h) una combinación de cualquiera de los anteriores.

Las diferentes combinaciones de oligonucleótidos pueden mezclarse en diferentes proporciones, por ejemplo una proporción seleccionada de 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:50; 1:100, 1:1000 o similar. En una forma de realización, la proporción de señuelo sintetizado químicamente a señuelo generado en la matriz se selecciona de 1:5, 1:10 o 1:20. Los oligonucleótidos de ADN o ARN pueden ser naturales o no naturales. En determinadas formas de realización, los señuelos incluyen uno o más nucleótidos no naturales para, por ejemplo, incrementar la temperatura de fusión. Entre los oligonucleótidos no naturales ejemplificativos se incluyen nucleótidos de ADN o ARN modificado. Entre los nucleótidos modificados ejemplificativos (por ejemplo, nucleótidos de ARN o ADN modificado) se incluyen, aunque sin limitarse a ellos, un ácido nucleico bloqueado (ANB), en el que la fracción de ribosa de un nucleótido de ANB se modifica con un puente adicional que conecta el oxígeno 2' y el carbono 4'; ácido péptido nucleico (APN), por ejemplo un APN compuesto de unidades repetidas de N-(2-aminoetil)-glicina unidas mediante enlaces peptídicas; un oligonucleótido de ADN o ARN modificado para capturar regiones de bajo contenido en GC; un ácido nucleico bicíclico (ANBi), un oligonucleótido entrecruzado, una 5-metil-desoxicitidina modificada y 2,6-diaminopurina. Se conocen en la técnica otros nucleótidos de ADN y ARN modificados.

En determinadas formas de realización, se obtiene una cobertura sustancialmente uniforme u homogénea de una secuencia diana (por ejemplo, un miembro diana). Por ejemplo, en cada categoría de conjunto de señuelos/diana, puede optimizarse la uniformidad de la cobertura mediante modificación de los parámetros de los señuelos, por ejemplo, mediante uno o más de:

(i) puede utilizarse el incremento/reducción de la representación o solapamiento de los señuelos para potenciar/reducir la cobertura de las dianas (por ejemplo, miembros de diana) que se encuentra infra/sobre-cubiertos respecto a otras dianas en la misma categoría,

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

(iii) puede modificarse una secuencia de señuelo para reducir la estructura secundaria del señuelo y potenciar su eficiencia de selección,

(iv) puede modificarse la longitud de un señuelo para equalizar la cinética de la hibridación de fusión de los diferentes señuelos dentro de la misma categoría. La longitud del señuelo puede modificarse directamente (mediante producción de señuelos de longitudes diferentes) o indirectamente (mediante la producción de señuelos de longitud consistente y sustituyendo los extremos de señuelos por secuencia arbitraria),

(v) la modificación de señuelos de diferente orientación para la misma región diana (es decir, cadena directa y cadena inversa) puede presentar diferentes eficiencias de unión. Puede seleccionarse el conjunto de señuelos con cualquiera de las orientaciones que proporcione una cobertura óptima de cada diana,

- (vi) modificar la cantidad de una entidad de unión, por ejemplo, una etiqueta de captura (por ejemplo, biotina), presente en cada señuelo puede afectar a su eficiencia de unión. El incremento/reducción del nivel de etiqueta de los señuelos con diana en una diana específica puede utilizarse para potenciar/reducir la cobertura relativa de la diana.
- (vii) modificar el tipo de nucleótido utilizado para diferentes señuelos puede alterarse para afectar a la afinidad de unión a la diana y potenciar/reducir la cobertura relativa de la diana, o
- (viii) utilizar señuelos oligonucleótidos modificados, por ejemplo con un apareamiento de bases más estable, puede utilizarse para ecualizar la cinética de la hibridación de fusión entre zonas de contenido de GC bajo o normal respecto a las de contenido de GC elevado.

Por ejemplo, pueden utilizarse diferentes tipos de conjuntos de señuelos oligonucleótidos. En una forma de realización, el valor de la eficiencia de selección se modifica mediante la utilización de diferentes tipos de señuelos oligonucleótidos para comprender regiones diana preseleccionadas. Por ejemplo, puede utilizarse un primer conjunto de señuelos (por ejemplo, un conjunto de señuelos basado en una matriz que comprende 10,000-50,000 señuelos de ARN o ADN) para cubrir una gran zona diana (por ejemplo, una zona diana total de 1-2 MB). Al primer conjunto de señuelos puede añadirse un segundo conjunto de señuelos (por ejemplo, un conjunto de señuelos de ARN o ADN sintéticos que comprende menos de 5,000 señuelos) para cubrir una región diana preseleccionada (por ejemplo, intervalos subgenómicos de interés seleccionados que comprende, por ejemplo, 250 kb o menos de una zona diana) y/o regiones de estructura secundaria superior, por ejemplo de mayor contenido en GC. Los intervalos subgenómicos de interés seleccionados pueden corresponder a uno o más de los genes o productos génicos indicados en la presente memoria, o a un fragmento de los mismos. El segundo conjunto de señuelos puede incluir aproximadamente 1-5,000, 2-5,000, 3-5,000, 10-5,000, 100-5,000, 500-5,000, 1000-5,000, 2,000-5,000 señuelos dependiendo del solapamiento de los señuelos deseado. En otras formas de realización, el segundo conjunto de señuelos puede incluir señuelos oligos seleccionados (por ejemplo, menos de 400, 200, 100, 50, 40, 30, 20, 10, 5, 4, 3, 2 o 1 señuelo) añadidos al primer conjunto de señuelos. El segundo conjunto de señuelos puede mezclarse en cualquier proporción de señuelos oligo individuales. Por ejemplo, el segundo conjunto de señuelos puede incluir señuelos individuales presentes en una proporción equimolar 1:1. Alternativamente, el segundo conjunto de señuelos puede incluir señuelos individuales presentes en una proporción diferente (por ejemplo 1:5, 1:10 o 1:20), por ejemplo para optimizar la captura de determinadas dianas (por ejemplo, determinadas dianas pueden presentar 5-10x más del segundo señuelo que de otras dianas).

En otras formas de realización, la eficiencia de selección se ajusta nivelando la eficiencia de los señuelos individuales en un grupo (por ejemplo, una primera, segunda o tercera pluralidad de señuelos), mediante el ajuste de la abundancia relativa de los señuelos, o la densidad de la entidad de unión (por ejemplo, la densidad de la etiqueta de hapteno o de afinidad) en referencia a la eficiencia diferencial de captura de secuencias observada al utilizar una mezcla equimolar de señuelos, seguido de la introducción de un exceso diferencial de grupo 1 internamente nivelado a la mezcla global de señuelos respecto al grupo 2 internamente nivelado.

En una forma de realización, el método comprende la utilización de una pluralidad de conjuntos de señuelos que incluye un conjunto de señuelos que selecciona un miembro tumoral, por ejemplo una moléculas de ácidos nucleicos que comprende un intervalo subgenómico procedente de una célula tumoral (también denominada en la presente memoria "conjunto de señuelos tumorales"). El miembro tumoral puede ser cualquier secuencia de nucleótidos presente en una célula tumoral, por ejemplo una secuencia de nucleótidos mutada, de tipo salvaje, PGx, de referencia o intrón, tal como se indica en la presente memoria, que se encuentra presente en una célula tumoral o de cáncer. En una forma de realización, el miembro tumoral incluye una alteración (por ejemplo, una o más mutaciones) que aparecen a una frecuencia baja, por ejemplo de aproximadamente 5% o menos de las células de la muestra tumoral alojan la alteración en su genoma. En otras formas de realización, el miembro tumoral incluye una alteración (por ejemplo, una o más mutaciones) que aparecen a una frecuencia de aproximadamente 10% de las células de la muestra tumoral. En otras formas de realización, el miembro tumoral incluye un intervalo subgenómico de un gen o producto génico PGx, una secuencia de intrón, por ejemplo una secuencia de intrón indicada en la presente memoria, una secuencia de referencia que se encuentra presente en una célula tumoral.

En otro aspecto, la invención proporciona un conjunto de señuelos indicado en la presente memoria, combinaciones de conjuntos de señuelos individuales indicados en la presente memoria, por ejemplo combinaciones indicadas en la presente memoria. El conjunto o conjuntos de señuelos pueden ser parte de un kit que puede comprender opcionalmente instrucciones, estándares, amortiguadores o enzimas u otros reactivos.

#### Selección de genes

Los intervalos subgenómicos preseleccionados, intervalos para el análisis, por ejemplo un grupo o conjunto de intervalos subgenómicos para conjuntos o grupos de genes y otras regiones, se indican en la presente memoria.

De esta manera, en formas de realización, un método comprende la selección y/o secuenciación de miembros de biblioteca que incluyen un intervalo subgenómico de por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta o más genes o productos génicos de la muestra de ácidos nucleicos adquirida, en el que los genes o productos génicos se seleccionan de: ABL1, AKT1, AKT2, AKT3, ALK, APC, AR, BRAF, CCND1, CDK4, CDKN2A, CEBPA, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FLT3, HRAS, JAK2, KIT, KRAS, MAP2K1, MAP2K2, MET, MLL, MYC, NF1, NOTCH 1, NPM1, NRAS, NTRK3, PDGFRA, PIK3CA, PIK3CG, PIK3R1, PTCH1, PTCH2, PTEN, RB1, RET, SMO, STK11, SUFU, o TP53, analizando de esta manera la muestra tumoral.

En otras formas de realización, el método comprende la selección y/o secuenciación de miembros de biblioteca que incluyen un intervalo subgenómico de por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta o más genes o productos génicos de la muestra, en la que los genes o productos génicos se seleccionan de: ABL1, AKT1, AKT2, AKT3, ALK, APC, AR, BRAF, CCND1, CDK4, CDKN2A, CEBPA, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FLT3, HRAS, JAK2, KIT, KRAS, MAP2K1, MAP2K2, MET, MLL, MYC, NF1, NOTCH1, NPM1, NRAS, NTRK3, PDGFRA, PIK3CA, PIK3CG, PIK3R1, PTCH1, PTCH2, PTEN, RB1, RET, SMO, STK11, SUFU, o TP53.

En otra forma de realización, se analizan intervalos subgenómicos de uno de los conjuntos o grupos siguientes. Por ejemplo, los intervalos subgenómicos asociados a un gen o producto génico de tumor o cáncer, un gen o producto génico de referencia (por ejemplo, de tipo salvaje), y un gen o producto génico de PGx, pueden proporcionar un grupo o conjunto de intervalos subgenómicos de la muestra tumoral.

En una forma de realización, el método comprende la selección y/o secuenciación de miembros de biblioteca de un conjunto de intervalos subgenómicos de la muestra tumoral, en la que los intervalos subgenómicos se seleccionan de por lo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 o la totalidad de los siguientes:

A) por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta o más intervalos subgenómicos de un gen o producto génico mutado o de tipo salvaje seleccionado de por lo menos cinco o más de: ABL1, AKT1, AKT2, AKT3, ALK, APC, AR, BRAF, CCND1, CDK4, CDKN2A, CEBPA, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FLT3, HRAS, JAK2, KIT, KRAS, MAP2K1, MAP2K2, MET, MLL, MYC, NF1, NOTCH1, NPM1, NRAS, NTRK3, PDGFRA, PIK3CA, PIK3CG, PIK3R1, PTCH1, PTCH2, PTEN, RB1, RET, SMO, STK11, SUFU, o TP53;

B) por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta, treinta y cinco, cuarenta, cuarenta y cinco, cincuenta, cincuenta y cinco, sesenta, sesenta y cinco, setenta, setenta y cinco, ochenta, ochenta y cinco, noventa, noventa y cinco, cien, ciento cinco, ciento diez, ciento quince, ciento veinte o más intervalos subgenómicos de un gen o producto génico mutado o de tipo salvaje seleccionado de por lo menos cinco o más de: ABL2, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BAP1, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRCA1, BRCA2, CBL, CARD11, CBL, CCND2, CCND3, CCNE1, CD79A, CD79B, CDH1, CDH2, CDH20, CDH5, CDK6, CDK8, CDKN2B, CDKN2C, CHEK1, CHEK2, CRKL, CRLF2, DNMT3A, DOT1L, EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, ERBB3, ERBB4, ERG, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FGFR4, FLT1, FLT4, FOXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GUCY1A2, HOXA3, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, JAK1, JAK3, JUN, KDM6A, KDR, LRP1B, LRP6, LTK, MAP2K4, MCL1, MDM2, MDM4, MEN1, MITM, MLH1, MPL, MRE11A, MSH2, MSH6, MTOR, MUTYH, MYCL1, MYCN, NF2, NKX2-1, NTRK1, NTRK2, PAK3, PAX5, PDGFRB, PKHD1, PLCG1, PRKDC, PTPN11, PTPRD, RAF1, RARA, RICTOR, RPTOR, RUNX1, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SOX10, SOX2, SRC, TBX22, TET2, TGFBR2, TMPRSS2, TNFAIP3, TNK, TNKS2, TOP1, TSC1, TSC2, USP9X, VHL, o WT1,

C) por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte o más intervalos subgenómicos de un gen o producto génico según la Tabla 1, 1A, 2, 3 o 4,

D) por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte o más intervalos subgenómicos de un gen o producto génico que está asociado a un tumor o cáncer (por ejemplo es un factor predictivo de respuesta positiva o negativa al tratamiento, es un factor pronóstico positivo o negativo, o permite el diagnóstico diferencial, de un tumor o cáncer, por ejemplo un gen o producto génico seleccionado de uno o más de: ABL1, AKT1, ALK, AR, BRAF, BRCA1, BRCA2, CEBPA, EGFR, ERBB2, FLT3, JAK2, KIT, KRAS, MET, NPM1, PDGFRA, PIK3CA, RARA, AKT2, AKT3, MAP2K4, NOTCH1 y TP53,

E) por lo menos cinco, seis, siete, ocho, nueve, diez o más intervalos subgenómicos, incluyendo un codón mutado o de tipo salvaje seleccionado de uno o más de: codón 315 del gen ABL1, codón 1114, 1338, 1450 o 1556 de APC, codón 600 de BRAF; codón 32, 33, 34, 37, 41 o 45 de CTNNB1; codón 719, 746-750, 768, 790, 858 o 861 de EGFR; codón 835 de FLT3; codón 12, 13, o 61 de HRAS; codón 617 de JAK2; codón 816 de KIT; codón 12, 13, o 61 de KRAS; codón 88, 542, 545, 546, 1047, o 1049 de PIK3CA; codón 130, 173, 233, o 267 de PTEN; codón 918 de RET; codón 175, 245, 248, 273, o 306 de TP53 (por ejemplo, por

lo menos cinco, diez, quince, veinte o más intervalos subgenómicos que incluyen uno o más de los codones mostrados en la Tabla 1).

F) Por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta o más de los intervalos subgenómicos de un gen o producto génico mutado o de tipo salvaje (por ejemplo, un polimorfismo de nucleótido único (SNP)) de un intervalo subgenómico que se encuentra presente en un gen o producto génico asociado a uno o más de metabolismo de fármaco, sensibilidad a fármaco, o toxicidad (también denominados en el mismo, genes "PGx") seleccionados de: ABCB1, BCC2, ABCC4, ABCG2, C1orf144, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DPYD, ERCC2, ESR2, FCGR3A, GSTP1, ITPA, LRP2, MAN1B1, MTHFR, NQO1, NRP2, SLC19A1, SLC22A2, SLCO1B3, SOD2, SULT1A1, TPMT, TYMS, UGT1A1 o UMPS,

G) por lo menos cinco, seis, siete, ocho, nueve, diez, quince, veinte, veinticinco, treinta o más de los intervalos subgenómicos de un gen o producto génico PGx mutado o de tipo salvaje (por ejemplo, un polimorfismo de nucleótido único (SNP)) de un intervalo subgenómico que se encuentra presente en un gen o producto génico asociado a uno o más de: (i) mejor supervivencia de un paciente de cáncer tratado con un fármaco (por ejemplo, mejor supervivencia de un paciente de cáncer de mama tratado con paclitaxel (por ejemplo, un gen ABCB1), (ii) metabolismo del paclitaxel (por ejemplo, genes CYP2C8 en diferentes loci y mutaciones mostradas en la Tabla 2; gen CYP3A4), (iii) toxicidad con un fármaco (por ejemplo, toxicidad de 6-MP tal como se observa con el gen ABCC4 (Tabla 2); toxicidad de 5-FU tal como se observa con el gen DPYD, el gen TYMS o el gen UMPS (Tabla 2); toxicidad de purinas tal como se observa con un gen TMPT (Tabla 2); toxicidad de daunorrubicina tal como se observa con el gen NRP2, el gen C1orf144 o el gen CYP1B1 (Tabla 2), o (iv) un efecto secundario de un fármaco (por ejemplo, los genes ABCG2, TYMS, UGT1A1, ESR1 y ESR2 (Tabla 2)),

H) una alteración de traslocación de por lo menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 110 o más genes o productos génicos según la Tabla 3,

I) una alteración de traslocación de por lo menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 110 o más genes o productos génicos según la Tabla 3 en una muestra de tumor sólido de los tipos de cáncer especificados en la misma,

J) una alteración de traslocación de por lo menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200 o más genes o productos génicos según la Tabla 4,

K) una alteración de traslocación de por lo menos 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200 o más genes o productos génicos según la Tabla 4 en una muestra de tumor de tipo heme de los tipos de cáncer especificados en la misma,

L) por lo menos cinco genes o productos génicos seleccionados de la Tabla 1-4, en la que una variación alélica, por ejemplo en la posición preseleccionada, se asocia a un tipo preseleccionado de tumor, y en el que dicha variación alélica se encuentra presente en menos de 5% de las células en dicho tipo tumoral,

M) por lo menos cinco genes o productos génicos seleccionados de la Tabla 1, 1A-4, que se encuentran incluidos en una región rica en GC, o

N) por lo menos cinco genes o productos génicos indicativos de un factor genético (por ejemplo, un riesgo de línea germinal) de desarrollar cáncer (por ejemplo, el gen o producto génico se selecciona de uno o más de BRCA1, BRCA2, EGFR, HRAS, KIT, MPL, ALK, PTEN, RET, APC, CDKN2A, MLH1, MSH2, MSH6, NF1, NF2, RB1, TP53, VHL o WT1).

En todavía otra forma de realización, el método comprende la selección y/o secuenciación de miembros de biblioteca que incluyen un conjunto de intervalos subgenómicos de la muestra tumoral, en la que los intervalos subgenómicos se seleccionan de una, dos, tres, cuatro, cinco, diez, quince o la totalidad de las alteraciones indicadas en la Tabla 1B.

En una forma de realización, el intervalo subgenómico incluye una alteración clasificada en una o más de las categorías A, B, C, D o E.

En otras formas de realización, el intervalo subgenómico incluye una alteración en KRAS G13D en una muestra tumoral, por ejemplo una muestra de tumor de colon, pulmón o mama.

En otras formas de realización, el intervalo subgenómico incluye una alteración en NRAS Q61K en una muestra tumoral, por ejemplo una muestra de melanoma o de tumor de colon.

En todavía otras formas de realización, el intervalo subgenómico incluye una alteración en BRAF V600E en una muestra tumoral, por ejemplo una muestra de melanoma, tumor de colon o tumor de pulmón.

5 En otras formas de realización, el intervalo subgenómico incluye una alteración en BRAF D594G en una muestra tumoral, por ejemplo una muestra de tumor de pulmón.

En otras formas de realización, el intervalo subgenómico incluye una alteración en PIK3CA H1047R en una muestra tumoral, por ejemplo una muestra de tumor de mama o de colon.

10 En todavía otras formas de realización, el intervalo subgenómico incluye una alteración en EGFR L858R o T790M en una muestra tumoral, por ejemplo una muestra de tumor de pulmón.

En otras formas de realización, el intervalo subgenómico incluye una alteración en ERBB2 en una muestra tumoral, por ejemplo una amplificación de ERBB2 en una muestra de tumor de mama.

15 En otras formas de realización, el intervalo subgenómico incluye una alteración en BRCA1 en una muestra tumoral, por ejemplo una inactivación bialélica de BRCA1 en una muestra de tumor de mama.

20 En otras formas de realización, el intervalo subgenómico incluye una alteración en BRCA2 en una muestra tumoral, por ejemplo una inactivación bialélica de BRCA2 en una muestra de tumor pancreático.

En otras formas de realización, el intervalo subgenómico incluye una alteración en ATM en una muestra tumoral, por ejemplo una inactivación bialélica de ATM en una muestra de tumor de mama.

25 En otras formas de realización, el intervalo subgenómico incluye una alteración en TSC en una muestra tumoral, por ejemplo una inactivación bialélica de TSC en una muestra de tumor de colon.

#### Breve descripción de los dibujos

30 La figura 1 representa la configuración típica de una biblioteca de moldes que conduce a la selección de moldes deseados con el método de captura de híbridos.

La figura 2A representa una estrategia convencional de bloqueo de oligonucleótidos utilizando los oligonucleótidos 102 como bloqueantes para hibridarse con secuencias 102 adaptadoras correspondientes que se encuentran en los moldes 203 bajo condiciones de temperatura que son favorables a la formación de dúplex 102:102. Observar que se capturan múltiples moldes 203 mediante unión de un señuelo oligonucleótido 204 y su interacción con un reactivo de captura sobre un soporte inmovilizado 205.

40 La figura 2 representa la estrategia de bloqueo con oligonucleótidos aumentados en  $T_m$  para el enriquecimiento de dianas de ADN deseadas sin coselección de secuencias de ADN no deseadas del reservorio complejo de moldes de NGS. En lugar de utilizar oligonucleótidos 102 como bloqueantes, la estrategia utiliza oligonucleótidos aumentados en  $T_m$  202 como bloqueantes para hibridarse con las secuencias 102 adaptadoras correspondientes que se encuentran en los moldes 203 bajo condiciones de temperatura que son favorables a la formación de dúplex 102:102. Debido a que los dúplex 202:102 resultan favorecidos respecto a los dúplex 102:102 a temperaturas próximas al valor de  $T_m$  aumentado óptimo, se capturan menos moldes no deseados 203 mediante unión de un señuelo oligonucleótido 204 y su interacción con un reactivo de captura sobre un soporte inmovilizado 205.

50 La figura 3A-3F es una ilustración mediante diagrama de flujo de un método para el análisis multigénico de una muestra tumoral.

La figura 3A representa un diagrama de flujo de la recepción de muestras, control de calidad y aislamiento de ADN.

55 La figura 3B representa un diagrama de flujo del control de calidad del AD y la generación de bibliotecas.

La figura 3C proporciona una ilustración mediante diagrama de flujo de la captura de híbridos y la secuenciación.

60 La figura 3D proporciona una ilustración mediante diagrama de flujo del control de calidad de los datos de secuencias y la llamada de mutaciones.

La figura 3E proporciona una ilustración mediante diagrama de flujo de la generación de informes.

65 La figura 3F proporciona una ilustración mediante diagrama de flujo de datos adicionales de la generación de informes.

La figura 4 representa el impacto de las expectativas previas y la profundidad de lectura sobre la detección de mutaciones.

La figura 5 representa las frecuencias de mutación en más de 100 muestras clínicas de cáncer.

La figura 6 es una representación lineal de un histograma de coberturas. El número de dianas (eje y) se ilustra como función de la cobertura (eje x). La línea nº 1 representa la cobertura utilizando un conjunto de señuelos que incluye señuelos oligonucleótidos de ADN derivados de matriz biotinilados con adición de señuelos oligonucleótidos de ADN sintetizados individualmente biotinilados (denominados en la presente memoria "conjunto de señuelos nº 1"). La línea nº 2 representa la cobertura utilizando un conjunto de señuelos que incluye señuelos oligonucleótidos de ADN derivados de matriz biotinilados únicamente (denominados en la presente memoria "conjunto de señuelos nº 2"). La cobertura media global utilizando el conjunto de señuelos nº 2 era de 924, mientras que la cobertura en zonas de elevado contenido de GC (aproximadamente 68%) utilizando el conjunto de señuelos nº 2 era de 73. En contraste, al utilizar el conjunto de señuelos nº 1, la cobertura global era de aproximadamente 918, pero la cobertura había mejorado hasta 183 en zonas de elevado contenido de GC.

La figura 7 es un histograma de coberturas que compara la uniformidad de la cobertura detectada con un conjunto de señuelos que consiste en señuelos oligonucleótidos de ADN sintetizados individualmente biotinilados (conjunto de señuelos nº 1) y un conjunto de señuelos que incluye señuelos oligonucleótidos de ADN derivados de matriz biotinilados con adición de señuelos oligonucleótidos de ADN sintetizados individualmente biotinilados ("conjunto de señuelos nº 2") en comparación con un conjunto de señuelos que incluye señuelos oligonucleótidos de ADN derivados de matriz biotinilados únicamente ("conjunto de señuelos nº 3"). Los conjuntos de señuelos se muestran como nº 1, 2 y 3 en la figura 7. Se detectaron varios huecos en la cobertura utilizando el conjunto de señuelos nº 3, aunque no se detectaron utilizando los conjuntos de señuelos nº 1-2, tal como se ilustra en la figura 7.

La figura 8 ilustra en forma de diagrama una configuración ejemplificativa de concatámeros no diana de los miembros de biblioteca. Las regiones no diana (por ejemplo, los adaptadores ilustrados como "P5" y "P7") se muestran como hibridantes con sus cadenas no diana complementarias (ilustradas como "rcP5" y "rcP7", respectivamente). Se muestra un señuelo etiquetado con biotina que se hibrida con una región complementaria de la inserción de diana del miembro de biblioteca.

La figura 9 es un gráfico de columnas que ilustra el porcentaje de selección de dianas utilizando oligos de bloqueo estándares y extendidos.

La figura 10 ilustra un histograma de cobertura de exones que muestra los resultados de captura utilizando bloqueantes estándares o extendidos.

La figura 11 ilustra el número de lecturas de secuencias por muestra como función de los diferentes tipos de bloqueante utilizados en la etapa de captura de híbridos de la forma de realización. De izquierda a derecha en la figura, el control no enriquecido (es decir, sin oligonucleótidos bloqueantes); oligos "P7/P5" (SEC ID nº 81 y nº 23, respectivamente); oligos "P7Comp 6xl/P5" (SEC ID nº 82 y nº 23, respectivamente); oligos "P7Comp Med ANB 6xl/P5 Med ANB" (SEC ID nº 84 y nº 85, respectivamente); oligos "P7Comp Alta ANB 6xl/P5 Alta ANB" (SEC ID nº 86 y nº 87, respectivamente); oligos "P7 6xl/P5Comp" (SEC ID nº 88 y nº 89, respectivamente); oligos "P7 Med ANB 6xl/P5Comp Med ANB" (SEC ID nº 90 y nº 91, respectivamente) y oligos "P7 Alta ANB 6xl/P5Comp Alta ANB" (SEC ID nº 92 y nº 93, respectivamente).

La figura 12 ilustra el porcentaje de superficie de diana cubierta 1x (es decir, más de una vez) como función de diferentes tipos de bloqueante utilizados en la etapa de captura de híbridos de la forma de realización. De izquierda a derecha en la figura, oligos "P7/P5" (SEC ID nº 81 y nº 23, respectivamente); oligos "P7Comp 6xl/P5" (SEC ID nº 82 y nº 23, respectivamente); oligos "P7Comp Med ANB 6xl/P5 Med ANB" (SEC ID nº 84 y nº 85, respectivamente); oligos "P7Comp Alta ANB 6xl/P5 Alta ANB" (SEC ID nº 86 y nº 87, respectivamente); oligos "P7 6xl/P5Comp" (SEC ID nº 88 y nº 89, respectivamente); oligos "P7 Med ANB 6xl/P5Comp Med ANB" (SEC ID nº 90 y nº 91, respectivamente) y oligos "P7 Alta ANB 6xl/P5Comp Alta ANB" (SEC ID nº 92 y nº 93, respectivamente); el porcentaje se expresa como función de la superficie de diana cubierta 2x, 10x, 20x y 30x (de izquierda a derecha para cada uno de los tipos ilustrados de emparejamiento de bloqueante sometidos a ensayo).

La figura 13 ilustra el número de lecturas de secuencia que se alinean en la diana como función de los diferentes tipos de bloqueante utilizados en la etapa de captura de híbridos de la forma de realización. De izquierda a derecha en la figura, oligos "P7/P5" (SEC ID nº 81 y nº 23, respectivamente); oligos "P7Comp 6xl/P5" (SEC ID nº 82 y nº 23, respectivamente); oligos "P7Comp Med ANB 6xl/P5 Med ANB" (SEC ID nº 84 y nº 85, respectivamente); oligos "P7Comp Alta ANB 6xl/P5 Alta ANB" (SEC ID nº 86 y nº 87, respectivamente); oligos "P7 6xl/P5Comp" (SEC ID nº 88 y nº 89, respectivamente); oligos "P7 Med ANB 6xl/P5Comp Med ANB" (SEC

ID nº 90 y nº 91, respectivamente) y oligos "P7 Alta ANB 6x1/P5Comp Alta ANB" (SEC ID nº 92 y nº 93, respectivamente).

La figura 14 ilustra el factor de enriquecimiento en secuencias en la diana como función de los diferentes tipos de bloqueante utilizados en la etapa de captura de híbridos de la forma de realización. De izquierda a derecha en la figura, oligos "P7/P5" (SEC ID nº 81 y nº 23, respectivamente); oligos "P7Comp 6x1/P5" (SEC ID nº 82 y nº 23, respectivamente); oligos "P7Comp Med ANB 6x1/P5 Med ANB" (SEC ID nº 84 y nº 85, respectivamente); oligos "P7Comp Alta ANB 6x1/P5 Alta ANB" (SEC ID nº 86 y nº 87, respectivamente); oligos "P7 6x1/P5Comp" (SEC ID nº 88 y nº 89, respectivamente); oligos "P7 Med ANB 6x1/P5Comp Med ANB" (SEC ID nº 90 y nº 91, respectivamente) y oligos "P7 Alta ANB 6x1/P5Comp Alta ANB" (SEC ID nº 92 y nº 93, respectivamente).

La figura 15 ilustra el porcentaje de lecturas de secuencias en la diana como función de los diferentes tipos de bloqueante utilizados en la etapa de captura de híbridos de la forma de realización. De izquierda a derecha en la figura, los oligos "0" no contienen grupos que aumentan  $T_m$  modificados con ANB (SEC ID nº 94 y nº 97, respectivamente); los oligos "8" contienen 8 grupos que aumentan  $T_m$  modificados con ANB en cada oligo bloqueante (SEC ID nº 95 y nº 98, respectivamente) y los oligos "22" oligos contienen 17 o 22 grupos que aumentan  $T_m$  modificados con ANB en cada oligo bloqueante, dependiendo del oligo bloqueante (SEC ID nº 96 y nº 97, respectivamente); los gráficos de columnas individuales para cada tipo de grupo bloqueante representan réplicas independientes de ensayo.

### Descripción detallada de la invención

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

Los términos utilizados en la presente memoria pretenden ser términos "abiertos" (por ejemplo, el término "incluyendo" debe interpretarse como "incluyendo, aunque sin limitación"; el término "presentando" debe interpretarse como "que presenta por lo menos"; el término "incluye" debe interpretarse como "incluye, aunque sin limitación", etc.).

Además, en aquellos casos en los que se utiliza una convención análoga a "por lo menos uno de A, B y C, etc.", en general dicha construcción posee el sentido que entendería el experto ordinario en la materia (por ejemplo, "un sistema que presenta por lo menos A, B y C" debería incluir, aunque sin limitación, sistemas que presentan A solo, B solo, C solo, A y B juntos, A y C juntos, B y C juntos y/o A, B y C juntos). El experto en la materia entenderá además que prácticamente cualquier término y/o expresión disyuntiva que presenta dos o más términos alternativos, en la descripción o en figuras, debe entenderse que contempla las posibilidades de incluir uno de los términos, cualquiera de los términos, o ambos términos. Por ejemplo, la expresión "A o B" se entenderá que incluye las posibilidades "A" o "B", o "A y B".

Todas las expresiones tales como "de", "a", "hasta", "por lo menos", "mayor que", "menor que", y similares, incluyendo el número indicado y se refieren a intervalos que después pueden descomponerse en subintervalos, tal como se ha comentado anteriormente.

Un intervalo incluye cada miembro individual. De esta manera, por ejemplo un grupo que presenta 1 a 3 miembros se refiere a grupos que presentan 1, 2 o 3 miembros. De manera similar, un grupo que presenta 6 miembros se refiere a grupos que presentan 1, 2, 3, 4 o 6 miembros, y de esta manera sucesivamente.

El verbo modal "puede" se refiere a la utilización o selección preferente de una o más opciones o selecciones de entre varias formas de realización o características indicadas contenidas en el mismo. En el caso de que no se den a conocer opciones o selecciones respecto a una forma de realización o característica particular contenida en las mismas, el verbo modal "puede" se refiere a un acto afirmativo respecto a cómo preparar o utilizar un aspecto de una forma de realización o característica indicada contenida en el mismo, o una decisión definitiva de utilizar una habilidad específica respecto a la forma de realización o característica indicada contenida en el mismo. En este último contexto, el verbo modal "puede" presenta el mismo significado y connotación que el verbo auxiliar "poder".

Tal como se utiliza en la presente memoria, los artículos "un" y "una" se refieren a uno o más de uno (por ejemplo, a por lo menos uno) del objeto gramatical del artículo.

"Aproximadamente" se referirá en general a un grado aceptable de error para la cantidad medida dada la naturaleza o precisión de las mediciones. Los grados ejemplificativos de error se encuentran dentro de 20 a 25 por ciento (%), típicamente, dentro de 10%, y más típicamente dentro de 5%, de un valor o intervalo de valores dado.

"Adquirir" o "adquisición" como términos utilizados en la presente memoria se refieren a obtener la posesión de una entidad física o un valor, por ejemplo un valor numérico, mediante la "adquisición directa" o "adquisición indirecta" de la entidad física o valor. "Adquisición directa" se refiere a llevar a cabo el procedimiento (por ejemplo, llevar a cabo un método sintético o analítico) para obtener la entidad física o valor. "Adquisición indirecta" se refiere

a recibir la entidad física o valor de otra parte o fuente (por ejemplo, un tercer laboratorio que ha adquirido directamente la entidad física o valor). Adquirir directamente una entidad física incluye llevar a cabo un procedimiento que incluye un cambio físico en una sustancia física, por ejemplo un material de partida. Entre los cambios ejemplificativos se incluyen producir una entidad física a partir de dos o un material de partida, cortar o fragmentar una sustancia, separar o purificar una sustancia, combinar dos o más entidades separadas en una mezcla, llevar a cabo una reacción química que incluye romper o formar un enlace covalente o no covalente. Adquirir directamente un valor incluye llevar a cabo un procedimiento que incluye un cambio físico en una muestra u otra sustancia, por ejemplo llevar a cabo un procedimiento analítico que incluye un cambio físico en una sustancia, por ejemplo una muestra, analito o reactivo (en ocasiones denominado en la presente memoria "análisis físico"), llevar a cabo un método analítico, por ejemplo un método que incluye uno o más de los siguientes: separar o purificar una sustancia, por ejemplo un analito o un fragmento u otro derivado del mismo, respecto de otra sustancia; combinar un analito o fragmento u otro derivado del mismo, con otra sustancia, por ejemplo un amortiguador, solvente o reactivo, o modificar la estructura de un analito, o un fragmento u otro derivado del mismo, por ejemplo mediante la rotura o formación de un enlace covalente o no covalente, entre un primer y un segundo átomo del analito, o mediante la modificación de la estructura de un reactivo, o un fragmento u otro derivado del mismo, por ejemplo mediante rotura o formación de un enlace covalente o no covalente, entre un primer y un segundo átomo del reactivo.

"Adquirir una secuencia" o "adquirir una lectura", tal como se utiliza el término en la presente memoria, se refiere a conseguir la posesión de una secuencia de nucleótidos o secuencia de aminoácidos, mediante la "adquisición directa" o "adquisición indirecta" de la secuencia o lectura. "Adquisición directa" de una secuencia o lectura se refiere a llevar a cabo un procedimiento (por ejemplo, llevar a cabo un método sintético o analítico) para obtener la secuencia, tal como llevar a cabo un método de secuenciación (por ejemplo, un método de secuenciación de nueva generación (NGS)). "Adquisición indirecta" de una secuencia o lectura se refiere a recibir información o conocimiento, o recibir, la secuencia de otra parte o fuente (por ejemplo, un tercer laboratorio que ha adquirido directamente la secuencia). La secuencia o lectura adquirida no necesita ser una secuencia completa, por ejemplo, la secuenciación de por lo menos un nucleótido, o la obtención de información o conocimiento que identifica una o más de las alteraciones dadas a conocer en la presente memoria como presente en un sujeto constituye adquirir una secuencia.

La adquisición directa de una secuencia o lectura incluye llevar a cabo un procedimiento que incluye un cambio físico en una sustancia física, por ejemplo un material de partida, tal como una muestra de tejido o celular, por ejemplo una biopsia, o una muestra de ácido nucleico aislado (por ejemplo, ADN o ARN). Entre los cambios ejemplificativos se incluyen producir 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 ácidos nucleicos a partir de un tejido); combinar dos o más entidades separadas en una mezcla, llevar a cabo una reacción química que incluye romper o formar un enlace covalente o no covalente. Adquirir directamente un valor incluye llevar a cabo un procedimiento que incluye un cambio físico en una muestra u otra sustancia, tal como se ha indicado anteriormente.

"Adquirir una muestra" tal como se utiliza la expresión en la presente memoria, se refiere a conseguir la posesión de una muestra, por ejemplo una muestra de tejido o una muestra de ácidos nucleicos, mediante la "adquisición directa" o "adquisición indirecta" de la muestra. "Adquirir directamente una muestra" se refiere a llevar a cabo un procedimiento (por ejemplo, llevar a cabo un método físico, tal como una cirugía o extracción) para obtener la muestra. "Adquisición indirecta de una muestra" se refiere a recibir la muestra de otra parte o fuente (por ejemplo, un tercer laboratorio que ha adquirido directamente la secuencia). Adquirir directamente una muestra incluye llevar a cabo un procedimiento que incluye 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 ha sido aislado anteriormente de un paciente. Entre los cambios ejemplificativos se incluyen producir una entidad física a partir de un material de partida, diseccionar o raspar un tejido; separar o purificar una sustancia (por ejemplo, una muestra de tejido o una muestra de ácidos nucleicos); combinar dos o más entidades separadas en una mezcla; llevar a cabo una reacción química que incluye romper o formar un enlace covalente o no covalente. Adquirir directamente una muestra incluye llevar a cabo un procedimiento que incluye un cambio físico en una muestra u otra sustancia, por ejemplo tal como se ha indicado anteriormente.

"Alteración" o "estructura alterada" tal como se utilizan en la presente memoria, de un gen o producto génico (por ejemplo, un gen o producto génico marcador) se refieren 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 de tipo salvaje. La alteración puede ser de cantidad, estructura y/o actividad en un tejido de cáncer o célula de cáncer, en comparación con su cantidad, estructura y/o cantidad en un tejido o célula normal o sano (por ejemplo, de control) y está asociado a un estado de enfermedad, tal como cáncer. Por ejemplo, una alteración que se asocia a cáncer, o predictiva de la capacidad de respuesta a terapéuticas anticáncer, puede presentar una alteración de la secuencia de nucleótidos (por ejemplo, una mutación), secuencia de aminoácidos, traslocación cromosómica, inversión intracromosómica, número de copia, nivel de expresión, nivel de proteínas, actividad de proteína o estado de metilación, en un tejido de cáncer o célula de cáncer, en comparación con un tejido o célula sano normal. Entre las mutaciones ejemplificativas se incluyen, aunque sin

limitarse a ellas, mutaciones puntuales (por ejemplo, silenciosas, contrasentido o sin sentido), deleciones, inserciones, inversiones, mutaciones de enlace, duplicaciones, traslocaciones, reorganizaciones intercromosómicas e intracromosómicas. Las mutaciones pueden estar presentes en la región codificante o no codificante del gen. En determinadas formas de realización, la alteración o alteraciones se detectan como una reorganización, por ejemplo una reorganización genómica que comprende uno o más intrones o fragmentos de los mismos (por ejemplo, una o más reorganizaciones en la UTR 5' y/o 3'). En determinadas formas de realización, las alteraciones están asociadas (o no asociadas) a un fenotipo, por ejemplo un fenotipo canceroso (por ejemplo, uno o más de riesgo de cáncer, progresión de cáncer, tratamiento de cáncer o resistencia a tratamiento del cáncer). En una forma de realización, la alteración está asociada a uno o más de: un factor de riesgo genético de cáncer, un factor predictivo de respuesta positiva al tratamiento, un factor predictivo de respuesta negativa al tratamiento, un factor pronóstico positivo, un factor pronóstico negativo o un factor diagnóstico.

"Señuelo", tal como se utiliza en la presente memoria, es un tipo de reactivo de captura de híbridos. Un señuelo puede ser una molécula de ácidos nucleicos, por ejemplo una molécula de ADN o ARN, que puede hibridarse con (por ejemplo, ser complementaria a), y de esta manera permitir la captura de, un ácido nucleico diana. En una forma de realización, un señuelo 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 formas de realización, un señuelo incluye una entidad de unión, por ejemplo, una etiqueta de afinidad, que permite la captura y separación, por ejemplo, mediante la unión a una entidad de unión, de un híbrido formado por un señuelo y un ácido nucleico hibridado con el señuelo. En una forma de realización, un señuelo resulta adecuado para la hibridación en fase solución.

La expresión "conjunto de señuelos", tal como se utiliza en la presente memoria, se refiere a una pluralidad de moléculas de señuelo.

"Entidad de unión" se refiere a cualquier molécula a la que pueden unirse directa o indirectamente etiquetas moleculares que es capaz de unirse específicamente a un analito. La entidad de unión puede ser una etiqueta de afinidad en cada secuencia de señuelo. En determinadas formas de realización, la entidad de unión permite la separación de los híbridos de señuelo/miembro a partir de la mezcla de hibridación mediante la unión a una pareja, tal como una molécula de avidina, o un anticuerpo que se une al hapteno o a un fragmento de unión a antígeno del mismo. Entre las entidades de unión ejemplificativas se incluyen, aunque sin limitarse a ellas, una molécula de biotina, un hapteno, un anticuerpo, un fragmento de unión a antígeno, un péptido y una proteína.

"Complementario" se refiere a la complementariedad de secuencia entre regiones de dos cadenas de ácidos nucleicos o entre dos regiones de la misma cadena de ácidos nucleicos. Es conocido que un residuo de adenina de una primera región de ácidos nucleicos es capaz de formar enlaces de hidrógeno específicos ("apareamiento de bases") con un residuo de una segunda región de ácidos nucleicos que es antiparalela respecto a la primera región si el residuo es timina o uracilo. De manera similar, es conocido que un residuo de citosina de una primera cadena de ácidos nucleicos es capaz de apareamiento de bases con un residuo de una segunda cadena de ácidos nucleicos que es antiparalela respecto a la primera cadena si el residuo es guanina. Una primera región de un ácido nucleico es complementaria a una segunda región del mismo ácido nucleico o de uno diferente en el caso de que, al disponer las dos regiones en orientaciones antiparalelas, por lo menos un residuo nucleotídico de la primera región es capaz de apareamiento de bases con un residuo de la segunda región. En determinadas formas de realización, la primera región comprende una primera parte y la segunda región comprende una segunda parte, de manera que, al disponer la primera y segunda parte en orientación antiparalela, por lo menos aproximadamente 50%, por lo menos aproximadamente 75%, por lo menos aproximadamente 90% o por lo menos aproximadamente 95% de los residuos nucleotídicos de la primera parte son capaces de apareamiento de bases con residuos nucleotídicos en la segunda parte. En otras formas de realización, todos los residuos nucleotídicos de la primera parte son capaces de apareamiento de bases con residuos nucleotídicos en la segunda parte.

El término "cáncer" o "tumor" se utiliza intercambiamente en la presente memoria. Estos términos se refieren a la presencia de células que poseen características típicas de las células causantes de cáncer, tales como una proliferación incontrolada, inmortalidad, potencial metastásico, rápido crecimiento y tasa de proliferación, y determinadas características morfológicas. Las células de cáncer con frecuencia se encuentran en forma de un tumor, aunque tales células pueden existir individualmente en un animal, o pueden ser una célula de cáncer no tumorigénica, tal como una célula leucémica. Entre estos términos se incluyen un tumor sólido, un tumor de tejido blando o una lesión metastásica. Tal como se utiliza en la presente memoria, el término "cáncer" incluye cánceres premalignos, así como cánceres malignos.

"Probablemente" o "probabilidad incrementada", tal como se utiliza en la presente memoria, se refiere a una probabilidad incrementada de que ocurra un ítem, objeto, cosa o persona. De esta manera, en un ejemplo, un sujeto que es probable que responda al tratamiento presenta una probabilidad incrementada de responder al tratamiento respecto a un sujeto o grupo de sujetos de referencia.

“Improbablemente” se refiere a una probabilidad reducida de que ocurra un suceso, ítem, objeto, cosa o persona con respecto a una referencia. De esta manera, un sujeto que es improbable que responda al tratamiento presenta una probabilidad reducida de responder al tratamiento respecto a un sujeto o grupo de sujetos de referencia.

5 “Miembro de control” se refiere a un miembro que presenta secuencia de una célula no tumoral.

10 “Selector de secuencia de alineación de indel”, tal como se utiliza en la presente memoria, se refiere a un parámetro que permite o dirige la selección de una secuencia con la que debe alinearse una lectura, con el caso de un indel preseleccionado. La utilización de dicha secuencia puede optimizar la secuenciación de un intervalo subgenómico preseleccionado que comprende un indel. El valor de un selector de secuencia de alineación de indel es una función de un indel preseleccionado, por ejemplo un identificador del indel. En una forma de realización, el valor es la identidad del indel.

15 Tal como se utiliza en la presente memoria, el término “biblioteca” se refiere a una colección de miembros. En una forma de realización, la biblioteca incluye una colección de miembros ácidos nucleicos, por ejemplo una colección de genomas completos, fragmentos subgenómicos, ADNc, fragmentos de ADNc, ARN, fragmentos de ARN o una combinación de los mismos. En una forma de realización, una parte o la totalidad de los miembros de biblioteca comprende una secuencia adaptadora no diana. La secuencia adaptadora puede estar localizada en un extremo o en ambos. La secuencia adaptadora puede resultar útil, por ejemplo, para un método de secuenciación (por ejemplo, un método de NGS), para la amplificación, para la transcripción inversa o para la clonación en un vector.

20 La biblioteca puede comprender una colección 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 biblioteca pueden proceder de un único individuo. En formas de realización, una biblioteca 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 pueden agruparse dos o más bibliotecas de diferentes sujetos para formar una biblioteca que presenta miembros de más de un sujeto. En una forma de realización, el sujeto es un ser humano que presenta, o está en riesgo de presentar, un cáncer o tumor.

25 “Captura de biblioteca” se refiere a un subconjunto de una biblioteca, por ejemplo un subconjunto enriquecido en intervalos subgenómicos preseleccionados, por ejemplo un producto capturado mediante hibridación con señuelos preseleccionados.

30 “miembro” o “miembro de biblioteca” u otro término similar, tal como se utiliza en la presente memoria, se refiere a una molécula de ácidos nucleicos, por ejemplo un ADN, ARN o una combinación de ellos, que es el miembro de una biblioteca. Típicamente, el miembro es una molécula de ADN, por ejemplo ADN genómico o ADNc. Un miembro puede ser ADN genómico fragmentado, por ejemplo, cortado o preparado enzimáticamente. Los miembros comprenden secuencia de un sujeto y pueden comprender además secuencia no derivada del sujeto, por ejemplo una secuencia no diana, tal como una secuencia adaptadora, una secuencia de cebador u otras secuencias que permitan la identificación, por ejemplo secuencias “de código de barras” o “índice”.

35 “Secuenciación de nueva generación o NGS, o secuenciación NGS”, tal como se utiliza en la presente memoria, se refiere a cualquier método de secuenciación que determina la secuencia de nucleótidos de moléculas individuales de ácidos nucleicos (por ejemplo, en la secuenciación de moléculas individuales) o sustitutos expandidos clonalmente para moléculas individuales de ácidos en un modo de alto rendimiento (por ejemplo, se secuencian simultáneamente más de  $10^3$ ,  $10^4$ ,  $10^5$  o más moléculas). En una forma de realización, la abundancia relativa de las especies de ácidos nucleicos en la biblioteca puede estimarse mediante el recuento del número relativo de incidencias de sus secuencias afines en los datos generados mediante el experimento de secuenciación. Los métodos de secuenciación de nueva generación son conocidos en la técnica y se describen en, por ejemplo, Metzker, M., Nature Biotechnology Reviews 11:31-46, 2010, incorporada en la presente memoria como referencia. La secuenciación de nueva generación puede detectar una variante presente en menos de 5% de los ácidos nucleicos en una muestra.

40 “Valor nucleótido” tal como se utiliza en la presente memoria, representa la identidad del nucleótido o nucleótidos que ocupan o están asignados a una posición nucleótida preseleccionada. Entre los valores nucleótido típicos se incluyen: faltante (por ejemplo, deleción); 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 nucleótido puede ser una frecuencia de 1 o más, por ejemplo 2, 3 o 4 bases (u otro valor indicado en la presente memoria, por ejemplo faltante o adicional) en una posición nucleótida. Por ejemplo, un valor nucleótido puede comprender una frecuencia de A y una frecuencia de G, en una posición nucleótida.

65 “O” se utiliza en la presente memoria para referirse y se utiliza intercambiamente con el término “y/o”, a menos que el contexto indique claramente lo contrario. La utilización del término “y/o” en algunos sitios en la presente

memoria no se refiere a que los usos del término “o” no son intercambiables con el término “y/o”, a menos que el contexto indique claramente lo contrario.

“Control primario” se refiere a un tejido no tumoral diferente de un tejido NAT (tejido contiguo normal) en una muestra tumoral. La sangre es un control primario típico.

“Selector de secuencia de alineación de reorganización”, tal como se utiliza en la presente memoria, se refiere a un parámetro que permite o dirige la selección de una secuencia con la que debe alinearse una lectura, con el caso de una reorganización preseleccionada. La utilización de dicha secuencia puede optimizar la secuenciación de un intervalo subgenómico preseleccionado que comprende una reorganización. El valor de un selector de secuencia de alineación de reorganización es una función de una reorganización preseleccionada, por ejemplo un identificador de la reorganización. En una forma de realización, el valor es la identidad de la reorganización. Un “selector de secuencia de alineación de indel” (también definido en otro sitio de la presente memoria) es un ejemplo de un selector de secuencia de alineación de reorganización.

“Muestra”, “muestra de tejido”, “muestra de paciente”, “muestra de célula o tejido de paciente” o “especimen” se refieren, cada uno, a una colección de células similares obtenida de un tejido, o células circulantes, de un sujeto o paciente. La fuente de la muestra de tejido puede ser tejido sólido, tal como de un órgano, muestra de tejido, biopsia o aspirado fresco, congelado y/o conservado; sangre o cualesquiera constituyentes sanguíneos; líquidos corporales, tales como líquido cefalorraquídeo, líquido amniótico, líquido peritoneal o líquido intersticial, o células de cualquier momento en la gestión o desarrollo del sujeto. La muestra de tejido puede contener compuestos que no se encuentran naturalmente entremezclados con el tejido naturalmente, tal como conservantes, anticoagulantes, amortiguadores, fijadores, nutrientes, antibióticos o similares. En una forma de realización, la muestra se conserva en forma de una muestra congelada o como preparación de tejido fijada con formaldehído o paraformaldehído incluida en parafina (FFPE). Por ejemplo, la muestra puede incluirse en una matriz, por ejemplo un bloque de FFPE o una muestra congelada.

En una forma de realización, la muestra es una muestra tumoral, por ejemplo, incluye una o más células premalignas o malignas. En determinadas formas de realización, la muestra, por ejemplo la muestra tumoral, se adquiere a partir de un tumor sólido, un tumor de tejidos blandos o una lesión metastásica. En otras formas de realización, la muestra, por ejemplo la muestra tumoral, incluye tejido o células procedentes de un margen quirúrgico. En otra forma de realización, la muestra, por ejemplo la muestra tumoral, incluye una o más células tumorales circulantes (CTC) (por ejemplo, una CTC adquirida de una muestra de sangre).

“Sensibilidad”, tal como se utiliza en la presente memoria, es una medida de la capacidad de un método de detectar una variante de secuencia preseleccionada en una población heterogénea de secuencias. Un método presenta una sensibilidad de S% para las variantes de F% en el caso de que, dada una muestra en la que la variante de secuencia preseleccionada se encuentra presente como por lo menos F% de las secuencias en la muestra, el método pueda detectar la secuencia preseleccionada con una confianza preseleccionada de C%, S% de las veces. A título de ejemplo, un método presenta una sensibilidad de 90% para las variantes de 5% en el caso de que, dada una muestra en la que la variante de secuencia preseleccionada se encuentra presente como por lo menos 5% de las secuencias en la muestra, el método pueda detectar la secuencia preseleccionada con una confianza preseleccionada de 99%, 9 de cada 10 veces (F=5%, C=99%, S=90%). Entre las sensibilidades ejemplificativas se incluyen aquellas de S=90%, 95%, 99% para variantes de secuencia con F=1%, 5%, 10%, 20%, 50% o 100% a niveles de confianza de C=90%, 95%, 99% y 99,9%.

“Especificidad”, tal como se utiliza en la presente memoria, es una medida de la capacidad de un método de distinguir una variante de secuencia preseleccionada verdaderamente existente respecto de artefactos de secuenciación u otras secuencias estrechamente relacionadas. Es la capacidad de evitar la detección de falsos positivos. La detección de falsos positivos puede producirse por errores introducidos en la secuencia de interés durante la preparación de la muestra, error de secuenciación o secuenciación involuntaria de secuencias estrechamente relacionadas tales como pseudogenes o miembros de una familia génica. Un método presenta una especificidad de X% en el caso de que, al aplicarse a un conjunto de muestras de  $N_{\text{Total}}$  secuencias, en el que  $X_{\text{Verdaderas}}$  secuencias son verdaderamente variantes y  $X_{\text{No verdaderas}}$  no son verdaderamente variantes, el método selecciona por lo menos X% de las secuencias no verdaderamente variantes como no variantes. Por ejemplo, un método presenta una especificidad de 90% en el caso de que, al aplicarse a un conjunto de muestras de 1,000 secuencias, en el que 500 secuencias son verdaderamente variantes y 500 no son verdaderamente variantes, el método selecciona por lo menos 90% de las 500 secuencias no verdaderamente variantes como no variantes. Entre las especificidades ejemplificativas se incluyen 90%, 95%, 98% y 99%.

Una “muestra de ácidos nucleicos tumorales” tal como se utiliza en la presente memoria se refiere a moléculas de ácidos nucleicos de una muestra de tumor o cáncer. Típicamente, es ADN, por ejemplo ADN genómico o ADNc derivado de ARN procedente de una muestra de tumor o cáncer. En determinadas formas de realización, la muestra de ácidos nucleicos tumorales se purifica o se aísla (por ejemplo, se extrae de su estado natural).

Una "muestra de ácidos nucleicos" "de control" o "de referencia", tal como se utiliza en la presente memoria se refiere a moléculas de ácidos nucleicos de una muestra de control o 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 formas de realización, la muestra de ácidos nucleicos de referencia o control es una secuencia de tipo salvaje o no mutada. En determinadas formas de realización, la muestra de ácidos nucleicos de referencia se purifica o se aísla (por ejemplo, se extrae de su estado natural). En otras formas de realización, la muestra de ácidos nucleicos de referencia es de una muestra no tumoral, por ejemplo un control de sangre, un tumor contiguo normal (NAT) o cualquier otra muestra no cancerosa procedente del mismo sujeto o de un sujeto diferente.

"Secuenciación" de una molécula de ácidos nucleicos requiere determinar la identidad de por lo menos 1 nucleótido en la molécula. En formas de realización, se determina la identidad de menos de la totalidad de los nucleótidos en una molécula. En otras formas de realización, se determina la identidad de una mayoría o de la totalidad de los nucleótidos en la molécula.

"Intervalo subgenómico" tal como se utiliza en la presente memoria, se refiere a una parte de secuencia genómica. En una forma de realización, un intervalo subgenómico puede ser una única posición nucleótida, por ejemplo una posición nucleótida variante asociada (positiva o negativamente) con un fenotipo tumoral. En una forma de realización, un intervalo subgenómico comprende más de una posición nucleótida. Entre dichas formas de realización se incluyen secuencias de por lo menos 2, 5, 10, 50, 100, 150 o 250 posiciones nucleótidas de longitud. Los intervalos subgenómicos pueden comprender un gen entero, o una parte preseleccionada del mismo, por ejemplo la región codificante (o partes de la misma), un intrón (o parte del mismo) o un exón (o parte 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 formas de realización, un intervalo subgenómico es una secuencia continua de una fuente genómica. En formas de realización, un intervalo subgenómico incluye secuencias que no son contiguas en el genoma, por ejemplo puede incluir uniones formadas que se encuentran en uniones exón-exón en ADNc.

En una forma de realización, un intervalo subgenómico comprende o consiste en: una única posición nucleótida, 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 intensificador, 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 SNP; una mutación somática, una mutación de línea germinal o ambas; una alteración, por ejemplo, una mutación puntual o una mutación única; una mutación por delección (por ejemplo, una delección en el marco de lectura, una delección intragénica o una delección de un gen completo); una mutación por inserción (por ejemplo, una inserción intragénica); una mutación por inversión (por ejemplo, una inversión intracromosómica); una mutación de unión; una mutación por inserción ligada; una mutación por duplicación invertida; una mutación en tándem (por ejemplo, una duplicación en tándem intracromosómica); una traslocación (por ejemplo, una traslocación cromosómica, una traslocación no recíproca); una reorganización (por ejemplo, una reorganización genómica (por ejemplo, una reorganización de uno o más intrones, o un fragmento de la misma; un intrón reorganizado puede incluir una 5'UTR y/o 3'UTR); un cambio del número de copia génica; un cambio en la expresión génica; un cambio en los niveles de ARN o una combinación de los mismos. El "número de copia de un gen" se refiere al número de secuencias de ADN en una célula codificantes de un producto génico particular. Generalmente, para un gen dado, un mamífero presenta dos copias de cada gen. El número de copia puede incrementarse mediante, por ejemplo, amplificación o duplicación génica, o reducirse mediante delección.

"Valor umbral", tal como se utiliza en la presente memoria, es un valor que es una función del número de lecturas que es necesario que haya presentes para asignar un valor de nucleótido a un intervalo subgenómico. Por ejemplo, es una función del número de lecturas que presenta un valor de nucleótido específico, por ejemplo A, en una posición nucleótida, requerido para asignar ese valor de nucleótido a la posición nucleótida en el intervalo subgenómico. El valor umbral puede expresarse, por ejemplo, como (o como función de) un número de lecturas, por ejemplo un número entero, o como una proporción de lecturas que presenta el valor preseleccionado. A título de ejemplo, en el caso de que el valor umbral sea X, y se encuentren presentes X+1 lecturas con el valor de nucleótido "A", el valor de "A" se asigna a la posición preseleccionada en el intervalo subgenómico. El valor umbral también puede expresarse como una función de la predicción de mutación o variante, frecuencia de mutación o de previo Bayesiano. En una forma de realización, una frecuencia de mutación preseleccionada requeriría un número preseleccionado o proporción de lecturas con un valor de nucleótido, por ejemplo A o G, en una posición preseleccionada, para considerar a ese valor el valor de nucleótido. En formas de realización, el valor umbral puede ser una función de la predicción de mutación, por ejemplo la frecuencia de mutación y el tipo tumoral. Por ejemplo, una variante preseleccionada en una posición nucleótida preseleccionada podría presentar un primer valor umbral en el caso de que el paciente presente un primer tipo tumoral y un segundo valor umbral en el caso de que el paciente presente un segundo tipo tumoral.

Tal como se utiliza en la presente memoria, "miembro diana" se refiere a una molécula de ácidos nucleicos que uno desea aislar a partir de la biblioteca de ácidos nucleicos. En una forma de realización, los miembros diana

puede ser un miembro tumoral, un miembro de referencia, un miembro de control o un miembro de PGx tal como se indica en la presente memoria.

“Miembro tumoral” u otra expresión similar (por ejemplo, un “miembro asociado a tumor o cáncer”) tal como se utiliza en la presente memoria se refiere a un miembro que presenta secuencia de una célula tumoral. En una forma de realización, el miembro tumoral incluye un intervalo subgenómico que presenta una secuencia (por ejemplo, una secuencia de nucleótidos) que presenta una alteración (por ejemplo, una mutación) asociada a un fenotipo canceroso. En otras formas de realización, el miembro tumoral incluye un intervalo subgenómico que presenta una secuencia de tipo salvaje (por ejemplo, una secuencia de nucleótidos de tipo salvaje). Por ejemplo, un intervalo subgenómico de un alelo de tipo salvaje heterocigótico u homocigótico presente en una célula de cáncer. Un miembro tumoral puede incluir un miembro de referencia o un miembro de PGx.

“Miembro de referencia” u otra expresión similar (por ejemplo, un “miembro de control”), tal como se utiliza en la presente memoria, se refiere a un miembro que comprende un intervalo subgenómico que presenta una secuencia (por ejemplo, una secuencia de nucleótidos) que no está asociada al fenotipo canceroso. En una forma de realización, el miembro de referencia incluye una secuencia de nucleótidos de tipo salvaje o no mutada de un gen o producto génico que, en caso de estar mutada, está asociada al fenotipo canceroso. El miembro de referencia puede encontrarse presente en una célula de cáncer o no de cáncer.

“Miembro de PGx” u otra expresión similar, tal como se utiliza en la presente memoria, se refiere a un miembro que comprende un intervalo subgenómico que está asociado al perfil farmacogenético o farmacogenómico de un gen. En una forma de realización, el miembro de PGx incluye un SNP (por ejemplo, un SNP tal como se indica en la presente memoria). En otras formas de realización, el miembro de PGx incluye un intervalo subgenómico según la Tabla 1 o la Tabla 2.

Tal como se utiliza en la presente memoria, una “nucleobase universal” se refiere a una nucleobase que muestra la capacidad de sustituir cualquiera de las cuatro nucleobases normales sin desestabilizar significativamente las interacciones de pares de bases vecinas. En el caso de que tales composiciones de nucleobases, incluyendo composiciones de nucleobases universales, se encuentren presentes en bloqueantes, ocupan una pluralidad de posiciones nucleótidas sustancialmente contiguas de longitudes comprendidas preferentemente entre aproximadamente 5 y aproximadamente 12 nucleótidos.

“Variante”, tal como se utiliza en la presente memoria, se refiere a una estructura que puede estar presente en un intervalo subgenómico que puede presentar más de una estructura, por ejemplo un alelo en un locus polimórfico.

Se presentan encabezamientos, por ejemplo (a), (b), (i), etc., meramente para facilitar la lectura de la especificación y las reivindicaciones. La utilización de encabezamientos en la especificación o reivindicaciones no requiere llevar a cabo las etapas o miembros en orden alfabético o numérico o en el orden en que se presentan.

La invención se refiere a nuevos oligonucleótidos aumentados en  $T_m$  como bloqueantes y señuelos para un enriquecimiento mejorado de la diana y una selección reducida de secuencias fuera de diana. Las composiciones de oligonucleótidos encuentran una aplicación robusta en la preparación de moldes de ácidos nucleicos para aplicaciones de secuenciación de nueva generación. Los oligonucleótidos resultan modificados por grupos aumentados en  $T_m$  para incrementar la afinidad de unión de los oligonucleótidos para sus dianas respectivas que permite llevar a cabo reacciones de hibridación/captura a temperaturas más elevadas y bajo condiciones de lavado más astringente que los oligonucleótidos no modificados. Para bloqueantes oligonucleótidos con secuencia idéntica a los adaptadores terminales de los moldes de NGS, la inclusión de oligonucleótidos aumentados en  $T_m$  como bloqueantes en el método de captura de híbridos reduce el nivel de secuencias contaminantes no deseadas que resultan de la formación de híbridos mediados por adaptadores entre los moldes de NGS (el “efecto de cadena conectada en serie”), incrementando de esta manera la eficiencia global del procedimiento de enriquecimiento para los moldes de NGS deseados. Se dan a conocer en mayor detalle posteriormente composiciones de oligonucleótidos aumentados en  $T_m$  como bloqueantes y señuelos, así como su utilización específica para el enriquecimiento mejorado en dianas y para una selección reducida de fueros de diana, incluyendo su utilización en aplicaciones tales como experimentos de secuenciación en paralelo masiva.

Haciendo referencia a la figura 1, el ADN de entrada 100 se fragmenta para proporcionar intervalos de tamaño apropiados. Los intervalos de tamaño preferentes para los fragmentos de ADN resultantes 101 dependerá de la aplicación particular y/o de la plataforma de NGS, aunque típicamente presentan una longitud entre 200 y 500 pb. El método preferente de fragmentar el ADN 100 es mediante rotura mecánica del ADN utilizando procedimientos de sonicación. Los sonicadores disponibles comercialmente y otros instrumentos de sonicación pueden utilizarse para fragmentar el ADN 100 hasta los intervalos de tamaño apropiados. Aunque la fragmentación del ADN 100 mediante rotura mecánica es el medio preferente de fragmentación, pueden utilizarse otros procedimientos de fragmentación, tales como la digestión parcial del ADN 100 utilizando endonucleasas (por ejemplo, ADNasas o endonucleasas de restricción).

Los fragmentos de ADN resultantes 101 se tratan enzimáticamente para preparar extremos romos a los que se ligan adaptadores oligonucleótidos 102 que presentan por lo menos un extremo romo, proporcionando los moldes de NGS 103. Típicamente, el ADN roto mecánicamente puede incluir una diversidad de extremos, tales como extremos romos, extremos 5'-protuberantes y extremos 3'-protuberantes. Pueden hacerse que fragmentos de ADN que incluyen extremos 5'-protuberantes presenten extremos romos mediante rellenado de los extremos 3' cortos utilizando una polimerasa adecuada (por ejemplo, la ADN polimerasa de T4, el fragmento grande (Klenow) de la ADN polimerasa I, la ADN polimerasa Vent o la ADN polimerasa Deep Vent, entre otros). Puede hacerse que aquellos fragmentos de ADN que incluyen un extremo 3'-protuberante presenten extremos romos mediante la utilización de la actividad 3'→5' exonucleasa de una ADN polimerasa, preferentemente en presencia de dNTP (por ejemplo, la ADN polimerasa de T4, el fragmento grande (Klenow) de la ADN polimerasa I o la polimerasa *Pfu*, entre otros). También puede hacerse que fragmentos de ADN que presentan extremos 5'-protuberantes y 3'-protuberantes presenten extremos romos utilizando nucleasas de cadenas sencillas (por ejemplo, la nucleasa (por ejemplo, la nucleasa de judía Mungo, la nucleasa PI o la nucleasa S1, entre otras). La utilización de una ADN polimerasa resulta preferible para la utilización en la preparación de extremos romos para los fragmentos 101.

Opcionalmente, los fragmentos resultantes 101 pueden manipularse enzimáticamente para incluir un extremo protuberante de un solo nucleótido (por ejemplo, un extremo protuberante 3'-dA) que puede facilitar la ligación con adaptadores 102 que presentan por lo menos un extremo con el extremo protuberante de un solo nucleótido complementario (en el ejemplo anterior, un extremo protuberante 3'-dT). Dichos fragmentos 101 se preparan típicamente con extremos romos tal como se ha indicado anteriormente. Y después se tratan seguidamente con un enzima que presenta actividad de 3'-polimerasa ("de formación de extremos") (por ejemplo, actividad de ADN polimerasa *Tth*, ADN polimerasa *Bst*, ADN polimerasa *Taq* o ADN polimerasa Klenow (exo-), entre otras).

Además, el ADN roto mecánicamente puede incluir roturas internas (por ejemplo, muescas) dentro de una de las dos cadenas complementarias que no resultan en la rotura completa de la estructura de ADN de doble cadena. Dichas roturas internas pueden repararse utilizando una ADN polimerasa que presenta actividad de traducción de muescas en presencia de dNTP (por ejemplo, ADN polimerasa de T4 o fragmento grande (Klenow) de la ADN polimerasa I, entre otras) o en presencia de una ligasa adecuada en presencia de ATP (por ejemplo, ADN ligasa de T4). Resulta preferible reparar cualesquiera roturas de cadena sencilla dentro del ADN roto mecánicamente de fragmentos 101, ya que los moldes finales 103 preferentemente incluye dos adaptadores 102 ligados en cada extremo de las dos cadenas continuas.

Los adaptadores 102 preferentemente se diseñan para incluir diferentes tipos de extremos. Este diseño preferente se selecciona para proporcionar una única copia de adaptador de doble cadena 102 para cada extremo de los moldes resultantes 103. Para los fragmentos 101 tratados enzimáticamente para incluir extremos romos, se diseñan adaptadores 102 para incluir un primer extremo con un extremo romo y un segundo extremo con un extremo protuberante. Para dichos adaptadores 102, el segundo extremo se diseña adicionalmente para incluir una o más características que impiden la ligación a otros adaptadores 102 (por ejemplo, no presentan un sustrato competente para ligasa, tal como un grupo 5'-fosfato, un grupo 3'-hidroxilo y/o complementariedad de la secuencia, entre otros). Para los fragmentos 101 tratados enzimáticamente para incluir extremos de nucleótido único, se diseñan adaptadores 102 para incluir un primer extremo con un extremo protuberante de un solo nucleótido complementario y un segundo extremo con un tipo diferente de extremo. Al igual que se ha indicado anteriormente, el segundo extremo de los últimos adaptadores 102 preferentemente están diseñados para incluir una o más características que impiden la ligación a otros adaptadores 102.

La composición de oligonucleótidos de adaptadores 102 preferentemente incluye nucleobases convencionales, en la que los enlaces internucleotídicos son fracciones fosfodiéster convencionales. Los adaptadores 102 preferentemente excluyen grupos químicos que muestran propiedades aumentadas para la  $T_m$ , tal como se explica en mayor detalle posteriormente. Las longitudes preferentes de adaptadores oligonucleótidos 102 están comprendidas entre aproximadamente 15 nucleótidos y aproximadamente 75 nucleótidos.

Para determinadas aplicaciones de NGS, resulta deseable incluir secuencias de "código de barras" para permitir la secuenciación multiplex en experimentos de secuenciación en paralelo masiva. Con este fin, los adaptadores 102 pueden incluir una pluralidad de posiciones nucleótidas que presentan composiciones mixtas de nucleobases (por ejemplo, una mezcla de dos o más nucleobases canónicas en una o más posiciones particulares), incluyendo composiciones de nucleobases "universales" (por ejemplo, inosina, 3'-nitropirrol, 5-nitroindol, entre otros) que representan las etiquetas de secuencia de código de barras. Tal como se utiliza en la presente memoria, una "nucleobase universal" se refiere a una nucleobase que muestra la capacidad de sustituir cualquiera de las cuatro nucleobases normales sin desestabilizar significativamente las interacciones de pares de bases vecinas. En el caso de que tales composiciones mixtas de nucleobases, incluyendo composiciones de nucleobases universales, se encuentren presentes en los adaptadores 102, ocupan una pluralidad de posiciones nucleótidas sustancialmente contiguas de longitudes comprendidas preferentemente entre aproximadamente 5 y aproximadamente 12 nucleótidos. Preferentemente, la pluralidad de posiciones nucleótidas sustancialmente contiguas que incluye dichas nucleobases está localizada dentro del oligonucleótido en una posición central alejada de los extremos.

La composición de la secuencia primaria de los adaptadores 102 puede depender de varias consideraciones. Una consideración es la plataforma de NGS utilizada para los experimentos de secuenciación en paralelo masiva. Por ejemplo, los instrumentos automáticos disponibles comercialmente utilizados para las aplicaciones de NGS presentan diferentes bibliotecas de moldes 103 que contienen diferentes adaptadores 102, de manera que la selección de composiciones de secuencia primaria para cualquier plataforma comercial dada de instrumentación de NGS dependerá de ese criterio. Otra consideración es el diseño composicional de la secuencia primaria del oligonucleótido aumentado en  $T_m$  complementario como el bloqueante. Tal como resultará evidente posteriormente, resultan preferentes determinadas composiciones de secuencia primaria para los bloqueantes, que pueden influir sobre el diseño de decisiones respecto a la composición de la secuencia primaria de los adaptadores complementarios 102.

En referencia a la figura 2A-B, el principio de oligonucleótidos aumentados en  $T_m$  como bloqueantes y señuelos se ilustra para una aplicación de NGS típica. Los moldes de doble cadena 203, bloqueantes oligonucleótidos aumentados en  $T_m$  202, señuelos oligonucleótidos biotinilados 204 y C<sub>o</sub>t-1 DNA® (no mostrado) se mezclan entre sí y se desnaturalizan térmicamente a 95°C en una mezcla de amortiguadores ajustada para incluir una concentración final de 5x de amortiguador de citrato sódico salino (SSC) (o amortiguador de hibridación similar, tal como es bien conocido por el experto en la materia) y se mantienen durante 2 horas a 3 días a una temperatura de hibridación inferior al valor de  $T_m$  medio predicho para los híbridos de señuelo:diana. A medida que la mezcla de hibridación se enfría de 95°C en la etapa de desnaturalización hasta la etapa de hibridación, se forman los híbridos de señuelo:diana. Los híbridos de diana:diana también se formarán, con regiones de unión de secuencias complementarias mediante interacción entre dominios repetidos que pueden ser comunes en diferentes ácidos nucleicos diana o dominios adaptadores, que son los mismos para todos los ácidos nucleicos diana en la biblioteca. Se añaden ácidos nucleicos bloqueantes (ADN C<sub>o</sub>t-1 para unir dominios repetidos y bloqueantes oligonucleótidos para unir dominios adaptadores) para competir con las reacciones de diana:diana no deseadas. Sin embargo, en el caso de que se utilicen oligonucleótidos bloqueantes de ADN no modificados, la  $T_m$  del bloqueante y de los dúplex de adaptador-adaptador serán idénticas y la única manera de evitar la formación de dúplex de diana:diana será mediante acción masiva mediante adición de un gran exceso de oligonucleótido bloqueante no modificado a la reacción de hibridación. La utilización de oligonucleótidos bloqueantes modificados aumentados en  $T_m$  evitará la formación de diana:diana mejor que la utilización de bloqueantes no modificados. Debido a que la  $T_m$  de los bloqueantes modificados es más elevada que la de los adaptadores no modificados, se formarán híbridos de bloqueante:adaptador antes de formarse híbridos de adaptador:adaptador, evitando de esta manera la formación de "cadenas conectadas en serie". Por ejemplo, en el caso de que la  $T_m$  del adaptador no modificado sea 65°C y la  $T_m$  del bloqueante modificado sea 75°C, los dúplex de bloqueante modificado:diana se formarán a 75°C y todos los adaptadores resultarán bloqueados antes de que la mezcla de hibridación se enfríe a 65°C, la temperatura a la que pueden formarse dúplex de diana:diana mediante interacciones de adaptador. A continuación, se añade la mezcla a un medio de soporte sólido 205 que contiene estreptavidina para permitir la captura de los híbridos 203:204. El medio de soporte/mezcla se lava bajo condiciones sucesivamente más astringentes (por ejemplo, 1xSSC seguido de 0,1xSSC) a una temperatura inferior al valor estimado de  $T_m$  de señuelo:diana y, preferentemente, superior al valor de la  $T_m$  de los adaptadores no modificados. Dado que los adaptadores habitualmente son mucho más cortos que los oligómeros de señuelo, la  $T_m$  del señuelo habitualmente es muy superior a la  $T_m$  del adaptador. Debido a que los bloqueantes 202 presentan valores de  $T_m$  aumentados en comparación con adaptadores no modificados observados en los moldes, los moldes 203 preferentemente se hibridarán con los bloqueantes 202 bajo las temperaturas incrementadas de hibridación, minimizado de esta manera la formación por parte de los moldes 203 de agregados encadenados en serie a través de sus secuencias adaptadoras respectivas. Tras los lavados astringentes a la temperatura de hibridación elevada, se llevó a cabo un lavado astringente final a temperatura ambiente y los moldes deseados se recuperan a partir del soporte inmovilizado 205.

Los bloqueantes oligonucleótidos típicos correspondientes a las secuencias adaptadoras pueden proporcionar un enriquecimiento de aproximadamente 60% de las secuencias diana deseadas obtenidas de la captura de híbridos. En contraste, los oligonucleótidos aumentados en  $T_m$  como bloqueantes pueden proporcionar un enriquecimiento superior a aproximadamente 80% en las secuencias diana deseadas obtenido de la captura de híbridos. La mejora resultante de enriquecimiento en la diana de los experimentos de captura de híbridos con los oligonucleótidos aumentados en  $T_m$  como intervalos de bloqueantes proporciona un incremento superior a aproximadamente 30% del rendimiento de los moldes diana deseados respecto al rendimiento obtenido con oligonucleótidos no modificados como bloqueantes.

A continuación, se indican diversas formas de realización del diseño de oligonucleótidos aumentados en  $T_m$  como bloqueantes y señuelos. Tal como se utiliza en la presente memoria, un "oligonucleótido aumentado en  $T_m$ " es un oligonucleótido que incluye por lo menos un grupo modificado ("grupo de aumento de  $T_m$ ") que proporciona un valor incrementado de la temperatura de fusión térmica ("valor de  $T_m$  aumentado") para un ácido nucleico dúplex que incluye como pareja de hibridación el oligonucleótido, respecto a un ácido nucleico dúplex que incluye como pareja de hibridación un oligonucleótido que presenta una composición de nucleobases idéntica y grupos no modificados. Típicamente, la utilización de modificaciones que incrementan la  $T_m$  también incrementa la afinidad de unión del bloqueante para la diana, incrementando la  $K_a$  o constante de asociación, y también puede reducir la  $K_d$ , o constante de disociación, de la reacción inversa.

Pueden utilizarse numerosos grupos que aumentan la  $T_m$  en el diseño de oligonucleótidos aumentados en  $T_m$ . Entre los ejemplos de grupos que aumentan la  $T_m$  adecuados para dicho fin se incluyen modificaciones de las nucleobases o fracciones de ribosa, incluyendo, por ejemplo, ácidos nucleicos bloqueados (ANB), ácidos nucleicos bicíclicos (ANBi, tales como ácidos etilnucleicos restringidos, de Isis Pharmaceuticals), bases pirimidina modificadas en C5 (por ejemplo, 5-metil-dC o propinil-pirimidinas, entre otros). También pueden utilizarse reacciones químicas del esqueleto alternativas, tales como ácidos péptido nucleicos (APN) o morfolinos, entre otros. También pueden utilizarse modificaciones no bases para incrementar la  $T_m$  (o afinidad de unión), tales como ligante de surco menor (MGB), espermina, pinza G o una caperuza antraquinona Uaq. El experto en la materia conoce muchas estrategias para incrementar la afinidad de unión y la utilización de la totalidad de dichas modificaciones se considera comprendida dentro del alcance de la invención.

Preferentemente, entre los oligonucleótidos aumentados en  $T_m$  se incluye una pluralidad de grupos que aumentan la  $T_m$ . El número preferente de grupos que aumentan la  $T_m$  es aquel número que proporciona un incremento del valor de la  $T_m$  óptima bajo condiciones astringentes (0,1xSSC) ("valor de  $T_m$  aumentado óptimo") de por lo menos aproximadamente 1,4°C para un ADN dúplex que contiene el oligonucleótido aumentado en  $T_m$  como una cadena complementaria. El número preferente de grupos que aumentan la  $T_m$  en un oligonucleótido aumentado en  $T_m$  proporciona un valor de la  $T_m$  aumentado óptimo comprendido entre aproximadamente 2°C y aproximadamente 25°C. Puede utilizarse más de un tipo de modificación aumentada en  $T_m$  en un único bloqueante modificado, tal como la combinación de nucleótidos de ANBI con un grupo MGB terminal.

Un enfoque preferente al diseño de un oligonucleótido aumentado en  $T_m$  para un enriquecimiento mejorado en molde en los métodos de captura de híbridos depende de los grupos que aumentan la  $T_m$  utilizados en el oligonucleótido. El valor de la  $T_m$  de un oligonucleótido aumentado en  $T_m$  que contiene cualquiera de los grupos que aumentan la  $T_m$  anteriormente mencionados puede determinarse utilizando métodos empíricos rutinarios. La utilización de grupos que aumentan la  $T_m$  de ANB o ANBi resulta preferente, ya que se encuentran disponibles métodos fiables para predecir con exactitud el valor de la  $T_m$  para los oligonucleótidos aumentados en  $T_m$  que contienen dichos últimos grupos que aumentan la  $T_m$ , que requieren una evaluación empírica mínima o reducida. Un ejemplo de uno de tales métodos para dicho propósito se proporciona en la publicación de patente US nº 2012/0029891 A1, publicada el 2 de feb., 2012, titulada "METHODS FOR PREDICTING STABILITY AND MELTING TEMPERATURES OF NUCLEIC ACID DUPLEXES", de Behlke.

Para determinadas formas de realización preferidas, entre los oligonucleótidos aumentados en  $T_m$  se incluye una etiqueta de secuencia de código de barras. Se incluyen con frecuencia miembros de código de barras en uno de los dos oligonucleótidos adaptadores unidos al ácido nucleico diana durante la construcción de bibliotecas. Un miembro de código de barras típicamente presenta 6 bases de longitud; también se utilizan miembros más largos, tales como de 8 bases o más largos. Típicamente, el adaptador de código de barras comprende únicamente uno de los dos adaptadores utilizados en la preparación de bibliotecas de NGS, siendo un adaptador "único y codificado" y siendo un adaptador "universal". También resulta posible situar los códigos de barras en ambos adaptadores. La utilización de adaptadores de código de barras permite mezclar múltiples muestras y procesarlas juntas en una única tanda de secuenciación multiplex, ofreciendo ahorros significativos de costes y un rendimiento incrementado. Las secuencias se desconvolucionan mediante análisis después de la secuenciación. Los experimentos multiplex pueden implicar la utilización de 2, 3, 4 o hasta cien o más secuencias adaptadoras modificadas de código de barras. Debido a que cada adaptador de código de barras diferente presenta una secuencia única, el oligonucleótido u oligonucleótidos bloqueantes más eficaces serían las secuencias que presentan una correspondencia complementaria perfecta con cada adaptador de código de barras único presente en el conjunto. Este enfoque garantizar la  $T_m$  más alta posible para los bloqueantes, ya que los emparejamientos incorrectos dentro del dominio de código de barras entre adaptador y bloqueante reducirán la  $T_m$ . Por lo tanto, por ejemplo utilizar 4 adaptadores de código de barras en una reacción 4-plex requeriría la utilización de 5 oligonucleótidos bloqueantes diferentes que comprenden 4 secuencias únicas para los 4 adaptadores de código de barras y 1 secuencia única para el adaptador universal común. Sin embargo, en el caso de que se utilicen muchos adaptadores de código de barras diferentes, este enfoque puede requerir la utilización de hasta cien o más oligonucleótidos bloqueantes únicos para experimentos multiplex de nivel elevado, lo que no resulta efectivo según los costes. Además, es probable que se produzca la hibridación incorrecta del bloqueante "A" con el adaptador "B", reduciendo la afinidad de unión de los oligonucleótidos bloqueantes y reduciendo la efectividad de la etapa de bloqueo. Una solución es incorporar un dominio "universal" en el oligonucleótido bloqueante que comprende un dominio N-mero aleatorio (por ejemplo, una secuencia de hexámero de bases mixtas NNNNNN) en una localización aproximada dentro del oligonucleótido adaptador para abarcar el dominio de código de barras en el adaptador. Con este enfoque puede utilizarse un único oligonucleótido bloqueante con un gran número de adaptadores con código de barras. La utilización de un dominio N-mero de 6 bases, se encuentran presentes 4096 secuencias diferentes en el reservorio de oligonucleótidos bloqueantes. Con la presencia de este gran número de códigos de barra resulta que la mayoría de parejas de bloqueante:adaptador incluirá emparejamientos incorrectos en el dominio de código de barras. Alternativamente, puede utilizarse una "base universal" en lugar de N-bases. Las bases universales son nucleobases modificadas que se hibridan con algunas o todas las bases naturales con menor coste termodinámico que los emparejamientos incorrectos de bases verdaderas, tales como las parejas G:A o T:T. Existen muchas bases universales, tales como inosina (I), 5-nitroindol ("5-NI"), etc., los cuales son bien

conocidos por el experto en la materia. El emparejamiento de un dominio de inosinas (IIIII) con un código de barras presentará de media una  $T_m$  más elevada que un dominio N-mero totalmente emparejado incorrectamente (NNNNNN). Por lo tanto, pueden utilizarse tres enfoques para preparar oligonucleótidos bloqueantes para adaptadores de código de barras: 1) sintetizar una serie de bloqueantes que son una correspondencia perfecta con cada adaptador, 2) sintetizar un único bloqueante con un dominio N-mero para emparejarse con el dominio de código de barras del adaptador, o 3) sintetizar un único bloqueante con un dominio de bases universales para emparejarse con el dominio de código de barras del adaptador. Puede calcularse una estimación suficientemente exacta del valor de la  $T_m$  para un oligonucleótido bloqueante aumentado en  $T_m$  particular que contiene grupos de ANB o ANBi con el adaptador de código de barras mediante la omisión de la contribución de secuencia atribuida a las secuencias de nucleobases mixtas o universales con el método anteriormente mencionado. El valor de  $T_m$  preciso para dichos oligonucleótidos puede determinarse a continuación con mayor precisión utilizando métodos empíricos rutinarios.

Tal como se ha mencionado anteriormente, los adaptadores 102 se encuentran presentes en forma de dos cadenas complementarias en los moldes 103. Tras la desnaturalización de la población de moldes de doble cadena 103 para la captura de híbridos, cada molde de cadena sencilla 103 incluirá una copia de cadena sencilla correspondiente de adaptador 102. Para evitar las interacciones entre diferentes moldes de cadena sencilla 103 que resultan en el agregado de cadenas conectadas en serie de muchos moldes no relacionados 103, sólo necesita bloquearse una de las dos cadenas adaptadoras 102 para la hibridación con otra cadena complementaria adaptadora 102. Por este motivo, y en formas de realización preferidas, sólo necesita incluirse una cadena oligonucleótida aumentada en  $T_m$  como bloqueante para conseguir un enriquecimiento mejorado en moldes en los métodos de captura de híbridos con los moldes de NGS 103.

El diseñado de la secuencia primaria del oligonucleótido aumentado en  $T_m$  como bloqueante se basa en las secuencias primarias de una de las dos cadenas complementarias de los adaptadores oligonucleótidos 102. Aunque una puede incluir como grupos que aumentan la  $T_m$  cualquiera o la totalidad de las nucleobases disponibles en un oligonucleótido aumentado en  $T_m$ , resulta preferible incluir únicamente un único tipo de nucleobase modificada o dos tipos diferentes de nucleobase modificada.

Los oligonucleótidos modificados con nucleobases que aumentan  $T_m$  presentan un riesgo incrementado de formación de horquillas o autodímeros. Pueden utilizarse algoritmos o calculadores de diseño de oligonucleótidos para modelar el potencial de horquilla y dímero de una secuencia, y deberían utilizarse para ayudar a cribar los patrones de modificación. Ver, por ejemplo, OligoAnalyzer, que se encuentra disponible públicamente en el sitio web de IDT: <http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>. Esta cuestión resulta de particular importancia si se utilizan modificaciones de ANB o ANBi. Las parejas de bases de ANB:ADN muestran un incremento de la  $T_m$  respecto a las parejas de ADN:ADN. Las parejas de bases de ANB:ANB muestran un incremento de la  $T_m$  respecto a las parejas de ANB:ADN. Cualesquiera parejas de ANB:ANB en sucesos de horquilla o autodímero resultan particularmente favorables (observar que sólo pueden formarse parejas ANB:ADN entre bloqueantes y dianas). Por lo tanto, los oligonucleótidos aumentados en  $T_m$  deben diseñarse cuidadosamente para evitar patrones que fomenten la formación de autodímeros u horquillas mediante sucesos de emparejamiento de ANB:ANB. Lo anterior se aplica igualmente a la modificación ANBi.

Un enfoque preferente para evitar dicho problema es utilizar únicamente un solo tipo de nucleobase modificada. Por ejemplo, puede prepararse un oligonucleótido bloqueante aumentado en  $T_m$  utilizando únicamente ANB-C o ANBi-C. Dependiendo de la composición de bases, la sustitución completa de un solo tipo de base podría no conseguir un incremento de la  $T_m$  suficientemente elevado para proporcionar un rendimiento óptimo. En este caso, pueden utilizarse dos nucleobases modificadas diferentes, tales como ANB-C y ANB-A o ANBi-C con ANBi-A. En general, puede utilizarse C modificada con A modificada o T modificada, pero no con G modificada. De manera similar, puede utilizarse A modificada con C modificada o G modificada, pero no con T modificada. La utilización de C modificada con G modificada o la utilización de A modificada con T modificada debería evitarse. Esta estrategia limita el riesgo de incrementar la formación de horquillas/dímeros mediante la limitación de la potencial interacción entre las bases modificadas. La modificación propinilo de pirimidinas sólo se encuentra disponible como bases pdU y pdC. En este caso, un oligonucleótido bloqueante modificado puede incluir una o muchas bases pdC. Alternativamente, el oligonucleótido bloqueante modificado puede incluir una mezcla de bases pdC y pdU y cumplir los criterios de diseño anteriormente establecidos.

Para los oligonucleótidos aumentados en  $T_m$ , el número preferente de grupos que aumentan  $T_m$  puede variar entre aproximadamente 2% y aproximadamente 50% de la composición del oligonucleótido. Generalmente, los oligonucleótidos que sirven de bloqueantes presentarán la misma longitud de una de las dos cadenas complementarias de los oligonucleótidos utilizados en los adaptadores 102 (por ejemplo, de entre -15 y aproximadamente ~75 nucleótidos de longitud). Por ejemplo, el número preferente de grupos que aumentan  $T_m$  puede encontrarse comprendido entre 1 y aproximadamente 25 para un oligonucleótido aumentado en  $T_m$  como bloqueante que presenta 50 nucleótidos. La utilización de una fracción más elevada de residuos modificados incrementará progresivamente la  $T_m$  y añadirá una mejora progresiva al "poder de bloqueo" de dicho reactivo. Sin embargo, la adición de residuos modificados incrementa el coste del oligonucleótido sintético e incrementa el riesgo de formación de autodímeros y horquillas, de manera que se recomienda la utilización crítica de dichos grupos. En

la mayoría de aplicaciones de NGS, sólo se utilizan oligonucleótidos aumentados en  $T_m$  como bloqueantes para conseguir las mejoras deseadas de enriquecimiento en la diana en experimentos de secuenciación en paralelo masiva.

Para los oligonucleótidos aumentados en  $T_m$  que actúan de señuelos, el número preferente de grupos que aumentan la  $T_m$  se encuentra comprendido en el mismo intervalo de porcentajes que el indicado para los oligonucleótidos que actúan de bloqueantes. Los oligonucleótidos que actúan de señuelos no modificados presentan un tamaño comprendido entre aproximadamente 60 y aproximadamente 200 nucleótidos de longitud, en donde la longitud de señuelo más comúnmente utilizada es de aproximadamente 120 nucleótidos. Mediante la inclusión de grupos que aumentan la  $T_m$  en oligonucleótidos utilizados como señuelos, sin embargo, pueden utilizarse señuelos más cortos con una longitud comprendida entre aproximadamente 20 y aproximadamente 100 nucleótidos. Para determinados experimentos de secuenciación en paralelo masiva en aplicaciones de NGS, se utiliza una población de cientos de oligonucleótidos como señuelos. Por lo tanto, dependiendo del número de señuelos requerido en determinadas aplicaciones, la utilización de oligonucleótidos aumentados en  $T_m$  más cortos para cada señuelo candidato en la población puede proporcionar ventajas económicas respecto a la utilización de oligonucleótidos no modificados utilizados como señuelos.

La utilización de señuelos cuya estructura y actividad ha sido verificada según una especificación de producto estandarizada con un procedimiento de control de la calidad proporciona ventajas adicionales. Aunque se encuentran disponibles otros procedimientos para preparar señuelos, resulta preferible preparar como reactivo de captura una composición que incluya una pluralidad de señuelos (es decir, un conjunto de oligonucleótidos de señuelo discretos), en la que cada miembro de la pluralidad de señuelos se prepara individualmente.

Tal como se utiliza en el contexto de la presente memoria, el número de miembros de la pluralidad de oligonucleótidos de señuelo incluye un intervalo entre aproximadamente 10 y 1.000.000. Este intervalo resulta preferente, dependiendo de la aplicación. Por ejemplo, las necesidades de "adición puntual" para reequilibrar los conjuntos sintetizados en matriz existentes pueden ser de tan sólo aproximadamente 10 a 100 oligonucleótidos de señuelo; los conjuntos completos de oligos dirigidos pueden comprender entre 500 y 25.000 oligonucleótidos de señuelo; un conjunto de exoma entero puede incluir aproximadamente 600.000 oligonucleótidos de señuelo.

Más preferentemente, cada miembro de la pluralidad de señuelos se sintetiza individualmente mediante un procedimiento químico en el que la calidad del producto puede monitorizarse durante la síntesis y después de la purificación. Todavía más preferentemente, cada miembro de la pluralidad se prepara mediante un procedimiento químico sintético y se purifica, en el que tanto la calidad de la síntesis como la purificación pueden evaluarse independientemente. Lo más preferentemente, cada miembro de la pluralidad de señuelos presenta una especificación de producto independiente de otros miembros de la pluralidad de manera que puede obtenerse la pluralidad de señuelos, en la que la estructura y actividad de cada miembro se normaliza respecto a otros miembros dentro de la pluralidad de señuelos. La utilización de una pluralidad de señuelos que presenta una actividad normalizada permite una cobertura más completa y uniforme de una diana de interés dada, particularmente para dianas que presentan regiones de elevado contenido de GC. Estas ventajas pueden hacerse realidad para señuelos oligonucleótidos de todos los tipos, es decir, señuelos oligonucleótidos no modificados, así como señuelos oligonucleótidos aumentados en  $T_m$ .

Los oligonucleótidos aumentados en  $T_m$  pueden incluir características adicionales, tales como modificaciones internas o terminales. Para los oligonucleótidos aumentados en  $T_m$  que sirven de bloqueantes, la recuperación de los moldes de NGS deseados tras la captura de híbridos puede resultar típicamente en la copurificación de los bloqueantes. Los bloqueantes se encontrarán sustancialmente diluidos en la población de moldes, a medida que transcurran las sucesivas etapas de amplificación y secuenciación por PCR. Sin embargo, resulta deseable limitar la participación de los bloqueantes como cebadores durante estas etapas sucesivas. Por ello, los oligonucleótidos aumentados en  $T_m$  pueden incluir grupos 3'-terminales (por ejemplo, 3'-dC, 2',3'-ddC, dT invertida, 3'-espaciador C3, entre otros) que impidan la disponibilidad de los bloqueantes como cebadores para la síntesis de ADN.

Los oligonucleótidos que sirven de señuelos incluyen por lo menos una modificación que permite la selección de híbridos de molde:señuelo deseados a partir de la población de moldes 103 durante la captura de híbridos. Un ejemplo de una modificación preferente incluye biotina que puede incorporarse en el señuelo oligonucleótido durante la síntesis química y utilizarse con medio de soporte sólido que contiene avidina o estreptavidina para la selección de híbridos. Pueden utilizarse otros ligandos de captura, tales como digoxigenina u otros grupos, tales como son bien conocidos por el experto en la materia.

Entre los ejemplos preferentes de oligonucleótidos aumentados en  $T_m$  como bloqueantes se incluyen SEC ID nº 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, 32, 34 y 36. Estas secuencias particulares, sus composiciones y los métodos de utilización de aplicaciones de secuenciación en paralelo masiva se describen en mayor detalle en los Ejemplos.

Selección de genes o productos génicos

Los genes o productos génicos seleccionados (también denominados en la presente memoria “genes o productos génicos diana”) pueden incluir intervalos subgenómicos que comprenden regiones intragénicas o regiones intergénicas. Por ejemplo, el intervalo subgenómico puede incluir un exón o un intrón, o un fragmento de los mismos, típicamente una secuencia exónica o un fragmento de la misma. El intervalo subgenómico puede incluir una región codificante o una región no codificante, por ejemplo un promotor, un intensificador, una región 5' no traducida (5-UTR) o una región 3' no traducida (3'-UTR) o un fragmento de las mismas. En otras formas de realización, el intervalo subgenómico incluye un ADNc o un fragmento del mismo. En otras formas de realización, el intervalo subgenómico incluye un SNP, por ejemplo tal como se indica en la presente memoria.

En otras formas de realización, entre los intervalos subgenómicos se incluyen sustancialmente todos los exones en un genoma, por ejemplo uno o más de los intervalos subgenómicos indicados en la presente memoria (por ejemplo, exones de genes o productos génicos de interés seleccionados (por ejemplo, genes o productos génicos asociados a un fenotipo canceroso tal como se indica en la presente memoria). En una forma de realización, el intervalo subgenómico incluye una mutación somática, una mutación de línea germinal o ambas. En una forma de realización, el intervalo subgenómico incluye una alteración, por ejemplo una mutación puntual o individual, una mutación por delección (por ejemplo, una delección dentro del marco, una delección intragénica o una mutación génica completa), una mutación por inserción (por ejemplo, una inserción intragénica), una mutación por inversión (por ejemplo, una inversión intracromosómica), una mutación de unión, una mutación por inserción ligada, una mutación por duplicación invertida, una duplicación en tándem (por ejemplo, una duplicación en tándem intracromosómica), una traslocación (por ejemplo, una traslocación cromosómica o una traslocación no recíproca), una reorganización, un cambio en el número de copia génica o una combinación de los mismos. En determinadas formas de realización, el intervalo subgenómico constituye menos de 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 formas de realización, los intervalos subgenómicos no participan en una enfermedad, por ejemplo, no están asociados a un fenotipo canceroso, tal como se indica en la presente memoria.

En una forma de realización, el gen o producto génico diana es un biomarcador. Tal como se utiliza en la presente memoria, un “biomarcador” o “marcador” es un gen, ARNm o proteína que puede encontrarse alterado, en donde dicha alteración está asociada a cáncer. La alteración puede ser de cantidad, estructura y/o actividad en un tejido de cáncer o célula de cáncer, en comparación con su cantidad, estructura y/o cantidad en un tejido o célula normal o sano (por ejemplo, de control) y está asociado a un estado de enfermedad, tal como cáncer. Por ejemplo, un marcador asociado a cáncer, o predictivo de la capacidad de respuesta a terapéuticas anticáncer, puede presentar una alteración de la secuencia de nucleótidos, secuencia de aminoácidos, traslocación cromosómica, inversión intracromosómica, número de copia, nivel de expresión, nivel de proteínas o estado de metilación, en un tejido de cáncer o célula de cáncer, en comparación con un tejido o célula sano normal. Además, un “marcador” incluye una molécula cuya estructura se encuentra alterada, por ejemplo mutada (contiene una mutación), por ejemplo difiere de la secuencia de tipo salvaje al nivel de nucleótidos o de aminoácidos, por ejemplo por sustitución, delección o inserción, en caso de encontrarse presente en un tejido o célula asociado a un estado de enfermedad, tal como el cáncer.

En una forma de realización, el gen o producto génico diana incluye un polimorfismo de nucleótido único (SNP). En otra forma de realización, el gen o producto génico presenta una delección pequeña, por ejemplo, una pequeña delección intragénica (por ejemplo, una delección dentro del marco o de desplazamiento del marco). En todavía otra forma de realización, la secuencia diana resulta de la delección de un gen entero. En todavía otra forma de realización, la secuencia diana presenta una pequeña inserción, por ejemplo una pequeña inserción intragénica. En una forma de realización, la secuencia diana resulta de una inversión, por ejemplo una inversión intracromosómica. En otra forma de realización, la secuencia diana resulta de una traslocación intercromosómica. En todavía otra forma de realización, la secuencia diana presenta una duplicación en tándem. En una forma de realización, la secuencia diana presenta una característica no deseable (por ejemplo, un elevado contenido de GC o un miembro repetido). En otra forma de realización, la secuencia diana presenta una parte de secuencia de nucleótidos que no puede ser ella misma la diana, por ejemplo debido a su naturaleza repetitiva. En una forma de realización, la secuencia diana resulta de corte y empalme alternativos. En otra forma de realización, la secuencia diana se selecciona de un gen o producto génico, o un fragmento del mismo según la Tabla 1, 1A, 2, 3 o 4.

Entre los cánceres se incluyen, aunque sin limitarse a ellos, cáncer de células B, por ejemplo mieloma múltiple, melanomas, cáncer de mama, cáncer de pulmón (tal como carcinoma pulmonar de células no pequeñas, o CPCNP), cáncer bronquial, cáncer colorrectal, cáncer de próstata, cáncer pancreático, cáncer de estómago, cáncer ovárico, cáncer de vejiga urinaria, cáncer cerebral o del sistema nervioso central, cáncer del sistema nervioso periférico, cáncer esofágico, cáncer cervical, uterino o endometrial, cáncer de la cavidad oral o faringe, cáncer hepático, cáncer renal, cáncer testicular, cáncer del tracto biliar, cáncer del intestino delgado o el apéndice, cáncer de las glándulas salivales, cáncer de la glándula tiroidea, cáncer de la glándula adrenal, osteosarcoma, condrosarcoma, cáncer de tejidos hematológicos, adenocarcinoma, tumores miofibroblásticos inflamatorios, tumor estromal gastrointestinal (TEG), cáncer de colon, mieloma múltiple (MM), síndrome mielodisplásico (SMD), trastorno mieloproliferativo (TMP), leucemia linfocítica aguda (LLA), leucemia mielocítica aguda (LMA), leucemia mielocítica crónica (LMC), policitemia vera, linfoma de Hodgkin, linfoma no de Hodgkin (LNH), sarcoma de los 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, carcinoma piliar, adenocarcinomas papilares, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma del conducto biliar, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, carcinoma vesicular, carcinoma epitelial, glioma, astrocitoma, meduloblastoma, craneofaringioma, ependimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, linfoma folicular, linfoma de células B grandes difusas, linfoma de células del manto, carcinoma hepatocelular, cáncer de tiroides, cáncer gástrico, cáncer de cabeza y cuello, cánceres de células pequeñas, trombocitemia esencial, metaplasia mieloide agnogénica, síndrome hipereosinofílico, mastocitosis sistémica, hipereosinofilia hereditaria, leucemia eosinofílica crónica, cánceres neuroendocrinos, tumores carcinoides y similares.

En una forma de realización, el gen o producto génico diana es una molécula de longitud completa o un fragmento de la misma seleccionada del grupo que consiste en ABCB1, ABCC2, ABCC4, ABCG2, ABL1, ABL2, AKT1, AKT2, AKT3, ALK, APC, AR, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRAF, BRCA1, BRCA2, C1orf144, CARD11, CBL, CCND1, CCND2, CCND3, CCNE1, CDH1, CDH2, CDH20, CDH5, CDK4, CDK6, CDK8, CDKN2A, CDKN2B, CDKN2C, CEBPA, CHEK1, CHEK2, CRKL, CRLF2, CTNNB1, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DNMT3A, DOT1L, DPYD, EGFR, EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, ERBB2, ERBB3, ERBB4, ERCC2, ERG, ESR1, ESR2, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FCGR3A, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, FDXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GSP1, GUCY1A2, HOXA3, HRAS, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, ITPA, JAK1, JAK2, JAK3, JUN, KDR, KIT, KRAS, LRP1B, LRP2, LTK, MAN1B1, MAP2K1, MAP2K2, MAP2K4, MCL1, MDM2, MDM4, MEN1, MET, MITM, MLH1, MLL, MPL, MRE11A, MSH2, MSH6, MTHFR, MTOR, MUTYH, MYC, MYCL1, MYCN, NF1, NF2, NKX2-1, NOTCH1, NPM1, NQO1, NRAS, NRP2, NTRK1, NTRK3, PAK3, PAX5, PDGFRA, PDGFRB, PIK3CA, PIK3R1, PKHD1, PLCG1, PRKDC, PTCH1, PTEN, PTPN11, PTPRD, RAF1, RARA, RB1, RET, RICTOR, RPTOR, RUNX1, SLC19A1, SLC22A2, SLC01B3, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMO, SOD2, SOX10, SOX2, SRC, STK11, SULT1A1, TBX22, TET2, TGFB2, TMPRSS2, TOP1, TP53, TPMT, TSC1, TSC2, TYMS, UGT1A1, UMPS, USP9X, VHL y WT1.

En una forma de realización, el gen o producto génico diana, o un fragmento del mismo, presenta uno o más SNP que son relevantes a la farmacogenética y farmacogenómica (PGx), por ejemplo, el metabolismo y toxicidad farmacológicos. Entre los genes o productos génicos ejemplificativos se incluyen, aunque sin limitación ABCB1, ABCC2, ABCC4, ABCG2, C1orf144, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DPYD, ERCC2, ESR2, FCGR3A, GSP1, ITPA, LRP2, MAN1B1, MTHFR, NQO1, NRP2, SLC19A1, SLC22A2, SLC01B3, SOD2, SULT1A1, TPMT, TYMS, UGT1A1 y UMPS.

En otra forma de realización, el gen o producto génico diana, o un fragmento del mismo, presenta uno o más codones que están asociados al cáncer. Entre los genes o productos génicos ejemplificativos se incluyen, aunque sin limitación, ABL1 (por ejemplo, el codón 315), AKT1, ALK, APC (por ejemplo, los codones 1114, 1338, 1450 y 1556), AR, BRAF (por ejemplo, el codón 600), CDKN2A, CEBPA, CTNNB1 (por ejemplo, los codones 32, 33, 34, 37, 41 y 45), EGFR (por ejemplo, 719, 746-750, 768, 790, 858 y 861), ERBB2, ESR1, FGFR1, FGFR2, FGFR3, FLT3 (por ejemplo, el codón 835), HRAS (por ejemplo, el codón 12, 13 y 61), JAK2 (por ejemplo, el codón 617), KIT (por ejemplo, el codón 816), KRAS (por ejemplo, los codones 12, 13 y 61), MET, MLL, MYC, NF1, NOTCH1, NPM1, NRAS, PDGFRA, PIK3CA (por ejemplo, los codones 88, 542, 545, 546, 1047 y 1049), PTEN (por ejemplo, los codones 130, 173, 233 y 267), RB1, RET (por ejemplo, el codón 918), TP53 (por ejemplo, 175, 245, 248, 273 y 306).

En todavía otra forma de realización, el gen o producto génico diana, o un fragmento del mismo, está asociado a cáncer. Entre los genes o productos génicos ejemplificativos se incluyen, aunque sin limitación, ABL2, AKT2, AKT3, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRCA1, BRCA2, CARD11, CBL, CCND1, CCND2, CCND3, CCNE1, CDH1, CDH2, CDH20, CDH5, CDK4, CDK6, CDK8, CDKN2B, CDKN2C, CHEK1, CHEK2, CRKL, CRLF2, DNMT3A, DOT1L, EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, ERBB3, ERBB4, ERG, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FGFR4, FLT1, FLT4, FDXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GUCY1A2, HOXA3, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, JAK1, JAK3, JUN, KDR, LRP1B, LTK, MAP2K1, MAP2K2, MAP2K4, MCL1, MDM2, MDM4, MEN1, MITM, MLH1, MPL, MRE11A, MSH2, MSH6, MTOR, MUTYH, MYCL1, MYCN, NF2, NKX2-1, NTRK1, NTRK3, PAK3, PAX5, PDGFRB, PIK3R1, PKHD1, PLCG1, PRKDC, PTCH1, PTPN11, PTPRD, RAF1, RARA, RICTOR, RPTOR, RUNX1, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMO, SOX10, SOX2, SRC, STK11, TBX22, TET2, TGFB2, TMPRSS2, TOP1, TSC1, TSC2, USP9X, VHL y WT1.

Entre las aplicaciones de los métodos anteriores se incluyen la utilización de una biblioteca de oligonucleótidos que contiene todas las variantes de secuencia conocidas (o un subconjunto de ellas) de un gen o genes particulares para la secuenciación en especímenes médicos.

Muestras de ácidos nucleicos

Una diversidad de muestras de tejidos puede ser la fuente de las muestras de ácidos nucleicos utilizadas en los presentes métodos. Pueden aislarse ácidos nucleicos genómicos o subgenómicos (por ejemplo, ADN o ARN) de la muestra de un sujeto (por ejemplo, una muestra tumoral, un tejido contiguo normal (NAT), una muestra de sangre, una muestra que contiene células tumorales circulantes (CTC) o cualquier control normal). En determinadas formas de realización, la muestra de tejido se conserva en forma de una muestra congelada o como preparación de tejido fijada con formaldehído o paraformaldehído incluida en parafina (FFPE). Por ejemplo, la muestra puede incluirse en una matriz, por ejemplo un bloque de FFPE o una muestra congelada. La etapa de aislamiento puede incluir la clasificación en flujo de cromosomas individuales, y/o la microdissección de la muestra de un sujeto (por ejemplo, una muestra tumoral, un NAT o una muestra de sangre).

Una molécula “aislada” de ácidos nucleicos es una que se separa de otras moléculas de ácidos nucleicos que se encuentran presentes en la fuente natural de la molécula de ácidos nucleicos. En determinadas formas de realización, una molécula “aislada” de ácidos nucleicos se encuentra libre de secuencias (tales como secuencias codificantes de proteínas) que flanquean naturalmente el ácido nucleico (es decir, secuencias situadas 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 formas de realización, la molécula aislada de ácidos nucleicos 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 flanquean naturalmente la molécula de ácidos nucleicos en el ADN genómico de la célula de la que se deriva el ácido nucleico. Además, una molécula de ácidos nucleicos “aislada”, tal como una molécula de ADNc, puede encontrarse sustancialmente libre de otro material celular o medio de cultivo al producirla mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otras sustancias químicas al sintetizarla químicamente.

La expresión “sustancialmente libre de otro material celular o medio de cultivo” incluye preparaciones de molécula de ácidos nucleicos en la que la molécula se separa de componentes celulares de las células a partir de las que se aísla o se produce recombinantemente. De esta manera, la molécula de ácidos nucleicos que se encuentra sustancialmente libre de material celular incluye preparaciones de molécula de ácidos nucleicos que presentan menos de aproximadamente 30%, menos de aproximadamente 20%, menos de aproximadamente 10% o menos de aproximadamente 5% (en peso seco) de otro material celular o medio de cultivo.

En determinadas formas de realización, el ácido nucleico se aísla a partir de una muestra envejecida, por ejemplo una muestra FFPE envejecida. La muestra envejecida puede ser, por ejemplo, de años de edad, por ejemplo de 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 de edad o más.

Puede obtenerse una muestra de ácidos nucleicos a partir de muestras de tejidos (por ejemplo, una biopsia o muestra FFPE) de diversos tamaños. Por ejemplo, el ácido nucleico puede aislarse a partir de una muestra de tejido de 5 a 200 µm o más grande. Por ejemplo, la muestra de tejido 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 DNA a partir de una muestra de tejido se proporcionan en el Ejemplo 1. Se dan a conocer métodos adicionales para aislar ácidos nucleicos (por ejemplo, de ADN) a partir de tejidos fijados con formaldehído o paraformaldehído incluidos en parafina (FFPE) en, por ejemplo, Cronin M. et al., Am. J. Pathol. 164(1):35-42, 2004; Masuda N. et al., Nucleic Acids Res. 27(22):4436-4443, 1999; Specht K. et al., Am. J. Pathol. 158(2):419-429, 2001, Ambion RecoverAll™ Total Nucleic Acid Isolation Protocol (Ambion, nº de cat. AM1975, septiembre de 2008), manual técnico del kit de purificación de ADN Maxwell® 16 FFPE Plus LEV (Promega, literatura nº TM349, febrero de 2011), manual del kit de ADN FFPE E.Z.N.A.® (OMEGA bio-tek, Norcross, GA, números de producto D3399-00, D3399-01 y D3399-02; junio de 2009) y manual de tejidos FFPE de ADN QIAamp® (Qiagen, nº de cat. 37625, octubre de 2007). El kit de aislamiento de ácidos nucleicos totales RecoverAll™ utiliza xileno a temperaturas elevadas para solubilizar muestras incluidas en parafina y un filtro de fibra de vidrio para capturar los ácidos nucleicos. Se utiliza el kit de purificación de ADN Maxwell® 16 FFPE Plus LEV con el instrumento Maxwell® 16 para la purificación del ADN genómico a partir de secciones de 1 a 10 µm de tejido FFPE. El ADN se purifica utilizando partículas paramagnéticas (PMP) revestidas de sílice y eluyendo en un volumen de elución pequeño. El kit de ADN FFPE E.Z.N.A.® utiliza una columna de centrifugación y un sistema de amortiguadores para el aislamiento del ADN genómico. El kit de tejidos FFPE para ADN QIAamp® utiliza tecnología micro de ADN QIAamp® para la purificación de ADN genómico y mitocondrial. Los protocolos para el aislamiento de ADN a partir de sangre se dan a conocer en, por ejemplo, el kit de ADN sanguíneo Maxwell® 16 LEV y en el manual técnico del kit de purificación de ADN de bajo volumen de elución (LEV) a partir de hisopo bucal Maxwell 16 (Promega, nº de literatura #TM333, 1 de enero de 2011).

Los protocolos para el aislamiento de ARN se dan a conocer en, por ejemplo, la ficha técnica del kit de purificación de ARN total Maxwell® 16 (Promega, nº de literatura TB351, agosto de 2009).

Las muestras aisladas de ácidos nucleicos (por ejemplo, muestras de ADN genómico) pueden fragmentarse o romperse mecánicamente mediante la puesta en práctica de técnicas rutinarias. Por ejemplo, el ADN genómico puede fragmentarse mediante métodos de rotura mecánica físicos, métodos de corte enzimático, métodos de corte químico y otros métodos bien conocidos por el experto en la materia. La biblioteca de ácidos nucleicos puede 5 contener la totalidad, o sustancialmente la totalidad, de la complejidad del genoma. La expresión "sustancialmente la totalidad" en el presente contexto se refiere a la posibilidad de que puede producirse, en la práctica, alguna pérdida no deseada de complejidad del genoma durante las etapas iniciales del procedimiento. Los métodos indicados en la presente memoria también resultan útiles en casos en que la biblioteca de ácidos nucleicos es una parte del genoma; es decir, en donde la complejidad del genoma se reduce por el diseño. En algunas formas de 10 realización, puede utilizarse cualquier parte seleccionada del genoma con los métodos indicados en la presente memoria. En determinadas formas de realización, se aísla el exoma entero o un subconjunto del mismo.

Entre los métodos proporcionados en la invención pueden incluirse además el aislamiento de una muestra de ácidos nucleicos para proporcionar una biblioteca (por ejemplo, una biblioteca de ácidos nucleicos tal como se 15 indica en la presente memoria). En determinadas formas de realización, la muestra de ácidos nucleicos incluye genomas completos, fragmentos subgenómicos, o ambos. Las muestras de ácidos nucleicos aisladas pueden utilizarse para preparar bibliotecas de ácidos nucleicos. De esta manera, en una forma de realización, los métodos proporcionados en la invención incluyen además el aislamiento de una muestra de ácidos nucleicos para proporcionar una biblioteca (por ejemplo, una biblioteca de ácidos nucleicos tal como se indica en la presente 20 memoria). Los protocolos para aislar y preparar bibliotecas a partir de genomas completos o fragmentos subgenómicos son conocidos en la técnica (por ejemplo, el kit de preparación de muestras de ADN genómico de Illumina). En determinadas formas de realización, el ADN genómico o fragmento de ADN subgenómico se aísla a partir de una muestra del sujeto (por ejemplo, una muestra tumoral, un tejido contiguo normal (NAT), una muestra de sangre o cualquier control normal). En una forma de realización, la muestra (por ejemplo, la muestra de tumor o NAT) es un espécimen conservado. Por ejemplo, la muestra se incluye en una matriz, por ejemplo un bloque de 25 FFPE o una muestra congelada. En determinadas formas de realización, la etapa de aislamiento incluye la clasificación en flujo de cromosomas individuales, y/o la microdissección de la muestra de un sujeto (por ejemplo, una muestra tumoral, un NAT o una muestra de sangre). En determinadas formas de realización, la muestra de ácidos nucleicos utilizada para generar la biblioteca de ácidos nucleicos pesa menos de 5 microgramos, menos de 30 1 microgramo o menos de 500 ng, menos de 200 ng, menos de 100 ng, menos de 50 ng, menos de 10 ng, menos de 5 ng o menos de 1 ng.

En todavía otras formas de realización, la muestra de ácidos nucleicos utilizada para generar la biblioteca incluye 35 ARN o ADNc derivado de ARN. En algunas formas de realización, el ARN incluye ARN celular total. En otras formas de realización, determinadas secuencias de ARN abundantes (por ejemplo, ARN ribosómicos) se han empobrecido. En algunas formas de realización, la fracción de ARNm con cola de poli(A) en la preparación de ARN total ha sido enriquecida. En algunas formas de realización, el ADNc se produce mediante métodos de síntesis de ADNc con cebadores aleatorios. En otras formas de realización, la síntesis de ADNc se inicia en la cola poli(A) de 40 ARNm maduros mediante el cebado con oligonucleótidos que contienen oligo(dT). Los métodos de empobrecimiento, enriquecimiento en poli(A) y síntesis de ADNc son bien conocidos por el experto en la materia.

El método puede incluir además amplificar la muestra de ácidos nucleicos mediante métodos de amplificación de ácidos nucleicos específicos o no específicos que son bien conocidos por el experto en la materia. En algunas 45 formas de realización, determinadas formas de realización, la muestra de ácidos nucleicos se amplifica mediante, por ejemplo, métodos de amplificación de genomas completos, tales como la amplificación por desplazamiento de cadena con cebadores aleatorios.

En otras formas de realización, la muestra de ácidos nucleicos se fragmenta o se rompe mecánicamente mediante 50 métodos físicos o enzimáticos y se liga a adaptadores sintéticos, se selecciona según tamaño (por ejemplo, mediante electroforesis en gel preparativa) y se amplifica (por ejemplo, mediante PCR). En otras formas de realización, el grupo fragmentado y ligado con adaptadores de ácidos nucleicos se utiliza sin selección de tamaño o amplificación explícita antes de la selección de híbridos.

En otras formas de realización, el ADN aislado (por ejemplo, el ADN genómico) se fragmenta o se rompe 55 mecánicamente. En algunas formas de realización, la biblioteca incluye menos de 50% de ADN genómico, tal como una subfracción de ADN genómico que es una representación reducida o una parte definida de un genoma, por ejemplo que ha sido subfraccionada por otros medios. En otras formas de realización, la biblioteca incluye la totalidad o sustancialmente la totalidad del ADN genómico.

En algunas formas de realización, la biblioteca incluye menos de 50% de ADN genómico, tal como una subfracción 60 de ADN genómico que es una representación reducida o una parte definida de un genoma, por ejemplo que ha sido subfraccionada por otros medios. En otras formas de realización, la biblioteca incluye la totalidad o sustancialmente la totalidad del ADN genómico. Los protocolos para aislar y preparar bibliotecas a partir de genomas completos o fragmentos subgenómicos son conocidos en la técnica (por ejemplo, el kit de preparación 65 de muestras de ADN genómico de Illumina) y se describen en la presente memoria como los Ejemplos 2A, 2B y 3. Se describen en la presente memoria métodos alternativos de rotura mecánica del ADN, en el Ejemplo 2B. Por

ejemplo, los métodos de rotura mecánica de ADN alternativos pueden ser más automatizables y/o más eficientes (por ejemplo, con muestras FFPE degradadas). Las alternativas a los métodos de rotura mecánica del ADN también pueden utilizarse para evitar la etapa de ligación durante la preparación de bibliotecas.

5 Los métodos indicados en la presente memoria pueden llevarse a cabo utilizando una pequeña cantidad de ácidos nucleicos, por ejemplo en el caso de que la cantidad de ADN fuente sea limitante (por ejemplo, incluso después de la amplificación de genomas completos). En una forma de realización, 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 ácidos nucleicos. Por ejemplo, uno puede iniciarse típicamente con  
10 50 a 100 ng de ADN genómico. Puede empezarse con menos; sin embargo, en el caso de que se amplifique el ADN genómico (por ejemplo, utilizando PCR) antes de la etapa de hibridación, por ejemplo la hibridación en solución. De esta manera resulta posible, aunque no resulta esencial, amplificar el ADN genómico antes de la hibridación, por ejemplo la hibridación en solución.

15 La muestra de ácidos nucleicos utilizada para generar la biblioteca puede incluir además ARN o ADNc derivado de ARN. En algunas formas de realización, el ARN incluye ARN celular total. En otras formas de realización, determinadas secuencias de ARN abundantes (por ejemplo, ARN ribosómicos) se han empobrecido. En otras formas de realización, la fracción de ARNm con cola de poli(A) en la preparación de ARN total ha sido enriquecida. En algunas formas de realización, el ADNc se produce mediante métodos de síntesis de ADNc con cebadores  
20 aleatorios. En otras formas de realización, la síntesis de ADNc se inicia en la cola poli(A) de ARNm maduros mediante el cebado con oligonucleótidos que contienen oligo(dT). Los métodos de empobrecimiento, enriquecimiento en poli(A) y síntesis de ADNc son bien conocidos por el experto en la materia.

25 El método puede incluir además amplificar la muestra de ácidos nucleicos mediante métodos de amplificación de ácidos nucleicos específicos o no específicos que son bien conocidos por el experto en la materia. La muestra de ácidos nucleicos puede amplificarse mediante, por ejemplo, métodos de amplificación de genomas completos, tales como la amplificación por desplazamiento de cadena con cebadores aleatorios.

30 La muestra de ácidos nucleicos puede fragmentarse o romperse mecánicamente mediante métodos físicos o enzimáticos tal como se indica en la presente memoria, y ligarse a adaptadores sintéticos, seleccionarse según tamaño (por ejemplo, mediante electroforesis en gel preparativa) y se amplificarse (por ejemplo, mediante PCR). El grupo fragmentado y ligado con adaptadores de ácidos nucleicos se utiliza sin selección de tamaño o amplificación explícita antes de la selección de híbridos.

### 35 *Miembros de biblioteca*

“Miembro” o “miembro de biblioteca” u otro término similar, tal como se utiliza en la presente memoria, se refiere a una molécula de ácidos nucleicos, por ejemplo un ADN, ARN, que es el miembro que pertenece a una biblioteca (o “captura de biblioteca”). El miembro de biblioteca puede ser uno o más de un miembro tumoral, un miembro de  
40 referencia o un miembro de PGx. Típicamente, un miembro es una molécula de ADN, por ejemplo una molécula de ADN genómico o ADNc. Un miembro puede ser fragmentado, por ejemplo enzimáticamente o mediante la rotura mecánica de ADN genómico. Los miembros pueden comprender una secuencia de ácidos nucleicos de un sujeto y puede comprender además una secuencia de nucleótidos no derivada del sujeto, por ejemplo cebadores o adaptadores (por ejemplo, para la amplificación por PCR o para la secuenciación) o secuencias que permiten la  
45 identificación de una muestra, por ejemplo secuencias “de código de barras”.

Tal como se utiliza en la presente memoria, “miembro diana” se refiere a una molécula de ácidos nucleicos que uno desea aislar a partir de la biblioteca de ácidos nucleicos. En una forma de realización, los miembros diana pueden ser un miembro tumoral, un miembro de referencia o un miembro de PGx, tal como se indica en la presente  
50 memoria. Los miembros que se seleccionan realmente de la biblioteca de ácidos nucleicos se denominan en la presente memoria “captura de biblioteca”. En una forma de realización, la captura de biblioteca incluye una selección o enriquecimiento en miembros de la biblioteca, por ejemplo la salida enriquecida o seleccionada de una biblioteca después de una o más rondas de captura de híbridos tal como se indica en la presente memoria.

55 Los miembros diana pueden ser un subgrupo de la biblioteca, es decir, que no todos los miembros de la biblioteca se seleccionan mediante cualquier uso particular de los procedimientos indicados en la presente memoria. En otras formas de realización, los miembros diana se encuentran dentro de una región diana deseada. Por ejemplo, los miembros diana pueden ser, en algunas formas de realización, un porcentaje de los miembros de biblioteca que es de tan solo 10% o de hasta 95%-98% o más elevado. En una forma de realización, la captura de biblioteca  
60 incluye por lo menos aproximadamente 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 99,9% o más de los miembros diana. En otra forma de realización, la biblioteca contiene 100% de los miembros diana. En una forma de realización, la pureza de la captura de biblioteca (porcentaje de lecturas que se alinean con las dianas) es de por lo menos aproximadamente 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 99,9% o más.

65

Los miembros diana (o la captura de biblioteca) obtenidos a partir del ADN genómico pueden incluir una fracción pequeña del ADN genómico total, de manera que incluye menos de aproximadamente 0,0001%, por lo menos aproximadamente 0,0001%, por lo menos aproximadamente 0,001%, por lo menos aproximadamente 0,01% o por lo menos aproximadamente 0,1% de ADN genómico o una fracción más significativa del ADN genómico total, de manera que incluye por lo menos aproximadamente 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% o 10% del ADN genómico o más de 10% de ADN genómico.

En una forma de realización, los miembros diana (o la captura de biblioteca) se seleccionan de una mezcla compleja de genoma. Por ejemplo, la selección del ADN de un tipo celular (por ejemplo, células de cáncer) procedente de una muestra que contiene el ADN de otros tipos celulares (por ejemplo, células normales). En dichas aplicaciones, el miembro diana puede incluir menos de 0,0001%, por lo menos 0,0001%, por lo menos aproximadamente 0,001%, por lo menos aproximadamente 0,01% o por lo menos aproximadamente 0,1% de la complejidad total de la secuencia de ácidos nucleicos presentes en la muestra compleja, o una fracción más significativa de manera que incluye por lo menos aproximadamente 1%, 2%, 5%, 10% o más de 10% de la complejidad total de secuencias de ácidos nucleicos presentes en la muestra compleja.

En una forma de realización, el miembro diana (o la captura de biblioteca) seleccionada mediante los métodos indicados en la presente memoria (por ejemplo, los métodos de selección de hibridación en solución) incluyen la totalidad o una parte de los exones en un genoma, tal como más de aproximadamente 0,1%, 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o 95% de los exones genómicos. En otra forma de realización, el miembro diana (o la captura de biblioteca) puede ser un grupo específico de exones, por ejemplo por lo menos aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 exones particulares, por ejemplo exones asociados a enfermedades particulares, tales como cáncer. En todavía otra forma de realización, el miembro diana (o la captura de biblioteca) contiene exones u otras partes de genes de interés seleccionados. La utilización de secuencias de señuelo específicas permite al profesional seleccionar secuencias diana (conjunto ideal de secuencias seleccionadas) y subgrupos de ácidos nucleicos (conjunto real de secuencias seleccionadas) que contienen tantos o tan pocos exones (u otras secuencias) de un grupo de ácidos nucleicos para una selección particular.

En una forma de realización, el miembro diana (o la captura de biblioteca) incluye un conjunto de ADNc. Los ADNc de captura pueden utilizarse, por ejemplo, para encontrar variantes de corte y empalme y para identificar transcritos de fusión (por ejemplo, de traslocaciones de ADN genómico). En otra forma de realización, el miembro diana (y la captura de biblioteca) se utiliza para encontrar cambios de bases únicas y otros cambios de la secuencia expresados en la fracción de ARN de una célula, tejido u órgano, por ejemplo en un tumor.

El miembro diana (o la captura de biblioteca) (por ejemplo, exones, ADNc y otras secuencias) pueden estar relacionadas o no relacionadas, según se desee. Por ejemplo, puede obtenerse un miembro diana seleccionado (y la captura de biblioteca) a partir de un grupo de ácidos nucleicos que son genes implicados en una enfermedad, tal como un grupo de genes implicado en una o más enfermedades, tales como cánceres, un grupo de ácidos nucleicos que contienen SNP específicos.

En una forma de realización, una parte o la totalidad de los miembros de biblioteca comprende una secuencia adaptadora a una no diana. La secuencia adaptadora puede resultar útil, por ejemplo, para un método de secuenciación (por ejemplo, un método de NGS), para la amplificación, para la transcripción inversa o para la clonación en un vector. La secuencia adaptadora puede estar localizada en un extremo o en ambos. Los adaptadores pueden ligarse en el extremo 5'- o 3'-3 de la inserción de biblioteca, por ejemplo tal como se indica en los Ejemplos adjuntos. Pueden obtenerse adaptadores de proveedores comerciales, tales como NimbleGen (Roche), Integrated DNA Technologies (IDT) para oligos de ADN, o Agilent Technologies.

Puede diseñarse un oligonucleótido bloqueante complementario a los adaptadores y prepararse mediante métodos conocidos en la técnica, por ejemplo métodos de síntesis de oligos. También pueden obtenerse oligonucleótidos bloqueantes a partir de proveedores comerciales, tales como NimbleGen (Roche), Integrated DNA Technologies (IDT) para oligos de ADN, o Agilent Technologies. La longitud y composición de dichos adaptadores pueden ajustarse para, por ejemplo, modificar la interacción de unión (por ejemplo, una  $T_m$  tal como se indica en la presente memoria) con el adaptador complementario siguiendo métodos conocidos en la técnica.

Los oligonucleótidos bloqueantes pueden incluir ADN, ARN o una combinación de ambos. Los oligonucleótidos de ADN o ARN pueden ser naturales o no naturales. En determinadas formas de realización, los oligonucleótidos bloqueantes incluyen uno o más nucleótidos no naturales para, por ejemplo, incrementar la temperatura de fusión. Entre los oligonucleótidos no naturales ejemplificativos se incluyen nucleótidos de ADN o ARN modificado. Un nucleótido de ARN modificado ejemplificativo es un ácido nucleico bloqueado (ANB), en el que la fracción ribosa de un nucleótido 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., "Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes", *Biochemistry* 45 (23): 7347-55, 2006). Entre otros nucleótidos de ADN y ARN ejemplificativos modificados se incluyen, aunque sin limitarse a ellos, ácido péptido nucleico (APN) compuesto de unidades de N-(2-aminoetil)-glicina repetitivas unidas mediante enlaces

peptídicos (Egholm, M. et al. Nature 365 (6446): 566-8, 1993); un oligonucleótido de ADN o ARN modificado para capturar regiones de bajo contenido de GC, un ácido nucleico bicíclico (ANBi) o un oligonucleótido entrecruzado, una 5-metil-desoxicitidina modificada y 2,6-diaminopurina. Se conocen en la técnica otros nucleótidos de ADN y ARN modificados.

#### Diseño y construcción de señuelos

Un señuelo puede ser una molécula de ácidos nucleicos, por ejemplo una molécula de ADN o ARN, que puede hibridarse con (por ejemplo, ser complementaria a), y de esta manera permitir la captura de, un ácido nucleico diana. En una forma de realización, el señuelo es una molécula de ARN. En otras formas de realización, un señuelo incluye una entidad de unión, por ejemplo, una etiqueta de afinidad, que permite la captura y separación, por ejemplo, mediante la unión a una entidad de unión, de un híbrido formado por un señuelo y un ácido nucleico hibridado con el señuelo. En una forma de realización, un señuelo resulta adecuado para la hibridación en fase solución.

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

Los señuelos de ARN pueden prepararse tal como se indica en otro sitio en la presente memoria, utilizando métodos conocidos en la técnica, incluyendo, aunque sin limitación, síntesis química *de novo* y transcripción de moléculas de ADN utilizando una ARN polimerasa dependiente de ADN. En una forma de realización, la secuencia de señuelo se produce utilizando métodos conocidos de amplificación de ácidos nucleicos, tales como PCR, por ejemplo, utilizando muestras de ADN humano o de ADN humano agrupado como el molde. A continuación, los oligonucleótidos pueden convertirse en señuelos de ARN. En una forma de realización, se utiliza la transcripción *in vitro*, por ejemplo basada en la adición de una secuencia de promotor de ARN polimerasa en un extremo del oligonucleótido. En una forma de realización, se añade la secuencia de promotor de ARN polimerasa en el extremo del señuelo mediante amplificación o reamplificación de la secuencia señuelo, por ejemplo utilizando PCR u otros métodos de amplificación de ácidos nucleicos, por ejemplo mediante adición a la cola de un cebador de cada pareja de cebadores específica de diana con una secuencia de promotor de ARN. En una forma de realización, la ARN polimerasa es una polimerasa de T7, una polimerasa SP6 o una polimerasa de T3. En una forma de realización, se marca el señuelo de ARN con una etiqueta, por ejemplo una etiqueta de afinidad. En una forma de realización, el señuelo de ARN se prepara mediante transcripción *in vitro*, por ejemplo utilizando UTP biotinilado. En otra forma de realización, el señuelo de ARN se produce sin biotina y después la biotina se entrecruza con la molécula de ARN utilizando métodos bien conocidos en la técnica, tal como el entrecruzamiento con psoraleno. En una forma de realización, el señuelo de ARN es una molécula de ARN resistente a ARNasa, que puede prepararse mediante, por ejemplo, la utilización de nucleótidos modificados durante la transcripción para producir una molécula de ARN que resiste a la degradación por ARNasa. En una forma de realización, el señuelo de ARN corresponde a únicamente una cadena de la diana de ADN de doble cadena. Típicamente, dichos señuelos de ARN no son autocomplementarios y resultan más eficaces como controladores de hibridación.

Los conjuntos de señuelos pueden diseñarse a partir de secuencias de referencia, de manera que los señuelos sean óptimos para seleccionar dianas de las secuencias de referencia. En algunas formas de realización, se diseñan secuencias de señuelo utilizando una base mixta (por ejemplo, degeneración). Por ejemplo, la base o bases mixtas pueden incluirse en la secuencia del señuelo en una o más posiciones de un SNP común o mutación, a fin de optimizar las secuencias de señuelo para capturar ambos alelos (por ejemplo, SNP y no SNP, mutantes y no mutantes). En algunas formas de realización, todas las variaciones de secuencia (o un subconjunto de ellas) pueden ser la diana con múltiples señuelos oligonucleótidos, en lugar de mediante la utilización de oligonucleótidos degenerados mixtos.

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

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

En algunas formas de realización, el conjunto de señuelos incluye una entidad de unión. La entidad de unión puede ser una etiqueta de afinidad en cada secuencia de señuelo. En algunas formas de realización, la etiqueta de afinidad es una molécula de biotina o un hapteno. En determinadas formas de realización, la entidad de unión permite la separación de los híbridos de señuelo/miembro a partir de la mezcla de hibridación mediante la unión a

una pareja, 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 formas de realización, los oligonucleótidos en el conjunto de señuelos contienen secuencias complementarias directa e inversa de la secuencia del mismo miembro diana, de manera que los oligonucleótidos con secuencias específicas de miembro complementario inverso también portan colas universales complementarias inversas. Lo anterior puede llevar a transcritos de ARN que son la misma cadena, es decir, que no son complementarias entre sí.

En otras formas de realización, el conjunto de señuelos incluye oligonucleótidos que contienen bases degeneradas o mixtas en una o más posiciones. En todavía otras formas de realización, el conjunto de señuelos incluye múltiples o sustancialmente la totalidad de las variantes de secuencia conocidas presentes en una población de una sola especie o comunidad de organismos. En una forma de realización, el conjunto de señuelos incluye múltiples o sustancialmente todas las variantes de secuencia conocidas presentes en una población humana.

En otras formas de realización, el conjunto de señuelos incluye secuencias de ADNc o se deriva de secuencias de ADNc. En otras formas de realización, el conjunto de señuelos incluye productos de amplificación (por ejemplo, productos de PCR) que se amplifican a partir de ADN genómico, ADNc o ADN clonado.

En otras formas de realización, el conjunto de señuelos incluye moléculas de ARN. En algunas formas de realización, el conjunto incluye moléculas de ARN química o enzimáticamente modificadas, o transcritas *in vitro*, incluyendo, aunque sin limitación, las que son más estables y resistentes a ARNasa.

En todavía otras formas de realización, los señuelos se producen mediante métodos descritos en el documento nº US 2010/0029498 y Gnirke A. et al., Nat. Biotechnol. 27(2):182-189, 2009) Por ejemplo, pueden producirse señuelos de ARN biotinilado mediante la obtención de un reservorio de oligonucleótidos largos sintéticos, sintetizados originalmente en una micromatriz, y la amplificación de los oligonucleótidos para producir las secuencias de señuelo. En algunas formas de realización, los señuelos se producen mediante adición de una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de señuelo y síntesis de las secuencias de ARN utilizando ARN polimerasa. En una forma de realización, pueden obtenerse bibliotecas de oligodesoxinucleótidos sintéticos a partir de proveedores comerciales, tales como Agilent Technologies, Inc., y amplificarse utilizando métodos de amplificación de ácidos nucleicos conocidos.

De acuerdo con lo anterior, se proporciona un método de producción del conjunto de señuelos anteriormente indicado. El método incluye seleccionar una o más secuencias oligonucleótidas de señuelo específicas de diana (por ejemplo, una o más secuencias oligonucleótidas de mutación, captura, referencia o control tal como se indican en la presente memoria); obtener un reservorio de secuencias oligonucleótidas de señuelo específicas de diana (por ejemplo, sintetizando el reservorio de secuencias oligonucleótidas de señuelo específicas de diana, por ejemplo mediante síntesis en micromatriz) y opcionalmente, amplificar los oligonucleótidos para producir el conjunto de señuelos.

En otras formas de realización, los métodos incluyen además amplificar (por ejemplo, mediante PCR) los oligonucleótidos utilizando uno o más cebadores biotinilados. En algunas formas de realización, los oligonucleótidos incluyen una secuencia universal en el extremo de cada oligonucleótido unido a la micromatriz. Los métodos pueden incluir además eliminar las secuencias universales respecto de los oligonucleótidos. Dichos métodos pueden incluir además la cadena complementaria de los oligonucleótidos, hibridar los oligonucleótidos y extender los oligonucleótidos. En algunas de dichas formas de realización, los métodos de amplificación (por ejemplo, mediante PCR) de los oligonucleótidos utilizan uno o más cebadores biotinilados. En algunas formas de realización, el método incluye además seleccionar según el tamaño los oligonucleótidos amplificados.

En una forma de realización, se produce un conjunto de señuelos de ARN. Los métodos incluyen producir un conjunto de secuencias de señuelo según el método descrito en la presente memoria, añadir una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de señuelo y sintetizar secuencias de ARN utilizando la ARN polimerasa. La ARN polimerasa puede seleccionarse de una ARN polimerasa de T7, una ARN polimerasa SP6 o una ARN polimerasa de T3. En otras formas de realización, la secuencia de promotor de ARN polimerasa se añade en los extremos de las secuencias de señuelo mediante amplificación (por ejemplo, mediante PCR) de las secuencias de señuelo. En formas de realización en las que las secuencias de señuelo se amplifican por PCR con parejas de cebadores específicos a partir de ADN genómico o ADNc añadiendo una secuencia de promotor de ARN al extremo 5' de uno de los dos cebadores específicos en cada pareja conducirá a un producto de PCR que puede transcribirse en un señuelo de ARN utilizando métodos estándares.

En otras formas de realización, pueden producirse conjuntos de señuelos utilizando muestras de ADN humano o de ADN humano agrupado a modo de molde. En dichas formas de realización, los oligonucleótidos se amplifican mediante reacción en cadena de la polimerasa (PCR). En otras formas de realización, los oligonucleótidos amplificados se reamplifican mediante amplificación en círculo rodante o amplificación en círculo rodante hiperramificado. También pueden utilizarse los mismos métodos para producir secuencias de señuelo utilizando

muestras de ADN humano o ADN humano agrupado a modo de molde. También pueden utilizarse los mismos métodos para producir secuencias de señuelo utilizando subfracciones de un genoma obtenido mediante otros métodos, incluyendo, aunque sin limitación, la digestión de restricción, la electroforesis en gel de campo pulsado, la separación en flujo, la centrifugación en gradiente de densidad de CsCl, la reasociación cinética selectiva, la microdissección de preparaciones de cromosomas y otros métodos de fraccionamiento conocidos por el experto en la materia.

En determinadas formas de realización, el número de señuelos en el conjunto de señuelos es inferior a 1,000. En otras formas de realización, el número de señuelos en el conjunto de señuelos es superior a 1,000, superior a 5,000, superior a 10,000, superior a 20,000, superior a 50,000, superior a 100,000 o superior a 500,000.

La longitud de la secuencia de señuelo puede ser de entre aproximadamente 70 nucleótidos y 1000 nucleótidos. En una forma de realización, la longitud de señuelo es de entre aproximadamente 100 y 300 nucleótidos, de entre 110 y 200 nucleótidos o de entre 120 y 170 nucleótidos de longitud. Además de las mencionadas anteriormente, pueden utilizarse longitudes de oligonucleótido 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 en los métodos indicados en la presente memoria. En algunas formas de realización, pueden utilizarse 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 señuelo puede incluir una secuencia de señuelo específica de diana (por ejemplo, específica de miembro) y colas universales en uno o en ambos extremos. Tal como se utiliza en la presente memoria, la expresión "secuencia de señuelo" puede referirse a la secuencia de señuelo específica de diana o al oligonucleótido entero, incluyendo la "secuencia de señuelo" específica de diana y otros nucleótidos del oligonucleótido. Las secuencias específicas de diana en los señuelos presentan una longitud de entre aproximadamente 40 nucleótidos y 1000 nucleótidos. En una forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 70 nucleótidos y 300 nucleótidos. En otra forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 100 nucleótidos y 200 nucleótidos. En todavía otra forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 120 nucleótidos y 170 nucleótidos, típicamente de 120 nucleótidos. También pueden utilizarse longitudes intermedias además de las indicadas anteriormente, en los métodos indicados en la presente memoria, 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 anteriormente indicadas.

En una forma de realización, el señuelo es un oligómero (por ejemplo, que comprende oligómeros de ARN, oligómeros de ADN o una combinación de los mismos) de entre aproximadamente 50 y 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 forma de realización, cada oligómero de señuelo incluye aproximadamente 120 o 170, o típicamente aproximadamente 120 nucleótidos, que son una secuencia de señuelo específica de diana. El señuelo puede comprender secuencias adicionales de nucleótidos no específicas de diana en uno o en ambos extremos. Las secuencias adicionales de nucleótidos pueden utilizarse para, por ejemplo, la amplificación por PCR o como un identificador del señuelo. En determinadas formas de realización, el señuelo comprende adicionalmente una entidad de unión tal como se indica en la presente memoria (por ejemplo, una etiqueta de captura, tal como una molécula de biotina). La entidad de unión, por ejemplo, la molécula de biotina, puede unirse al señuelo, por ejemplo en el extremo 5', 3' o internamente (por ejemplo, mediante la incorporación de un nucleótido biotinilado) del señuelo. En una forma de realización, la molécula de biotina se une al extremo 5' del señuelo.

En una forma de realización ejemplificativa, el señuelo es un oligonucleótido de aproximadamente 150 nucleótidos de longitud, de los que 120 nucleótidos son "secuencia de señuelo" específica de diana. Los otros 30 nucleótidos (por ejemplo, 15 nucleótidos en cada extremo) son colas arbitrarias universales utilizadas para la amplificación por PCR. Las colas pueden ser cualquier secuencia seleccionada por el usuario. Por ejemplo, el reservorio de oligonucleótidos sintéticos puede incluir oligonucleótidos de secuencia 5'-ATCGCACCAGCGTGTN<sub>120</sub>CACTGCGGCTCCTCA-3' (SEC ID n° 100), en donde N<sub>120</sub> indica las secuencias de señuelo específicas de diana.

Las secuencias de señuelo indicadas en la presente memoria pueden utilizarse para la selección de exones y secuencias diana cortas. En una forma de realización, el señuelo presenta una longitud de entre aproximadamente 100 nucleótidos y 300 nucleótidos. En otra forma de realización, el señuelo presenta una longitud de entre aproximadamente 130 nucleótidos y 230 nucleótidos. En todavía otra forma de realización, el señuelo presenta una longitud de entre aproximadamente 150 nucleótidos y 200 nucleótidos. Las secuencias específicas de diana en los señuelos, por ejemplo para la selección de exones y secuencias diana cortas, presentan una longitud de entre aproximadamente 40 nucleótidos y 1000 nucleótidos. En una forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 70 nucleótidos y 300 nucleótidos. En otra forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 100 nucleótidos y 200 nucleótidos. En todavía otra forma de realización, la secuencia específica de diana presenta una longitud de entre aproximadamente 120 nucleótidos y 170 nucleótidos.

En algunas formas de realización, los oligonucleótidos largos pueden minimizar el número de oligonucleótidos necesarios para capturar las secuencias diana. Por ejemplo, puede utilizarse un oligonucleótido por cada exón. Es conocido en la técnica que las medias y medianas de longitud de los exones codificantes de proteína en el genoma humano son de aproximadamente 164 y 120 pares de bases, respectivamente. Los señuelos más largos pueden ser más específicos y capturar mejor que los más cortos. En consecuencia, la tasa de éxito por cada secuencia de señuelo oligonucleótido es más elevada que con oligonucleótidos cortos. En una forma de realización, la secuencia cubierta por señuelo mínima es el tamaño de un señuelo (por ejemplo, 120 a 170 bases), por ejemplo para capturar dianas del tamaño de un exón. Para determinar la longitud de las secuencias de señuelo, también puede tomarse en consideración que los señuelos innecesariamente largos capturan más ADN no deseado directamente contiguo a la diana. Los señuelos oligonucleótidos más largos también pueden ser más tolerantes a polimorfismos en la región diana en las muestras de ADN que señuelos oligonucleótidos más cortos. Típicamente, las secuencias de señuelo se derivan de una secuencia genómica de referencia. En el caso de que la secuencia diana en la muestra de ADN real se aparte de la secuencia de referencia, por ejemplo en el caso de que contenga un polimorfismo de nucleótido único (SNP), puede hibridarse menos eficientemente con el señuelo y, por lo tanto, puede encontrarse infrarrepresentado o completamente ausente en las secuencias hibridadas con las secuencias de señuelo. Las pérdidas alélicas debido a SNP pueden ser menos probables con las moléculas de señuelo sintéticas más largas por el motivo de que un único apareamiento incorrecto en, por ejemplo, 120 a 170 bases puede presentar menos efecto sobre la estabilidad del híbrido que un único apareamiento incorrecto en 20 o 70 bases, que son las longitudes típicas de señuelo o cebador en la amplificación multiplex y la captura en micromatriz, respectivamente.

Para la selección de dianas que son largas en comparación con la longitud de los señuelos de captura, dichas regiones genómicas, las longitudes de secuencia de señuelo se encuentran típicamente en el mismo intervalo de tamaños que los señuelos para dianas cortas indicados anteriormente, excepto en el caso de que no exista necesidad de limitar el tamaño máximo de secuencias de señuelo con el único propósito de minimizar el reconocimiento de secuencias contiguas. Alternativamente, pueden titularse oligonucleótidos en una ventana mucho más amplia (típicamente, 600 bases). Dicho método puede utilizarse para capturar fragmentos de ADN que son mucho más grandes (por ejemplo, de aproximadamente 500 bases) que un exón típico. En consecuencia, se seleccionan muchas más secuencias no diana flanqueantes no deseadas.

#### Síntesis de señuelos

Los señuelos pueden ser cualquier tipo de oligonucleótido, por ejemplo de ADN o ARN. Los señuelos de ADN o ARN ("señuelos oligo") pueden sintetizarse individualmente o pueden sintetizarse en una matriz, en forma de un conjunto de señuelos de ADN o ARN ("señuelos de matriz"). Un señuelo oligo, proporcionado en un formato de matriz, o en forma de oligo aislado, típicamente es de cadena sencilla. El señuelo puede comprender adicionalmente una entidad de unión tal como se indica en la presente memoria (por ejemplo, una etiqueta de captura, tal como una molécula de biotina). La entidad de unión, por ejemplo, la molécula de biotina, puede unirse al señuelo, por ejemplo en el extremo 5' o 3' del señuelo, típicamente en el extremo 5' del señuelo.

En algunas formas de realización, pueden añadirse señuelos oligo individuales a un conjunto de señuelos en matriz. En estos casos, los señuelos oligo pueden diseñarse para reconocer las mismas zonas que las reconocidas por los señuelos de matriz, y pueden diseñarse señuelos oligo adicionales y añadirse a los señuelos de matriz estándares para conseguir una cobertura mejorada, o más completa, en determinadas zonas del genoma. Por ejemplo, pueden diseñarse señuelos oligo adicionales para reconocer zonas de baja cobertura de secuenciación después de una ronda inicial de secuenciación con un conjunto estándar de señuelos en matriz. En algunas formas de realización, los señuelos oligo están diseñados para presentar un efecto de embaldosado en la zona de cobertura para el conjunto de señuelos en matriz, o un efecto de embaldosado sobre la zona de cobertura para otros señuelos oligo.

En una forma de realización, los señuelos oligo individuales son oligos de ADN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ADN, o una combinación de los mismos (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente). En otras formas de realización, los señuelos oligo individuales son oligos de ADN que se utilizan para complementar un conjunto de señuelos de oligo de ARN o ADN, o una combinación de los mismos, que es una colección de oligos diseñados y sintetizados individualmente. En una forma de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ARN o ADN, o una combinación de los mismos (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente). En otras formas de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ARN o ADN, o una combinación de los mismos, que es una colección de oligos diseñados y sintetizados individualmente.

En todavía otra forma de realización, los señuelos oligo individuales son oligos de ADN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ADN (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente) y, en otras formas de realización, los señuelos oligo individuales son oligos de

ADN que se utilizan para complementar un conjunto de señuelos oligos de ADN que es una colección de oligos diseñados y sintetizados individualmente.

5 En todavía otra forma de realización, los señuelos oligo individuales son oligos de ADN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ARN (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente) y, en otras formas de realización, los señuelos oligo individuales son oligos de ADN que se utilizan para complementar un conjunto de señuelos oligos de ARN que es una colección de oligos diseñados y sintetizados individualmente.

10 En todavía otra forma de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ARN (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente) y, en otras formas de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos oligos de ARN que es una colección de oligos diseñados y sintetizados individualmente.

15 En todavía otra forma de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos en matriz de oligos de ADN (por ejemplo, un conjunto de señuelos en matriz disponible comercialmente) y, en otras formas de realización, los señuelos oligo individuales son oligos de ARN que se utilizan para complementar un conjunto de señuelos oligos de ADN que es una colección de oligos diseñados y sintetizados individualmente.

20 En una forma de realización, los señuelos oligo se diseñan para presentar como diana secuencias en genes de interés particular, tal como para conseguir una cobertura de secuenciación incrementada de conjuntos de genes expandidos.

25 En otra forma de realización, los señuelos oligo se diseñan para presentar como diana secuencias que representan un subconjunto del genoma y se mezclan y se utilizan como un reservorio en lugar de o adicionalmente a señuelos en matriz.

30 En una forma de realización, un primer conjunto de señuelos oligo se diseña para presentar como diana zonas de baja cobertura de secuenciación y un segundo conjunto de señuelos oligo se diseña para presentar como diana genes de interés particular. A continuación, ambos conjuntos de señuelos oligo se combinan y, opcionalmente, se mezclan con un conjunto estándar de señuelos en matriz para la utilización en la secuenciación.

35 En una forma de realización, se utiliza una mezcla de señuelos oligo, por ejemplo para secuenciar simultáneamente paneles génicos diana y para cribar un panel de polimorfismos de nucleótidos únicos (SNP) creados, tal como para el propósito de detectar reorganizaciones genómicas y alteraciones del número de copia (equivalente a la CGH (por sus siglas en inglés, hibridación genómica comparativa) en matriz) Por ejemplo, en primer lugar puede crearse un panel de SNP mediante el método de matrices en forma de señuelos en matriz, y después pueden diseñarse  
40 señuelos oligonucleótidos de ADN adicionales con diana en zonas de baja cobertura de secuenciación, con diana en un conjunto diana de genes. A continuación, puede repetirse la secuenciación de la colección de SNP con el conjunto original de señuelos en matriz más los señuelos oligo adicionales para conseguir la cobertura de secuenciación pretendida total.

45 En algunas formas de realización, se añaden señuelos oligo a un conjunto estándar de señuelos en matriz para conseguir una cobertura de secuenciación más completa. En una forma de realización, se diseñan señuelos oligo para reconocer zonas de baja cobertura de secuenciación después de una ronda inicial de secuenciación con un conjunto estándar de señuelos en matriz.

50 En otra forma de realización, se diseñan señuelos oligo con diana en secuencias de genes de interés particular. Dichos señuelos oligo pueden añadirse a un conjunto estándar de señuelos en matriz o a conjuntos existentes de híbridos de oligos/señuelos en matriz para conseguir, por ejemplo, una cobertura de secuenciación incrementada de conjuntos genes expandidos sin pasar por un ciclo entero de rediseño de reservorio de señuelos en matriz.

55 Pueden obtenerse señuelos oligo de un proveedor comercial, tal como NimbleGen (Roche) o Integrated DNA Technologies (IDT) para oligos de ADN. Los oligos también pueden obtenerse de Agilent Technologies. Los protocolos de enriquecimiento se encuentran disponibles públicamente, por ejemplo *SureSelect Target*.

#### *Sistema de enriquecimiento.*

60 Pueden producirse señuelos mediante métodos descritos en el documento nº US 2010/0029498 y Gnirke A. et al., Nat. Biotechnol. 27(2):182-1, 2009. Por ejemplo, pueden producirse señuelos de ARN biotinilado mediante la obtención de un reservorio de oligonucleótidos largos sintéticos, sintetizados originalmente en una micromatriz, y la amplificación de los oligonucleótidos para producir las secuencias de señuelo. En algunas formas de realización,  
65 los señuelos se producen mediante adición de una secuencia de promotor de ARN polimerasa en un extremo de las secuencias de señuelo y síntesis de las secuencias de ARN utilizando ARN polimerasa. En una forma de

realización, pueden obtenerse bibliotecas de oligodesoxinucleótidos sintéticos a partir de proveedores comerciales, tales como Agilent Technologies, Inc., y amplificarse utilizando métodos de amplificación de ácidos nucleicos conocidos.

5 Por ejemplo, puede generarse una gran colección de señuelos a partir de un reservorio personalizado de oligonucleótidos sintéticos sintetizado originalmente en una matriz de oligonucleótidos, por ejemplo una micromatriz de ADN programable de Agilent. De acuerdo con lo anterior, pueden sintetizarse simultáneamente por lo menos aproximadamente 2,500, 5,000, 10,000, 20,000, 3,000, 40,000, 50,000 o 60,000 oligonucleótidos únicos.

10 En una forma de realización, se selecciona un conjunto mínimo de oligonucleótidos únicos y se añaden copias adicionales (por ejemplo, alternantes entre los complementos inversos y las cadenas directas originales) hasta alcanzar la capacidad máxima de la matriz de oligonucleótidos sintéticos, por ejemplo para señuelos diseñados para capturar un conjunto preseleccionado de dianas (por ejemplo, un conjunto preseleccionado de exones). En otra forma de realización, la diana está representada por lo menos dos veces, por ejemplo mediante la síntesis de oligonucleótidos tanto diana como complementarios inversos. La síntesis de oligonucleótidos directos y complementarios inversos para una diana dada puede proporcionar una mejor redundancia en la etapa de síntesis que la síntesis de exactamente la misma secuencia dos veces. En todavía otra forma de realización, el producto de PCR o señuelo es el mismo para los oligonucleótidos directo y complementario inverso.

20 Los oligonucleótidos de los chips se sintetizan una vez y después pueden amplificarse para crear un conjunto de oligonucleótidos que puede utilizarse muchas veces. Este enfoque genera un reactivo universal que puede utilizarse como señuelo para un gran número de experimentos de selección, amortizando de esta manera el coste del chip para que resulte una fracción pequeña del coste de la secuenciación. Alternativamente, pueden producirse secuencias de señuelo utilizando métodos conocidos de amplificación de ácidos nucleicos, tales como PCR, por ejemplo, utilizando muestras de ADN humano o de ADN humano agrupado como el molde.

Tras la síntesis, los oligonucleótidos pueden liberarse (por ejemplo, arrancarse) de la matriz mediante escisión química seguido de la eliminación de los grupos de protección, y amplificarse por PCR en ADN de doble cadena utilizando cebadores universales. Puede utilizarse una segunda ronda de PCR para incorporar un sitio de promotor (por ejemplo, un promotor de T7, SP6 o de T3) en el amplicón, que se utiliza para transcribir el ADN en ARN de cadena sencilla.

35 En una forma de realización, los señuelos se embaldosan a lo largo de las secuencias (por ejemplo, exones) sin huecos o solapamientos. Por ejemplo, los señuelos pueden iniciarse en la base codificante más "a la izquierda" de la secuencia del genoma de referencia mostrada en el buscador de genoma UCSC (por ejemplo, 5' a 3' o 3' a 5' a lo largo de la secuencia codificante, dependiendo de la orientación del gen) y se añaden señuelos adicionales hasta cubrir todas las bases codificantes. En otra forma de realización, se diseñan por lo menos dos, tres, cuatro o cinco señuelos para cada diana, solapantes en por lo menos aproximadamente 15, 30, 45 o 60 bases. Tras la síntesis de oligonucleótidos y la amplificación por PCR utilizando cebadores universales, una de las colas del ADN de doble cadena puede monitorizarse enzimáticamente mediante la degradación de una de las cadenas. Los productos de cadena sencilla pueden hibridarse, convertirse en totalmente de doble cadena mediante rellenado y amplificarse mediante PCR. De esta manera, resulta posible producir señuelos que contienen por lo menos aproximadamente 300, 400, 500 o 600 bases específicas de diana contiguas, que son más de las que se pueden sintetizar químicamente. Tales señuelos largos pueden resultar útiles para aplicaciones que requieren una especificidad y sensibilidad elevadas, o para aplicaciones que no se benefician necesariamente de limitar la longitud de los señuelos (por ejemplo, la captura de regiones genómicas contiguas largas).

50 En una forma de realización, puede evaluarse la cobertura de cada diana y agruparse dianas que proporcionan una cobertura similar. Pueden crearse conjuntos diferentes de secuencias de señuelo para cada grupo de dianas, mejorando adicionalmente la representatividad. En otra forma de realización, se someten a ensayo oligonucleótidos de chips de micromatriz para eficacia de hibridación y ordenarse una ronda de producción de chips de micromatriz sobre la que se agrupan los oligonucleótidos según su eficacia de captura, compensando de esta manera para la variabilidad en la eficacia del señuelo. En todavía otra forma de realización, pueden agregarse reservorios de oligonucleótidos para formar un número relativamente pequeño de reservorios compuestos, de manera que haya poca variación de eficacia de captura entre ellos.

60 Los señuelos indicados en la presente memoria pueden marcarse con una etiqueta, por ejemplo una etiqueta de afinidad. Entre las etiquetas de afinidad ejemplificativas se incluyen, aunque sin limitación, moléculas de biotina, partículas magnéticas, haptenos u otras moléculas de etiqueta que permiten el aislamiento de los señuelos etiquetados con la molécula de etiqueta. Tales moléculas y métodos de unirlos a ácidos nucleicos (por ejemplo, los señuelos utilizados en los métodos dados a conocer en la presente memoria) son bien conocidos en la técnica. Se describen métodos ejemplificativos para producir señuelos biotinilados en, por ejemplo, Gnirke A. et al., Nat. Biotechnol. 27(2):182-9, 2009.

65 También se conocen en la técnica moléculas, partículas o dispositivos que se unen o son capaces de separar el conjunto de señuelos etiquetados respecto de la mezcla de hibridación. En una forma de realización, la molécula,

partícula o dispositivo se une a la etiqueta (por ejemplo, la etiqueta de afinidad). En una forma de realización, a molécula, partícula o dispositivo es una molécula de avidina, un imán o un anticuerpo o fragmento de unión a antígeno del mismo. En una forma de realización, los señuelos etiquetados se separan utilizando una perla magnética recubierta con moléculas de estreptavidina.

Se describen métodos ejemplificativos para preparar bibliotecas de oligonucleótidos en, por ejemplo, Gnirke A. et al., Nat. Biotechnol. 2009:182-9, 2, y Blumenstiel B. et al., Curr. Protoc. Hum. Genet. Capítulo 18: Unidad 18.4, 2010.

Los métodos y composiciones proporcionados en la invención implican el ajuste de la cobertura relativa de las secuencias de cada categoría de conjunto de señuelos/diana. Entre los métodos para implementar diferencias de cobertura relativa de secuencia en el diseño de señuelos se incluyen uno o más de:

- (i) representación diferencial de diferentes conjuntos de señuelos. El diseño del conjunto de señuelos para capturar una diana dada (por ejemplo, un miembro diana) puede incluirse en un número mayor/menor de copias para potenciar/reducir las profundidades relativas de cobertura de la diana,
- (ii) solapamiento diferencial de subconjuntos de señuelos. El diseño del conjunto de cebadores para capturar una diana dada (por ejemplo, un miembro diana) puede incluir un solapamiento más largo o más corto entre señuelos vecinos para potenciar/reducir las profundidades relativas de cobertura de la diana,
- (iii) parámetros diferenciales del sueño. El diseño del conjunto de cebadores para capturar una diana dada (por ejemplo, un miembro diana) puede incluir modificaciones de la secuencia/una longitud más corta a fin de reducir la eficiencia de captura y reducir las profundidades relativas de cobertura de la diana,
- (iv) mezcla de diferentes conjuntos de señuelos. Los conjuntos de señuelos que están diseñados para capturar diferentes conjuntos diana pueden mezclarse en diferentes proporciones molares para potenciar/reducir las profundidades relativas de cobertura de la diana,
- (v) utilizar diferentes tipos de conjuntos de señuelos oligonucleótidos. En determinadas formas de realización, el conjunto de señuelos puede incluir:
  - (a) uno o más señuelos sintetizados químicamente (por ejemplo, no enzimáticamente) (por ejemplo, sintetizados individualmente),
  - (b) uno o más señuelos sintetizados en una matriz,
  - (c) uno o más señuelos 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
  - (h) una combinación de cualquiera de los anteriores.

Las diferentes combinaciones de oligonucleótidos pueden mezclarse en diferentes proporciones, por ejemplo una proporción seleccionada de 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:50, 1:100, 1:1000 o similar. En una forma de realización, la proporción de señuelo sintetizado químicamente a señuelo generado en la matriz se selecciona de 1:5, 1:10 o 1:20. Los oligonucleótidos de ADN o ARN pueden ser naturales o no naturales. En determinadas formas de realización, los señuelos incluyen uno o más nucleótidos no naturales para, por ejemplo, incrementar la temperatura de fusión. Entre los oligonucleótidos no naturales ejemplificativos se incluyen nucleótidos de ADN o ARN modificado. Un nucleótido de ARN modificado ejemplificativo es un ácido nucleico bloqueado (ANB), en el que la fracción ribosa de un nucleótido 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., "Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes", Biochemistry 45 (23): 7347-55, 2006). Entre otros nucleótidos de ADN y ARN ejemplificativos modificados se incluyen, aunque sin limitarse a ellos, ácido péptido nucleico (APN) compuesto de unidades de N-(2-aminoetil)-glicina repetitivas unidades mediante enlaces peptídicos (Egholm, M. et al. Nature 365 (6446): 566-8, 1993); un oligonucleótido de ADN o ARN modificado para capturar regiones de bajo contenido de GC, un ácido nucleico bicíclico (ANBi) o un oligonucleótido entrecruzado, una 5-metil-desoxicitidina modificada y 2,6-diaminopurina. Se conocen en la técnica otros nucleótidos de ADN y ARN modificados.

En determinadas formas de realización, se obtiene una cobertura sustancialmente uniforme u homogénea de una secuencia diana (por ejemplo, un miembro diana). Por ejemplo, en cada categoría de conjunto de señuelos/diana, puede optimizarse la uniformidad de la cobertura mediante modificación de los parámetros de los señuelos, por ejemplo, mediante uno o más de:

- (i) puede utilizarse el incremento/reducción de la representación o solapamiento de los señuelos para potenciar/reducir la cobertura de las dianas (por ejemplo, miembros de diana) que se encuentra infra/sobre-cubiertos respecto a otras dianas en la misma categoría,
- (ii) para una cobertura baja, las secuencias diana difíciles de capturar (por ejemplo, secuencias de elevado contenido de GC), expanden la región que es la diana de los conjuntos de señuelos para cubrir, por ejemplo, secuencias contiguas (por ejemplo, secuencias contiguas menos ricas en GC),
- (iii) puede modificarse una secuencia de señuelo para reducir la estructura secundaria del señuelo y potenciar su eficiencia de selección,
- (iv) puede modificarse la longitud de un señuelo para ecualizar la cinética de la hibridación de fusión de los diferentes señuelos dentro de la misma categoría, La longitud del señuelo puede modificarse directamente (mediante producción de señuelos de longitudes diferentes) o indirectamente (mediante la producción de señuelos de longitud consistente y sustituyendo los extremos de señuelos por secuencia arbitraria),
- (v) la modificación de señuelos de diferente orientación para la misma región diana (es decir, cadena directa y cadena inversa) puede presentar diferentes eficiencias de unión Puede seleccionarse el conjunto de señuelos con cualquiera de las orientaciones que proporcione una cobertura óptima de cada diana,
- (vi) modificar la cantidad de una entidad de unión, por ejemplo, una etiqueta de captura (por ejemplo, biotina), presente en cada señuelo puede afectar a su eficiencia de unión. El incremento/reducción del nivel de etiqueta de los señuelos con diana en una diana específica puede utilizarse para potenciar/reducir la cobertura relativa de la diana.
- (vii) modificar el tipo de nucleótido utilizado para diferentes señuelos puede alterarse para afectar a la afinidad de unión a la diana y potenciar/reducir la cobertura relativa de la diana, o
- (viii) utilizar señuelos oligonucleótidos modificados, por ejemplo con un apareamiento de bases más estable, puede utilizarse para ecualizar la cinética de la hibridación de fusión entre zonas de contenido de GC bajo o normal respecto a las de contenido de GC elevado.

Por ejemplo, pueden utilizarse diferentes tipos de conjuntos de señuelos oligonucleótidos.

En una forma de realización, el valor de la eficiencia de selección se modifica mediante la utilización de diferentes tipos de señuelos oligonucleótidos para comprender regiones diana preseleccionadas. Por ejemplo, puede utilizarse un primer conjunto de señuelos (por ejemplo, un conjunto de señuelos basado en una matriz que comprende 10,000-50,000 señuelos de ARN o ADN) para cubrir una gran zona diana (por ejemplo una zona diana total de 1-2 MB). Al primer conjunto de señuelos puede añadirse un segundo conjunto de señuelos (por ejemplo, un conjunto de señuelos de ARN o ADN sintéticos que comprende menos de 5,000 señuelos) para cubrir una región diana preseleccionada (por ejemplo, intervalos subgenómicos de interés seleccionados que comprende, por ejemplo, 250 kb o menos de una zona diana) y/o regiones de estructura secundaria superior, por ejemplo de mayor contenido en GC. Los intervalos subgenómicos de interés seleccionados pueden corresponder a uno o más de los genes o productos génicos indicados en la presente memoria, o a un fragmento de los mismos. El segundo conjunto de señuelos puede incluir aproximadamente 2,000 a 5,000 señuelos según el solapamiento deseado de los señuelos. En todavía otras formas de realización, el segundo conjunto de señuelos puede incluir señuelos oligos seleccionados (por ejemplo, menos de 400, 200, 100, 50, 40, 30, 20 o 10 señuelo) añadidos al primer conjunto de señuelos. El segundo conjunto de señuelos puede mezclarse en cualquier proporción de señuelos oligo individuales. Por ejemplo, el segundo conjunto de señuelos puede incluir señuelos individuales presentes en una proporción equimolar 1:1. Alternativamente, el segundo conjunto de señuelos puede incluir señuelos individuales presentes en una proporción diferente (por ejemplo 1:5, 1:10 o 1:20), por ejemplo para optimizar la captura de determinadas dianas (por ejemplo, determinadas dianas pueden presentar 5-10x más del segundo señuelo que de otras dianas).

#### *Condiciones de hibridación*

Los métodos presentados en la invención incluyen la etapa de poner en contacto la biblioteca (por ejemplo, la biblioteca de ácidos nucleicos) con una pluralidad de señuelos para proporcionar una captura de biblioteca seleccionada. La etapa de puesta en contacto puede llevarse a cabo en la hibridación en solución. En determinadas formas de realización, el método incluye repetir la etapa de hibridación en una o más rondas adicionales de hibridación en solución. En algunas formas de realización, los métodos incluyen además someter la captura de

biblioteca a una o más rondas adicionales de hibridación en solución con la misma colección de señuelos o una diferente.

En otras formas de realización, los métodos presentados en la invención incluyen además la amplificación de la captura de biblioteca (por ejemplo, mediante PCR). En otras formas de realización, la captura de biblioteca no se amplifica.

En todavía otras formas de realización, los métodos incluyen además la etapa de someter la captura de biblioteca a genotipado, identificando de esta manera el genotipo de los ácidos nucleicos seleccionados.

Más específicamente, una mezcla de varios miles de secuencias de señuelos puede hibridarse eficazmente con ácidos nucleicos complementarios en un grupo de ácidos nucleicos y que dichos ácidos nucleicos hibridados (el subgrupo de ácidos nucleicos) puede separarse y recuperarse eficazmente. En una forma de realización, los métodos indicados en la presente memoria utilizando un conjunto de secuencias de señuelo que contienen más de aproximadamente 1,000 secuencias de señuelo, más de aproximadamente 2,000 secuencias de señuelo, más de aproximadamente 3,000 secuencias de señuelo, más de aproximadamente 4,000 secuencias de señuelo, más de aproximadamente 5,000 secuencias de señuelo, más de aproximadamente 6,000 secuencias de señuelo, más de aproximadamente 7,000 secuencias de señuelo, más de aproximadamente 8,000 secuencias de señuelo, más de aproximadamente 9,000 secuencias de señuelo, más de aproximadamente 10,000 secuencias de señuelo, más de aproximadamente 15,000 secuencias de señuelo, más de aproximadamente 20,000 secuencias de señuelo, más de aproximadamente 30,000 secuencias de señuelo, más de aproximadamente 40,000 secuencias de señuelo, o más de aproximadamente 50,000 secuencias de señuelo,

En algunas formas de realización, el procedimiento de selección se repite en el subgrupo seleccionado de ácidos nucleicos, por ejemplo con el fin de incrementar el enriquecimiento en ácidos nucleicos seleccionados. Por ejemplo, tras una ronda de hibridación, puede observarse un enriquecimiento de los ácidos nucleicos en un factor de varios miles. Tras una segunda ronda, el enriquecimiento puede elevarse, por ejemplo, hasta aproximadamente 15.000 veces de enriquecimiento medio, lo que puede proporcionar una cobertura cientos de veces de la diana en una única tanda del secuenciador. De esta manera, para experimentos que requieren factores de enriquecimiento que no pueden conseguirse en una única ronda de selección de híbridos, los métodos típicamente incluyen someter el subgrupo aislado de ácidos nucleicos (es decir, una parte o la totalidad de las secuencias diana) a una o más rondas adicionales de hibridación en solución con el conjunto de secuencias de señuelo.

La selección secuencial de híbridos con dos secuencias de señuelo diferentes (señuelo 1 y señuelo 2) puede utilizarse para aislar y secuenciar la "intersección", es decir, el subgrupo de secuencias de ADN que se une al señuelo 1 y al señuelo 2, que se utiliza para, por ejemplo, aplicaciones que incluyen, aunque sin limitación, el enriquecimiento para secuencias inter cromosómicas. Por ejemplo, la selección de ADN procedente de una muestra tumoral con un señuelo específico para secuencias en el cromosoma 1 seguido de la selección del producto de la primera selección de secuencias que se hibridan con un señuelo específico para el cromosoma 2, puede enriquecer en secuencias en uniones de traslocación cromosómica que contienen secuencias de ambos cromosomas.

La molaridad del subgrupo seleccionado de ácidos nucleicos puede controlarse de manera que la molaridad de cualquier ácido nucleico particular se encuentre dentro de una variación pequeña de la molaridad media de todos los ácidos nucleicos seleccionados en el subgrupo de ácidos nucleicos. Entre los métodos para controlar y optimizar la uniformidad de la representación de dianas se incluyen, aunque sin limitarse a ellos, el diseño tradicional de secuencias de señuelo basadas en reglas físicoquímicas, además de empíricas, de diseño de sondas bien conocidas en la técnica, y reservorios de señuelos en las que las secuencias que se conoce o se sospecha que presentan un rendimiento inferior al esperado están sobrerrepresentadas para compensar para sus debilidades intrínsecas. En algunas formas de realización, por lo menos aproximadamente 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90% o 95% del subgrupo aislado de ácidos nucleicos se encuentra dentro de aproximadamente 20 veces, 15 veces, 10 veces, 5 veces, 3 veces o 2 veces la molaridad media. En una forma de realización, por lo menos aproximadamente 50% del subgrupo aislado de ácidos nucleicos se encuentra dentro de aproximadamente 3 veces la molaridad media. En otra forma de realización, por lo menos aproximadamente 90% del subgrupo aislado de ácidos nucleicos se encuentra dentro de aproximadamente 10 veces la molaridad media.

Pueden ajustarse adicionalmente las variaciones en la eficiencia de selección mediante la alteración de la concentración de los señuelos. En una forma de realización, la eficiencia de selección se ajusta nivelando la eficiencia de los señuelos individuales en un grupo (por ejemplo, una primera, segunda o tercera pluralidad de señuelos), mediante el ajuste de la abundancia relativa de los señuelos, o la densidad de la entidad de unión (por ejemplo, la densidad de la etiqueta de hapteno o de afinidad) en referencia a la eficiencia diferencial de captura de secuencias observada al utilizar una mezcla equimolar de señuelos, seguido de la introducción de un exceso diferencial de grupo 1 internamente nivelado a la mezcla global de señuelos respecto al grupo 2 internamente nivelado.

En determinadas formas de realización, los métodos indicados en la presente memoria pueden conseguir una cobertura uniforme de las secuencias diana. En una forma de realización, el porcentaje de bases diana que

presenta por lo menos aproximadamente 50% de la cobertura esperada es de por lo menos aproximadamente 60%, 70%, 80% o 90%, por ejemplo para dianas cortas, tales como exones codificantes de proteína. En otra forma de realización, el porcentaje de bases diana que presenta por lo menos aproximadamente 50% de la cobertura esperada es de por lo menos aproximadamente 80%, 90% o 95%, por ejemplo para dianas que son excesivamente largas en comparación con la longitud de los señuelos de captura, tal como regiones genómicas.

Antes de la hibridación, los señuelos pueden desnaturalizarse según métodos bien conocidos en la técnica. En general, las etapas de hibridación comprenden añadir un exceso de ADN bloqueante a la composición de señuelo marcado, poner en contacto la composición de señuelo bloqueado bajo condiciones de hibridación con las secuencias diana que deben detectarse, eliminar mediante lavado los señuelos no hibridados y detectar la unión de la composición de señuelo a la diana.

Los señuelos se hibridan o aparean con las secuencias diana bajo condiciones hibridantes. "Condiciones hibridantes" son condiciones que facilitan la hibridación entre un señuelo y el ácido nucleico diana. Debido a que el apareamiento de diferentes señuelos variará dependiendo de la longitud de la sonda, la concentración de las bases y similares, el apareamiento se facilita modificando la concentración del señuelo, la temperatura de hibridación, la concentración salina y otros factores bien conocidos en la técnica.

Las condiciones de hibridación se facilitan mediante la modificación de las concentraciones, composiciones de las bases, complejidades y longitudes de los señuelos, así como concentraciones salinas, temperaturas y duración de la incubación. Por ejemplo, pueden llevarse a cabo en amortiguador de hibridación que contiene 5x SSPE, 5x solución de Denhardt, EDTA 5 mM y SDS al 0,1% y ADN bloqueante para suprimir la hibridación no específica. Pueden utilizarse inhibidores de ARNasa en el caso de que el señuelo sea de ARN. En general, las condiciones de hibridación, tal como se han indicado anteriormente, incluyen temperaturas de entre aproximadamente 25°C y aproximadamente 65°C, típicamente de aproximadamente 65°C y duraciones de incubación de entre aproximadamente 0,5 horas y aproximadamente 96 horas, típicamente de aproximadamente 66 horas. Se proporcionan condiciones de hibridación ejemplificativas adicionales en el Ejemplo 12A-12C y en la Tabla 14 en la presente memoria.

Los métodos indicados en la presente memoria son adaptables a métodos y dispositivos estándares de manipulación de líquidos. En algunas formas de realización, el método se lleva a cabo utilizando tecnología automatizada de manipulación de líquidos, tal como es conocido en la técnica, tal como dispositivos que manipulan placas multipocillo (ver, por ejemplo, Gnirke A. et al., Nat. Biotechnol. 27(2):182-189, 2009) Lo anterior puede incluir, aunque sin limitación, la construcción automatizada de bibliotecas y etapas de hibridación en solución, incluyendo la configuración y lavados posteriores a la hibridación en solución. Por ejemplo, puede utilizarse un aparato para llevar a cabo dichos métodos automatizados para las etapas de captura de perlas y de lavado después de la reacción de hibridación en solución. Entre los aparatos ejemplificativos pueden incluirse, aunque sin limitación, las posiciones siguientes: una posición para una placa multipocillo que contiene perlas magnéticas recubiertas con estreptavidina, una posición para la placa multipocillo que contiene las reacciones de selección de híbridos en solución, bloques térmicos de control I/O para precalentar reactivos y para llevar a cabo etapas de lavado a una temperatura definida por el usuario, una posición para una gradilla de puntas de pipeta, una posición con imanes dispuesta en determinadas configuraciones que facilitan la separación de sobrenadantes respecto de perlas inmovilizadas magnéticamente, una estación de lavado que lava las puntas de pipeta y desecha los residuos y posiciones para otras soluciones y reactivos, tales como amortiguadores de lavado de baja y alta astringencia o la solución para la elución alcalina de la captura final. En una forma de realización, el aparato está diseñado para procesar hasta 96 híbridos seleccionados de la etapa de captura de perlas hasta la etapa de neutralización de la captura en paralelo. En otra forma de realización, una o más posiciones presentan una función doble. En todavía otra forma de realización, el protocolo pide al usuario que intercambie una placa por otra.

El ácido nucleico seleccionado directamente puede concatenarse y romperse mecánicamente, lo que se lleva a cabo para superar las limitaciones de las lecturas de secuenciación cortas. En una forma de realización, cada diana de secuenciación de tamaño de exón se captura con una única molécula de señuelo que presenta aproximadamente el tamaño de la diana y presenta puntos terminales próximos a los puntos terminales de la diana. Sólo los híbridos que forman moléculas de doble cadena que presentan aproximadamente 100 o más pares de bases contiguas sobreviven a los lavados astringentes posteriores a la hibridación. Como resultado, el subgrupo seleccionado de ácidos nucleicos (es decir, la "captura") se enriquece en fragmentos de ADN genómico roto mecánicamente de manera aleatoria cuyos extremos se localizan próximos a los extremos de las moléculas de señuelo. La mera secuenciación terminal de la "captura" con lecturas de secuenciación muy cortas puede proporcionar una cobertura más elevada en proximidad al extremo (o incluso fuera) de la diana y reducir la cobertura en proximidad a la parte intermedia.

La concatenación de moléculas de "captura" mediante ligación y seguido de la rotura mecánica aleatoria y secuenciación aleatoria ("shotgun") es un método de obtener cobertura de secuencia a lo largo de la longitud entera de la secuencia diana. Este método produce un porcentaje más elevado de bases secuenciadas que están en la diana (en lugar de próximas a la diana) que la secuenciación terminal con lecturas muy cortas. Los métodos para concatenar moléculas mediante coligación son bien conocidas en la técnica. La concatenación puede llevarse a

cabo mediante simple ligación de extremos romos. Los extremos cohesivos para una ligación eficiente pueden producirse mediante una diversidad de métodos, incluyendo la amplificación por PCR de la "captura" con cebadores de PCR que presentan sitios de restricción próximos a sus extremos 5', seguido de digestión con el enzima de restricción correspondiente (por ejemplo, NotI) o mediante estrategias similares a las utilizadas comúnmente para la clonación independiente de ligación de productos de PCR, tales como la digestión parcial de una cadena ('chew-back') mediante ADN polimerasa de T4 (Aslanidis y de Jong, Nucleic Acids Res. 18:6069-6074, 1990) o el tratamiento de productos de PCR que contienen uracilo utilizando UDG glucosilasa y liasa endo VIII (por ejemplo, New England Biolabs, n° de cat. E5500S).

En otra forma de realización, se utiliza un conjunto escalonado de moléculas de señuelo para reconocer una región, obteniendo extremos de señuelo frecuentes en toda la región diana. En algunas formas de realización, la mera "captura" de secuenciación terminal (es decir, sin concatenación y rotura mecánica) proporciona una cobertura de secuencia bastante uniforme a lo largo de la región entera que está cubierta por el señuelo, incluyendo la diana de secuenciación misma (por ejemplo, un exón). Debido a que el escalonado de las moléculas de señuelo amplía el segmento cubierto por el señuelo, las bases secuenciadas se distribuyen en una zona más amplia. En consecuencia, la proporción de secuencias en la diana respecto a las secuencias próximas a la diana es más baja que para selecciones con señuelos no solapantes, los cuales con frecuencia requieren únicamente un solo señuelo por diana.

En otra forma de realización, la secuenciación terminal con lecturas ligeramente más largas (por ejemplo, de 76 bases) es el método típico para secuenciar dianas seleccionadas cortas (por ejemplo, exones). Al contrario que la secuenciación terminal con lecturas muy cortas, este método conduce a un perfil de cobertura unimodal sin una bajada de cobertura en la parte intermedia. Este método resulta más fácil de llevar a cabo que el método de concatenación y rotura mecánica indicado anteriormente, resulta en una cobertura relativamente uniforme a lo largo de las dianas y genera un porcentaje elevado de bases secuenciadas que se encuentran en el señuelo y en la diana misma.

En una forma de realización, el subgrupo seleccionado de ácidos nucleicos se amplifica (por ejemplo, mediante PCR) antes del análisis mediante secuenciación o genotipado. En otra forma de realización, el subgrupo se analiza sin una etapa de amplificación, por ejemplo al analizar el subgrupo seleccionado mediante métodos analíticos sensibles que pueden leer moléculas individuales.

### Secuenciación

La invención incluye además métodos de secuenciación de ácidos nucleicos. En estos métodos, se aíslan miembros de una biblioteca de ácidos nucleicos mediante la utilización de los métodos indicados en la presente memoria, por ejemplo utilizando la hibridación en solución, proporcionando de esta manera una captura de biblioteca. La captura de biblioteca o un subgrupo de la misma puede secuenciarse. De acuerdo con lo anterior, los métodos presentados en la invención incluyen además analizar la captura de biblioteca. En una forma de realización, la captura de biblioteca se analiza mediante un método de secuenciación, por ejemplo un método de secuenciación de nueva generación tal como se indica en la presente memoria. Los métodos incluyen aislar una captura de biblioteca mediante hibridación en solución y someter la captura de biblioteca a secuencia de ácidos nucleicos. En determinadas formas de realización, la captura de biblioteca puede resecuenciarse.

Puede utilizarse cualquier método de secuenciación conocido en la técnica. La secuenciación de ácidos nucleicos aislados mediante métodos de selección se lleva a cabo típicamente utilizando secuenciación de nueva generación (NGS). La secuenciación de nueva generación incluye cualquier método de secuenciación que determina la secuencia de nucleótidos de moléculas de ácidos nucleicos individuales o sustitutos expandidos clonalmente de moléculas individuales de ácidos nucleicos de una manera altamente paralela (por ejemplo, se secuencian simultáneamente más de  $10^5$  moléculas). En una forma de realización, la abundancia relativa de las especies de ácidos nucleicos en la biblioteca puede estimarse mediante el recuento del número relativo de incidencias de sus secuencias afines en los datos generados mediante el experimento de secuenciación. Los métodos de secuenciación de nueva generación son conocidos en la técnica y se describen en, por ejemplo, Metzker, M., Nature Biotechnology Reviews 11:31-46, 2010, incorporada en la presente memoria como referencia.

En una forma de realización, la secuenciación de nueva generación permite la determinación de la secuencia de nucleótidos de una molécula individual de ácidos nucleicos (por ejemplo, el sistema de secuenciación génica HeliScope de Helicos BioSciences y el sistema PacBio RS de Pacific Biosciences). En otras formas de realización, el método de secuenciación determina la secuencia de nucleótidos de sustitutos clonalmente expandidos de moléculas individuales de ácidos nucleicos (por ejemplo, el secuenciador Solexa, Illumina Inc., San Diego, Calif.; 454 Life Sciences (Branford, Conn.) e Ion Torrent). Por ejemplo, la secuenciación de lecturas cortas en paralelo masiva (por ejemplo, el secuenciador Solexa, Illumina Inc., San Diego, Calif.), que genera más bases de secuencia por cada unidad de secuenciación que otros métodos de secuenciación que generan menos lecturas, aunque más largas. Entre otros métodos o aparatos para la secuenciación de nueva generación se incluyen, aunque sin limitarse a ellos, los secuenciadores proporcionados por 454 Life Sciences (Branford, Conn.), Applied Biosystems

(Foster City, Calif.; secuenciador SOLiD), Helicos BioSciences Corporation (Cambridge, Mass.) y nanogotas de tecnología de secuenciación en emulsión y de microfluidos (por ejemplo, las gotas GnuBio).

Entre las plataformas de secuenciación de nueva generación se incluyen, aunque sin limitación, el secuenciador genómico (GS) de Roche/454, el analizador genómico (GA) de Illumina/Solexa, el sistema de detección de ligación de oligonucleótidos de soporte de Life/APG (SOLiD), el sistema G.007 de Polonator, sistema de secuenciación génica HeliScope de Helicos BioSciences y el sistema PacBio RS de Pacific Biosciences.

Las tecnologías de NGS pueden incluir una o más etapas, por ejemplo preparación de moldes, secuenciación y obtención de imágenes, y análisis de datos.

Se conocen en la técnica metodologías de secuenciación ejemplificativas adicionales, por ejemplo algunas de las cuales se describen en los documentos de patente de propiedad común USSN nº 13/339,986 y PCT nº US11/67725, ambos presentados el 29 de diciembre de 2011, que se han publicado como el documento nº WO2012/092426.

### Alineación

La alineación es el procedimiento de hacer corresponder una lectura con una localización, por ejemplo una localización genómica. La alineación defectuosa (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, la presencia de una secuencia repetitiva) de lecturas en torno a una mutación de un cáncer renal real puede conducir a la reducción de la sensibilidad de detección de la mutación, ya que las lecturas del alelo alternativo puede encontrarse desplazadas fuera de la acumulación principal de lecturas de alelo alternativas. En el caso de que el contexto de la secuencia problemática se encuentre en donde no existe ninguna mutación, la alineación defectuosa puede introducir lecturas artefactuales de alelos "mutados" mediante la introducción de lecturas reales de bases del genoma de referencia en una localización errónea. Debido a que los algoritmos de llamada de mutaciones para el análisis multigénico multiplicado deben ser sensibles a incluso mutaciones de baja abundancia, dichas alineaciones defectuosas pueden incrementar las tasas de descubrimiento falso positivo/reducir la especificidad.

Tal como se expone en la presente memoria, la sensibilidad reducida para las mutaciones reales puede controlarse mediante la evaluación de la calidad de las alineaciones (manualmente o de una manera automatizada) en torno a los sitios de mutación esperados en los genes que se analizan. Los sitios que deben evaluarse pueden obtenerse de bases de datos de mutaciones de cáncer (por ejemplo, COSMIC). Las regiones que se identifican como problemáticas pueden remediarse utilizando un algoritmo seleccionado para proporcionar un mejor rendimiento en el contexto de secuencia relevante, por ejemplo mediante optimización de alineaciones (o realineación) utilizando algoritmos de alineación más lentos aunque más exactos, tales como la alineación de Smith-Waterman. En los casos en que los algoritmos de alineación general no pueden remediar el problema, pueden crearse enfoques de alineación personalizada mediante, por ejemplo: el ajuste de los parámetros de penalización por apareamiento incorrecto de diferencia máxima para genes con una probabilidad elevada de contener sustituciones; el ajuste de los parámetros de penalización por apareamiento incorrecto específico basados en tipos de mutación específicos que son comunes en determinados tipos tumorales (por ejemplo, C→T en el melanoma); o el ajuste de los parámetros de penalización por apareamiento incorrecto específico basado en tipos de mutación específicos que son comunes en determinados tipos de muestra (por ejemplo, las sustituciones que son comunes en las muestras FFPE).

La especificidad reducida (tasa incrementada de falsos positivos) en las regiones génicas evaluadas debido a la alineación incorrecta puede evaluarse mediante examen manual o automatizado de todas las llamadas de mutación en las muestras secuenciadas. Aquellas regiones que se observe que presentan mayor tendencia a llamadas de mutación espuria debido a alineación incorrecta pueden someterse a los remedios de alineación que se han indicado anteriormente. En los casos en que se encuentre que ningún remedio algorítmico resulta posible, las "mutaciones" de las regiones problema pueden clasificarse o excluirse del panel de ensayo.

### Inserciones/Deleciones (indels)

Generalmente, la detección exacta de mutaciones indel es un ejercicio de alineación, ya que la tasa de indel espuria en las plataformas de secuenciación en la presente memoria es relativamente baja (de esta manera, hasta un puñado de observaciones de indel correctamente alineadas puede constituir una prueba sólida de mutación). Sin embargo, la alineación exacta en presencia de indels puede resultar difícil (especialmente a medida que se incrementa la longitud del indel). Además de los problemas generales asociados a la alineación, por ejemplo de sustituciones, el indel mismo puede causar problemas con la alineación. (Por ejemplo, una deleción de 2 pb de una repetición dinucleótida no puede localizarse definitivamente con facilidad). Tanto la sensibilidad como la especificidad pueden reducirse mediante la colocación incorrecta de lecturas que aparentemente contienen indel más cortas (<15 pb). Los indels más grandes (de magnitud que se aproxima a la longitud de las lecturas

individuales: 36 pb en el presente procedimiento) pueden hacer fracasar la alineación de la lectura, imposibilitando la detección del indel en el conjunto estándar de lecturas alineadas.

Las bases de datos de mutaciones de cáncer pueden utilizarse para resolver dichos problemas y mejorar los rendimientos. Para reducir el descubrimiento de falsos positivos de indel (para mejorar la especificidad), pueden examinarse regiones en torno a los indel esperados comúnmente, para alineaciones problemáticas debido al contexto de la secuencia y resolverse de manera similar a las sustituciones, anteriormente. Para mejorar la sensibilidad de la detección de indel, pueden utilizarse varios enfoques diferentes de utilización de información sobre los indel esperados en el cáncer. Por ejemplo, pueden simularse lecturas cortas que contienen indel esperados e intentarse la alineación. Las alineaciones pueden estudiarse y ajustarse los parámetros de alineación de las regiones de indel problemáticas, por ejemplo mediante la reducción de las penalizaciones por apertura/extensión de hueco o mediante alineación de lecturas parciales (por ejemplo, la primera o segunda mitad de una lectura).

Alternativamente, puede intentarse la alineación inicial no sólo con el genoma de referencia normal, sino también con versiones alternativas del genoma, que contienen cada una de las mutaciones indel de cáncer conocidas o probables. En este enfoque, las lecturas de indel que inicialmente no pudieron alinearse o se alinearon incorrectamente se colocan con éxito en la versión alternativa (mutada) del genoma.

Se conocen en la técnica metodologías de secuenciación ejemplificativas adicionales, por ejemplo algunas de las cuales se describen en los documentos de patente de propiedad común USSN nº 13/339,986 y PCT nº US11/67725, ambos presentados el 29 de diciembre de 2011.

#### Llamada de mutaciones

La llamada de bases se refiere a la salida en bruto obtenida en 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 nucleótida que se está secuenciando. Típicamente, las lecturas de secuenciación (o llamadas de base) para una posición proporcionarán más de un valor, por ejemplo algunas lecturas proporcionarán una T y algunas proporcionarán una G. La llamada de mutaciones es el procedimiento de asignación de un valor de nucleótido, por ejemplo, uno de dichos valores, a la secuencia. Aunque se denomina llamada "de mutaciones", puede aplicarse a la asignación de un valor de nucleótido a cualquier posición nucleótida, por ejemplo posiciones correspondientes a alelos mutantes, alelos de tipo salvaje, alelos que no se han caracterizado como mutantes o de tipo salvaje, o a posiciones no caracterizadas según la variabilidad. Los métodos para la llamada de mutaciones pueden incluir uno o más de los siguientes: la realización de llamadas independientes basadas en la información de cada posición en la secuencia de referencia (por ejemplo, el examen de las lecturas de secuencia; el examen de las llamadas de base y las puntuaciones de calidad; el cálculo de la probabilidad de las bases observadas y las puntuaciones de calidad proporcionadas a un genotipo potencial; y la asignación de genotipos (por ejemplo, utilizando la regla de Bayes)); la eliminación de los falsos positivos (por ejemplo, utilizando umbrales de profundidad para rechazar SNP con una profundidad de lectura muy inferior o superior a la esperada; la realineación local para eliminar falsos positivos debido a indels pequeños) y la realización de análisis basado en desequilibrios de ligamiento (LD)/imputación para afinar las llamadas.

Se indican ecuaciones para calcular la probabilidad de un genotipo asociada a un genotipo y posición específicos en, por ejemplo Li H. y Durbin R. *Bioinformatics*; 26(5): 589-95, 2010. La expectativa previa para una mutación particular en un tipo determinado de cáncer puede utilizarse al evaluar muestras de ese tipo de cáncer. Dicha probabilidad puede derivarse de bases de datos públicas de mutaciones de cáncer, por ejemplo, el Catálogo de mutaciones somáticas en cáncer (COSMIC), la Base de datos de mutaciones genéticas humanas (HGMD), la base de datos de mutaciones de cáncer de mama (BIC) del Consorcio SNP y la base de datos de genes del cáncer de mama (BCGD).

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

#### Llamada de mutaciones: sustituciones

Tras la alineación, puede llevarse a cabo la detección de sustituciones utilizando un método de llamada, por ejemplo el método de llamada de mutaciones bayesiano, que se aplica a cada base en cada uno de los intervalos subgenómicos, por ejemplo exones del gen que debe evaluarse, en donde se observa la presencia de alelos alternativos. Dicho método compara la probabilidad de observar los datos de lecturas en presencia de una mutación con la probabilidad de observar los datos de lecturas en presencia de solo error de llamada de bases. Las mutaciones pueden llamarse en el caso de que dicha comparación respalde con suficiente fuerza la presencia de una mutación.

Los métodos se han desarrollado para incluir desviaciones limitadas respecto a frecuencias de 50% o 100% para el análisis de ADN de cáncer (por ejemplo, SNVMix -Bioinformatics. 15 de marzo 2010; 26(6): 730-736). Los métodos dados a conocer en la presente memoria, sin embargo, permiten considerar la posibilidad de la presencia de un alelo mutante en cualquier sitio entre 1% y 100% del ADN de la muestra, y especialmente a niveles inferiores a 50%. Este enfoque resulta particularmente importante para la detección de mutaciones en muestras FFPE de baja pureza de ADN tumoral natural (multiclonal).

Una ventaja del enfoque de detección de mutaciones bayesiano es que la comparación de la probabilidad de presencia de una mutación con la probabilidad de error en la llamada de bases únicamente puede ponderarse con la expectativa previa de presencia de una mutación en el sitio. En el caso de que algunas lecturas de un alelo alternativo se observen en un sitio frecuentemente mutado para el tipo de cáncer dado, la presencia de una mutación puede llamarse fiablemente, aunque la cantidad de pruebas de mutación no satisfaga los umbrales habituales. Esta flexibilidad puede utilizarse en este caso para incrementar la sensibilidad de detección para mutaciones incluso más raras/muestras de pureza más baja, o para que la prueba resulte más robusta frente a reducciones de la cobertura de lecturas. La probabilidad de que un par de bases aleatorio en el genoma esté mutado en el cáncer es  $\sim 10^{-6}$ . La probabilidad de mutaciones específicas en muchos sitios en un panel genómico de cáncer multigénico típico puede ser órdenes de magnitud más elevada. Estas probabilidades pueden obtenerse de bases de datos públicas de mutaciones de cáncer (por ejemplo, COSMIC).

#### Llamada de mutaciones: indels

La llamada de indel es un procedimiento para encontrar bases en los datos de secuenciación que difieren respecto de la secuencia de referencia por inserción o delección, incluyendo típicamente una puntuación de confianza o métrica de evidencia estadística asociada.

Entre los métodos de llamada de indel pueden incluirse las etapas de identificar un indel candidato, calcular la probabilidad de un genotipo mediante realineación local y la realización de inferencia y llamada de genotipos basada en el LD. Típicamente, se utiliza un enfoque bayesiano para obtener potenciales candidatos de indel y después se someten a ensayo estos candidatos junto con la secuencia de referencia en un marco bayesiano.

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

Entre los métodos para generar llamadas de indel y probabilidades de genotipo a nivel individual se incluyen, por ejemplo, el algoritmo Dindel (Albers C.A. et al., Genome Res. 27 de oct., 2010. [pub. elec. anterior a la impresión]). Por ejemplo, puede utilizarse el algoritmo EM bayesiano para analizar las lecturas, realizar llamadas de indel iniciales, y generar probabilidades de genotipo para cada indel candidato, seguido de la imputación de genotipos utilizando, por ejemplo, QCALL (Le S.Q. y Durbin R., Genome Res. 27 de oct., 2010. [pub. elec. anterior a la impresión]). Pueden ajustarse los parámetros, tales como las expectativas previas de observar el indel (por ejemplo, incrementarse o reducirse), basándose en el tamaño o localización de los indel.

Se conocen en la técnica metodologías de secuenciación ejemplificativas adicionales, por ejemplo algunas de las cuales se describen en los documentos de patente de propiedad común USSN nº 13/339,986 y PCT nº US11/67725, ambos presentados el 29 de diciembre de 2011, que se han publicado como el documento nº WO2012/092426.

#### **Ejemplos**

La presente invención se describe adicionalmente haciendo referencia a los ejemplos a continuación, que son proporcionados a título ilustrativo y no limitativo de la invención. Pueden utilizarse técnicas estándares bien conocidas en la técnica o técnicas específicamente descritas a continuación.

#### **Ejemplo 1. Hibridación de sonda de ADN con los productos de captura**

El procedimiento a continuación resume las etapas necesarias para la hibridación de la sonda de ADN con los productos de captura.

##### A. Hibridación

Se combinaron cien nanogramos de señuelos biotinilados agrupados, 500 ng de biblioteca de ADN adaptada, 2 µg de ADN C<sub>6</sub>t-1, 2 nanomoles de bloqueantes oligonucleótidos en 2,0 µl, en un volumen de 10 µl y se mezclaron con 10 µl de amortiguador Genisphere precalentado (amortiguador de hibridación a base de SDS 2x: NaPO<sub>4</sub> 0,50 M, SDS al 1%, EDTA 2 mM, 2x SSC, 4x solución de Denhardt). Tras la mezcla con vórtex de la mezcla, se aplicó una capa superior de 40 µl de aceite mineral y la mezcla se desnaturalizó en un termociclador a 95°C durante 5 minutos con una lenta reducción hasta 71°C. La mezcla se incubó a 71°C durante 48 horas.

**B. Unión a perlas con estreptavidina**

Las perlas con estreptavidina se prepararon de la manera siguiente, antes de añadirlas a la mezcla de hibridación.

5 Las perlas con estreptavidina se dejaron en reposo a temperatura ambiente durante 30 minutos. Para cada reacción de hibridación, se lavaron 50 µl de perlas con estreptavidina Invitrogen M270 (magnéticas) dos veces con 2x amortiguador de unión y lavado (Tris-HCl 10 mM (pH 7,5), NaCl 2 M y EDTA 1 mM). Las perlas se resuspendieron en 80 µl que incluían 50 µl de amortiguador de unión y lavado y 30 µl de agua.

10 Al final del periodo de hibridación de 48 horas, se extrajeron los 20 µl de líquido de hibridación de bajo el aceite mineral añadido a los 80 µl de perlas, proporcionando un volumen total de 100 µl. La mezcla se centrifugó en un rotador de tubos durante 30 minutos para permitir que se produjese la unión entre la biotina en los complejos hibridados de molde: señoelo y la estreptavidina sobre las perlas.

**C. Lavado de las perlas con estreptavidina**

Tras el periodo de rotación, las muestras se introdujeron en una gradilla de separación magnética. Se dejó que las perlas se separasen del sobrenadante, y el sobrenadante que contenía ADN que no se unía a las sondas de captura se separó y se desechó. Las perlas unidas a sonda se lavaron secuencialmente con las soluciones siguientes. Para cada lavado, se añadió la solución de lavado que había sido preequilibrada hasta la temperatura dada, se introdujo en un rotador durante el tiempo indicado, se sedimentó brevemente (imán) y se recolectó el sobrenadante y se desechó. El primer lavado fue con 1000 µl 1x SSC/SDS al 0,1% durante 5 minutos, a 71°C con rotación. El segundo lavado fue con 1000 µl 1x SSC/SDS al 0,1% durante 5 minutos, a 71°C con rotación. El tercer lavado fue con 1000 µl 1x SSC/SDS al 0,1% durante 5 minutos, a 71°C con rotación. El cuarto lavado fue con 1000 µl 1x SSC/SDS al 0,1% durante 5 minutos, a temperatura ambiente (TA) con rotación. El quinto lavado fue con 1000 µl 0,2 x SSC durante 30 minutos, a TA con el tubo todavía en el imán. La solución de lavado final se eliminó por completo antes del procesamiento posterior, tal como se explica después.

Después del lavado final, se añadieron 50 µl de NaOH 0,125 N y la mezcla se incubó a TA durante 10 minutos, con agitación con vórtex cada 2 minutos para mantener las perlas en solución. El tubo con las perlas se introdujo nuevamente en el imán durante 1 minuto. Mientras las perlas se encontraban sobre el imán, se añadió una alícuota de 50 µl de Tris-HCl 1 M (pH 8,8) a un nuevo tubo de 1,5 ml de PCR libre de ARNasa/ADNasa. El sobrenadante del tubo en el imán (NaOH 0,125 N) se añadió al tubo que contenía Tris-HCl 1 M (pH 8,8) para neutralizar la solución. Los fragmentos de molde recuperados se purificaron con perlas AMPure utilizando 1,5 x volumen y eluyendo en 20 µl de amortiguador EB (Tris-Cl 10 mM, pH 8,5).

**Ejemplo 2. Reacciones de PCR con material de molde de cadena sencilla****A. Enriquecimiento de PCR final**

Los moldes de cadena sencilla recuperados (16 µl) se prepararon en un volumen total de 50 µl con los componentes siguientes de la mezcla de reacción (mezcla maestra KAPA HiFi (25 µl): cebador 1 10 µM (2,5 µl), cebador 2 10 µM (2,5 µl), agua (4 µl)). El ADN se agitó con vórtex brevemente y se recolectó en forma de una solución después de una centrifugación breve. Las reacciones se introdujeron en un termociclador con el programa siguiente: 98°C (5.0 min); 98° C (20 s); 60° C (15 s); 72° C (20 s); 77° C (5.0 min) durante cinco o más ciclos. Los productos amplificados se purificaron con perlas AMPure utilizando 1,5 x volumen y eluyendo en 20 µl de amortiguador EB (Qiagen) (Tris-Cl 10 mM, pH 8,5). La concentración resultante del ADN se midió con un fluorímetro Qubit y se diluyó para la utilización con la plataforma de secuenciación de NGS apropiada.

Se utilizaron cinco ciclos de amplificación para las bibliotecas Ion Torrent posteriores a la captura y típicamente se utilizaron no más de 18 ciclos de amplificación para las bibliotecas de Illumina posteriores a la captura. El protocolo estándar de Illumina se optimizó utilizando el procedimiento de PCR siguiente. Los moldes de ADN de cadena sencilla recuperados (2 ml) se combinaron en un volumen final de 50 µl que incluía 25 µl de mezcla maestra SyberGreen, 8 pmoles de cebador 1 y 8 pmoles de cebador 2. Las reacciones se prepararon en una placa de qPCR de 96 pocillos para imitar el enriquecimiento de PCR final y se implementó el programa siguiente: 95° C (5 min) seguido de 30 ciclos de 95° C (30 s) y 60° C (45 s). Se ajustó manualmente el umbral para encontrar el punto medio de la curva (a mitad de camino entre donde se inicia la amplificación y el nivel de meseta) y se restaron 3 ciclos de dicho valor para determinar el número de ciclos que debían ejecutarse para el enriquecimiento de PCR final. Se restaron tres ciclos debido a que la cantidad de ADN introducida en la optimización es 8x inferior a la que se introducirá en la reacción de enriquecimiento final; 2 µl del producto capturado neutralizado entra en la reacción de optimización de PCR y 16 µl entran en el enriquecimiento de PCR final.

### Ejemplo 3. Oligonucleótidos aumentados en $T_m$ para la utilización en la plataforma de secuenciación de Illumina con bases de inosina para dominios de código de barras

En la Tabla I, se diseñaron los oligonucleótidos bloqueantes siguientes para la utilización en experimentos de captura de híbridos para bibliotecas de moldes de ADN con la plataforma de secuenciación de Illumina. Los oligonucleótidos aumentados en  $T_m$  se prepararon utilizando ANB ("C" o "+A") o ANBi ("iANBi-meC/" o "iANBi-A") como grupos que aumentan  $T_m$ . Todos los oligonucleótidos se prepararon utilizando métodos químicos de fosforamida. Los valores de  $T_m$  se estiman para bases de ANB en amortiguador de NaCl 750 mM (fuerza iónica similar a 5x SSC) y para amortiguador de NaCl 15 mM (fuerza iónica similar a 0,1x SSC) utilizando el método de Owczarzy (Biochemistry 2011 50:9352-9367), que se incorpora como referencia en su totalidad. La modificación de ANBi presenta efectos termodinámicos similares a la modificación de ANB, por lo que las predicciones presentadas en la presente memoria se aplican a ambas clases de bloqueantes modificados y las modificaciones de ANB/ANBi pueden sustituirse en todos los ejemplos. Por ejemplo, el modelado termodinámico en los ejemplos, posteriormente, se llevó a cabo utilizando parámetros de vecino más próximo derivado de ANB, mientras que la síntesis de oligonucleótidos se llevó a cabo utilizando bases de ANBi.

Tabla I. Oligonucleótidos bloqueantes aumentados en  $T_m$

SEC ID nº	Secuencia	nº ANB	$T_m$ (°C) 750 mM (Na)	$T_m$ (°C) 15 mM (Na)
1	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT	0	90,3	62,0
2	AGATCGGAAGAGCGT+CGTGTAGGGAAAGAGTGTAGATCT+CGGTGGTCGCCGTATCATT	2	92,0	63,4
3	AGAT+CGGAAGAGCGT+CGTGTAGGGAAAGAGTGTAGATCT+CGGTGGTCGC+CGTATCATT	4	93,9	65,0
4	AGAT+CGGAAGAGCGT+CGTGTAGGGAAAGAGTGTAGATCT+CGGTGGT+CGC+CGTAT+CATT	6	>95	66,6
5	AGAT+CGGAAGAG+CGT+CGTGTAGGGAAAGAGTGTAGAT+CT+CGGTGGT+CGC+CGTAT+CATT	8	>95	68,3
6	AGAT+CGGA+AGAG+CGT+CGTGT+AGGG+AAAGAGTGTAGAT+CT+CGGTGGT+CG+C+CGTAT+CATT	12	>95	71,1
7	AG+AT+CGGA+AGAG+CGT+CGTGT+AGGG+A+A+AGAGTGTAG+AT+CT+CGGTGGT+CG+C+CGTAT+CATT	16	>95	73,3
8	AG+AT+CGG+A+AG+AG+CGT+CGTGT+AGGG+A+A+AG+AGTGT+AG+AT+CT+CGGTGGT+CG+C+CGT+AT+C+ATT	22	>95	77,4

En la Tabla I, los grupos que aumentan  $T_m$  ANB-C se incluyen inicialmente en los oligonucleótidos aumentados en  $T_m$  hasta que todas las posiciones de C (es decir, 9 posiciones que presentan C) han sido exhaustivamente sustituidas, tras la inclusión de grupos que aumentan  $T_m$  ANB-A en las posiciones de A, posteriormente.

En la Tabla II, posteriormente, se presenta el diseño de una serie de oligonucleótidos aumentados en  $T_m$  para la utilización como bloqueante contra el adaptador que contiene la secuencia de código de barras (8 inosinas). Tal como se explica en la descripción detallada, no existe ningún modo de modelar el valor de  $T_m$  aumentado con inosinas (definida en las secuencias de la Tabla II como "I" y "ideoxil/") apareado con bases variables. Por lo tanto, las bases de inosina no se incluyeron en el análisis de la  $T_m$ , aunque se encuentran presentes en la secuencia final. Sin embargo, el valor de  $T_m$  aumentado preciso para las secuencias reales se determina fácilmente mediante métodos empíricos rutinarios. Los oligonucleótidos aumentados en  $T_m$  se prepararon utilizando ANB ("C" o "+A") o ANBi ("iANBi-meC/" o "iANBi-A") como grupos que aumentan  $T_m$ . Todos los oligonucleótidos se prepararon utilizando métodos químicos de fosforamida.

Tabla II. Bloqueantes oligonucleótidos aumentados en  $T_m$  con secuencias de código de barras

SEC ID nº	Secuencia	nº ANB	$T_m$ (°C) 750 mM (Na)	$T_m$ (°C) 15 mM (Na)
9	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (III IIIII) ATCTCGTATGCCGTCTTCTGCTTG	0	89,8	61,7
10	GATCGGAAGAGCACACGTCTGAACTCCAGT+CAC (II IIIII) ATCT+CGTATGCCGTCTTCTGCTTG	2	91,5	63,1
11	GATCGGAAGAGCACACGTCTGAA+CTCCAGT+CAC (IIIIIIII) ATCT+CGTATGC+CGTCTTCTGCTTG	4	93,9	65,2

SEC ID nº	Secuencia	nº ANB	T <sub>m</sub> (°C) 750 mM (Na)	T <sub>m</sub> (°C) 15 mM (Na)
12	GATCGGAAGAG+CACACGTCTGAA+CTCCAGT+CAC (IIIIIIII) ATCT+CGTATGC+CGTCTTCTG+CTTG	6	>95	66,8
13	GAT+CGGAAGAG+CACACGTCTGAA+CTCCAGT+CA+ C (IIIIIIII) ATCT+CGTATGC+CGTCTTCTG+CTTG	8	>95	68,7
14	GAT+CGGAAGAG+CACA+CGT+CTGAA+CTC+CAGT+ CAC (IIIIIIII) ATCT+CGTATGC+CGT+CTT+CTG +CTTG	12	>95	72,7
15	GAT+CGGAAGAG+CA+CA+CGT+CTGAA+CT+C+CAG T+CA+C (IIIIIIII) AT+CT+CGTATG+C+CGT+CT T+CTG+CTTG	17	>95	77,4
16	GAT+CGG+A+AGAG+CA+CA+CGT+CTG+AA+CT+C+ CAGT+CA+C (IIIIIIII) +AT+CT+CGT+ATG+C+C GT+CTT+CTG+CTTG	22	>95	80,4

En la Tabla II, los grupos que aumentan T<sub>m</sub> ANB-C se incluyen inicialmente en los oligonucleótidos aumentados en T<sub>m</sub> hasta que todas las posiciones de C (es decir, 17 posiciones que presentan C) han sido exhaustivamente sustituidas, tras la inclusión de grupos que aumentan T<sub>m</sub> ANB-A en las posiciones de A, posteriormente. En el presente ejemplo, se incorporan bases de inosina para cubrir el dominio de código de barras. Una mezcla aleatoria de bases "N" de la totalidad de las 4 nucleobases u otra base universal, tal como 5-nitroindol, podría utilizarse tal como se ha indicado anteriormente en la especificación.

#### Ejemplo 4. Oligonucleótidos bloqueantes aumentados en T<sub>m</sub> para la utilización en la plataforma de secuenciación de Illumina con bases mixtas (base "N") para dominios de código de barras

En la Tabla III, se diseñaron los oligonucleótidos bloqueantes siguientes para la utilización en experimentos de captura de híbridos para bibliotecas de moldes de ADN con la plataforma de secuenciación de Illumina. Los oligonucleótidos aumentados en T<sub>m</sub> se prepararon utilizando ANB ("C", "T" o "A") o ANBi ("iANBi-meC/", "iANBi-T/" o "iANBi-A/") como grupos que aumentan T<sub>m</sub>. Todos los oligonucleótidos se prepararon utilizando métodos químicos de fosforamida.

Tabla III. Bloqueantes oligonucleótidos aumentados en T<sub>m</sub> con secuencias de código de barras ("N")

SEC ID nº	Secuencia	nº ANB	T <sub>m</sub> (°C) 750 mM (Na)	T <sub>m</sub> (°C) 15 mM (Na)
17	IndexBlock:	0	-90	61.8
	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGA CTG GAGTT CAGACGTGTGCTCTTCCGATCT			
18	IndexBlock +10-ANB:	10	>90	70.9
	CAAG+CAGAAGA+CGG+CATA+CGAGATNNNNNNGTG A+CTGGAGTT+CAGA+CGTGTG+CTCTT+C+CGATCT			
19	IndexBlock +20-ANB:	20	>90	76.8
	CAAG+CAGAAGA+CGG+CA+TA+CGAGA+TNNNNNNG +TGA+C+TGGAG+T+T+CAGA+CG+TG+TG+CTC+TT +C+CGA+TCT			
20	IndexBlock RevComp:	0	-90	61.8
	AGATCGGAAGACACACGTCTGAACTCCAGTCACNNN NNNATCTCGTATGCGCTCTTCTGCTTG			
21	IndexBlock RevComp +10-ANB:	10	>90	70.9
	AGAT+CGGAAGAG+CACA+CGTCTGAA+CTCCAGT+C A+CNNNNNNATCT+CGTATGC+CGT+CTTCTG+CTTG			
22	IndexBlock_RevComp +20-ANB:	20	>90	80.0
	AGA+T+CGGAAGAG+CA+CA+CGT+CTGAA+CT+C+C AGT+CA+CNNNNNNAT+CT+CG+TA+TG+C+CGT+CT T+CTG+CTTG			
23	PE1.0:	0	-90	62.0
	AATGATACGGCGACACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT			
24	PE1.0 +10-ANB:	10	>90	72.6
	AATGATA+CGG+CGA+CCA+CCGAGAT+CTACA+CTC TTTC+CCTACACGACGCT+CTTC+CGAT+CT			

SEC ID nº	Secuencia	nº ANB	T <sub>m</sub> (°C) 750 mM (Na)	T <sub>m</sub> (°C) 15 mM (Na)
25	PE1.0 +17-ANB:	17	>90	81.0
	AATGATA+CGG+CGA+CCA+CCGAGAT+CTA+CA+CT+CTTT+CC+CTA+CA+CGA+CG+CT+CTTC+CGAT+CT			
26	PE1.0_Rev Comp:	0	-90	62.0
	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT			
27	PE1.0_Rev Comp +9-ANB:	9	>90	69.0
	AGAT+CGGAAGAG+CGT+CGTGTAGGGAAAGAGTGTAGAT+CT+CGGTGGT+CG+C+CGTAT+CATT			
28	PE1.0_Rev Comp +20-ANB:	20	>90	77.3
	AG+AT+CGG+AAG+AG+CGT+CGTGT+AGGG+AA+AG+AGTGT+AG+AT+CT+CGGTGGT+CG+C+CGT+AT+C+ATT			

La Tabla III proporciona ejemplos en los que se han diseñado oligonucleótidos aumentados en T<sub>m</sub> utilizando cualquiera de las cadenas de la secuencia adaptadora como bloqueante. La cadena preferente para la utilización como el oligonucleótido aumentado en T<sub>m</sub> como bloqueante es una que proporciona el máximo “poder bloqueante” por cada grupo modificado (es decir, el valor de T<sub>m</sub> aumentada óptimo más elevado) con inclusión del menor número de grupos que aumentan la T<sub>m</sub>. Por ejemplo, comparar las SEC ID nº 19 y nº 22 (siendo preferente la SEC ID nº 22) y las SEC ID nº 25 y nº 28 (siendo preferente la SEC ID nº 25).

#### Ejemplo 5. Oligonucleótidos aumentados en T<sub>m</sub> para la utilización en la plataforma de secuenciación PGM de Ion Torrent.

En la Tabla IV, se diseñaron los oligonucleótidos siguientes para la utilización en experimentos de captura de híbridos para bibliotecas de moldes de ADN con la plataforma de secuenciación PGM de Ion Torrent. Los oligonucleótidos aumentados en T<sub>m</sub> se prepararon utilizando ANB (“+C” o “+A”) o ANBi (“/iANBi-meC/” o “/iANBi-A/”) como grupos que aumentan T<sub>m</sub>. Todos los oligonucleótidos se prepararon utilizando métodos químicos de fosforamidita.

Tabla IV. Bloqueantes oligonucleótidos aumentados en T<sub>m</sub> para adaptadores de Ion Torrent

SEC ID nº	Secuencia	nº ANB	T <sub>m</sub> (°C) 750 mM (Na)	T <sub>m</sub> (°C) 15 mM (Na)
29	Ion P1 parte superior:	0	89.2	62.3
	CCACTACGCCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT			
30	Ion P1 parte superior 11C:	11	>90	77.9
	C+CA+CTA+CGC+CTC+CG+CTTTC+CT+CT+CTATGGG+CAGT+CGGTGAT			
31	Ion P1 parte inferior:	0	89.6	62.1
	ATCACC GACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT			
32	Ion P1 parte inferior 6C8A:	14	>90	77.8
	AT+CAC+CGA+CTGC+CC+AT+AG+AG+AGGA+A+AG+CGG+AGG+CGT+AGTGGTT			
33	Ion A parte superior:	0	83.2	57.4
	CCATCTCATCCCTGCGTGTCTCCGACTCAG			
34	Ion A parte superior 11C:	11	>90	76.1
	C+CAT+CT+CAT+C+CCTG+CGTGT+CT+C+CGA+CT+CAG			
35	Ion A parte inferior:	0	84.0	57.4
	CTGAGTCGGAGACACGCAGGGATGAGATGGTT			
36	Ion A parte inferior 5C5A:	10	>90	74.3
	+CTG+AGT+CGG+AGA+CA+CG+CAGGG+ATG+AG+ATTGGTT			

La Tabla IV proporciona ejemplos adicionales en los que se han diseñado oligonucleótidos aumentados en  $T_m$  utilizando cualquiera de las cadenas de la secuencia adaptadora como bloqueante. La cadena preferente para la utilización como el oligonucleótido aumentado en  $T_m$  como bloqueante es una que proporciona el máximo “poder bloqueante” por cada grupo modificado (es decir, el valor de  $T_m$  aumentada óptimo más elevado) con inclusión del menor número de grupos que aumentan la  $T_m$ . El presente ejemplo muestra además que, dependiendo de la cadena seleccionada como el oligonucleótido aumentado en  $T_m$ , ANB-C presenta un “poder bloqueante” superior por cada grupo que aumenta  $T_m$  que ANB-A. Por ejemplo, comparar las SEC ID nº 30 y nº 32 (siendo preferente la SEC ID nº 30) y las SEC ID nº 34 y nº 36 (siendo preferente la SEC ID nº 34).

#### **Ejemplo 6. Utilización de oligonucleótidos aumentados en $T_m$ como oligonucleótidos bloqueantes en secuenciación de Illumina**

Con el fin de validar la eficacia de los bloqueantes aumentados en  $T_m$ , se prepararon ocho bibliotecas con códigos de barras individualizados (indexados) a partir de la misma fuente de ADN NA07034 ((Coriell Institute for Medical Research, Camden, NJ) utilizando los adaptadores TruSeq de Illumina (01, 03, 05, 06, 08, 09, 10, 11) (Illumina, San Diego, CA), siguiendo el protocolo de preparación de bibliotecas de Illumina. Se rompió mecánicamente un microgramo de ADN genómico utilizando un ultrasonificador S220 Focused (Covaris, Woburn, MA) a una potencia máxima de 175 W, factor de servicio 2.0 y 200 ciclos/ráfaga. El ADN roto mecánicamente se sometió a purificación utilizando un sistema Agencourt AMPure XP (Beckman Coulter), se repararon los extremos y añadieron colas de A utilizando un kit de prep. de muestras de ADN TruSeq de Illumina.

Siete de las bibliotecas se sometieron a enriquecimiento basado en captura de híbridos utilizando los diversos bloqueantes listados en la Tabla V, utilizando el protocolo delineado en el Ejemplo 1, con 500 ng de cada biblioteca y 1 nanomol de cada oligonucleótido bloqueante. El reservorio de oligonucleótidos de captura consistía en aproximadamente 11,000 oligonucleótidos 120-meros 5'-biotinilados (sondas Lockdown™, Integrated DNA Technologies, Coralville, Iowa) con diana en 265 genes. Las secuencias capturadas/enriquecidas se amplificaron, se normalizaron y se agruparon con la inclusión de la biblioteca no enriquecida. El material agrupado se secuenció utilizando un secuenciador de laboratorio MiSeq de Illumina. Los resultados de las secuencias se analizaron utilizando la plataforma de biología computacional Galaxy. La figura 11 muestra el número de secuencias únicas obtenido para cada biblioteca indexada. El número de secuencias estaba comprendido entre 4.3 millones y 3.1 millones, y no se observó ninguna reducción significativa con cualquiera de las secuencias de oligo bloqueante dadas.

Tabla V. Oligonucleótidos utilizados con los adaptadores TruSeq de Illumina.

SEC ID nº	Nombre	Secuencia
81	TruSeq P7 Índice 5	GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGT ATGCCGTCTTCTGCTTG
23	TruSeq P5 Universal	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT
82	TruSeq P7 Índice complementario 6 x I	CAAGCAGAAGACGGCATACGAGAT/ideoxyI//ideoxyI//ideoxyI//i deoxyI//ideoxyI/GTGACTGGAGTTACAGACGTGTGCTCTTCCGATCT/3Sp C3/
83	TruSeq P5 Universal	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT/3SpC3/
84	TruSeq P7 ANBi complementario Med 6-I	/5BNA-mcC/AAG/iBNA-mcC/AGAAGA/iBNA-mcC/GG/iBNA- mcC/ATA/iBNA-mcC/GAGAT/ideoxyI//ideoxyI//ideoxyI// /ideoxyI//ideoxyI/GTGA/iBNA-mcC/TGGAGTT/iBNA-mcC/AGA/iBNA- mcC/GTGTG/iBNA-mcC/T/iBNA-mcC/TT/iBNA-mcC/iBNA- mcC/GAT/iBNA-mcC/T/3SpC3/
85	TruSeq P5 ANBi Universal Med	AATGATA/iBNA-mcC/GG/iBNA-mcC/GA/iBNA-mcC/CAC/iBNA- mcC/GAGAT/iBNA-mcC/TA/iBNA-mcC/ACT/iBNA-mcC/TTT/iBNA- mcC/CCTA/iBNA-mcC/ACGA/iBNA-mcC/GCT/iBNA-mcC/TTT/iBNA- mcC/GAT/iBNA-mcC/T/3SpC3/
86	TruSeq P7 ANB complementario Alto 6-I	/5BNA-mcC/AAG/iBNA-mcC/AGA/iBNA-A/GA/iBNA-mcC/GG/iBNA- mcC/ATA/iBNA-mcC/G/iBNA-A/G/iBNA-A/T/ideoxyI//ideoxyI//ideoxyI// /ideoxyI//ideoxyI//ideoxyI/GTG/iBNA-A//iBNA-mcC/TGG/iBNA- A/GTT/iBNA-mcC/AG/iBNA-A//iBNA-mcC/GTGTG/iBNA-mcC/T/iBNA- mcC/TT/iBNA-mcC//iBNA-mcC/G/iBNA-A/T/iBNA-mcC/T/3SpC3/
87	TruSeq P5 ANBi Universal Alto	AATGATA/iBNA-mcC/GG/iBNA-mcC/GA/iBNA-mcC/CAC/iBNA-mcC/G AGAT/iBNA-mcC/TA/iBNA-mcC/A/iBNA-mcC/T/iBNA-mcC/TTT /iBNA-mcC/C/iBNA-mcC/TA/iBNA-mcC/A/iBNA-mcC/GA/iBNA-mcC/G /iBNA-mcC/T/iBNA-mcC/TT/iBNA-mcC//iBNA-mcC/GAT/iBNA-mcC/T /3SpC3/

88	P7 Índice 6-1	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC/ideoxyI//ideoxyI//ideoxyI//ideoxyI/ATCTCGTATGCCGTCTTCTGCTTG/3SpC3/
89	P5 complementario	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGT CGCCGTATCATT/3SpC3/
90	P7 Med ANBi índice 6-l	GAT/iBNA-mC/GGAAGAG/iBNA-mC/ACACGT/iBNA-mC/TGAA/iBNA-mC/TCCAGT/iBNA-mC/A/iBNA-mC//ideoxyI//ideoxyI//ideoxyI//ideoxyI/AT/iBNA-mC/T/iBNA-mC/GTATGC/iBNA-mC/GTCTTCTG/iBNA-mC/TTG/3SpC3/
91	P5 complementario Med ANBi	AGAT/iBNA-mC/GGAAGAG/iBNA-mC/GT/iBNA-mC/GTG TAGGGAAAGAGTGTAGAT/iBNA-mC/T/iBNA-mC/GGTGGT/iBNA-mC/G/iBNA-mC//iBNA-mC/GTAT/iBNA-mC/ATT/3SpC3/
92	P7 Alto ANBi índice 6-l	GAT/iBNA-mC/GGAAGAG/iBNA-mC/A/iBNA-mC/A/iBNA-mC/GT/iBNA-mC/TGAA/iBNA-mC/T/iBNA-mC//iBNA-mC/AGT/iBNA-mC/A/iBNA-mC//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/AT/iBNA-mC/T/iBNA-mC/GTATG/iBNA-mC//iBNA-mC/
		GT/iBNA-mC/TT/iBNA-mC/TG/iBNA-mC/TTG
93	P5 complementario alto ANBi	AG/iBNA-A/T/iBNA-mC/GG/iBNA-A//iBNA-A/G/iBNA-A/G/iBNA-mC/GT/iBNA-mC/GTGT/iBNA-A/GGG/iBNA-A//iBNA-A//iBNA-A/G/iBNA-A/GTGT/iBNA-A/G/iBNA-A/T/iBNA-mC/T/iBNA-mC/GGTGGT/iBNA-mC/G/iBNA-mC//iBNA-mC/GT/iBNA-A/T/iBNA-mC//iBNA-A/TT/3SpC3/
94	PEblock_0-ANBi	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGT CGCCGTATCATT
95	PEblock_8-ANBi	AGAT/iBNA-mC/GGAAGAG/iBNA-mC/GT/iBNA-mC/GTGTAGGGAA AGAGTGTAGAT/iBNA-mC/T/iBNA-mC/GGTGGT/iBNA-mC/GC/iBNA-mC/GTAT/iBNA-mC/ATT
96	PEblock_22-ANBi	AG/iBNA-A/T/iBNA-mC/GG/iBNA-A//iBNA-A/G/iBNA-A/G/iBNA-mC/GT/iBNA-mC/GTGT/iBNA-A/GGG/iBNA-A//iBNA-A//iBNA-A/G/iBNA-A/GTGT/iBNA-A/T/iBNA-mC/T/iBNA-mC/GGTGGT/iBNA-mC/G/iBNA-mC//iBNA-mC/GT/iBNA-A/T/iBNA-mC//iBNA-A/TT
97	InBlock_0-ANBi	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC/ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/ATCTCGTATGCCGTCTTCTGCTTG
98	Inblock_8-ANBi	GAT/iBNA-mC/GGAAGAG/iBNA-mC/ACACGTCTGAA/iBNA-mC/TC CAGT/iBNA-mC/A/iBNA-mC//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/ATCT/iBNA-mC/GTATGC/iBNA-mC/GTCTT CTG/iBNA-mC/TTG
99	Inblock_17-ANBi	GAT/iBNA-mC/GGAAGAG/iBNA-mC/A/iBNA-mC/A/iBNA-mC/GT/iBNA-mC/TGAA/iBNA-mC/T/iBNA-mC//iBNA-mC/AGT/iBNA-mC/A/iBNA-mC//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/AT/iBNA-mC/T/iBNA-mC/GTATG/iBNA-mC//iBNA-mC/GT/iBNA-mC/TT/iBNA-mC/TG/iBNA-mC/TTG

Una métrica de rendimiento es el porcentaje de secuencias que se alinean (se localizan en el mapa) en las secuencias diana deseada a una profundidad de cobertura de 20-30 veces ("20x-30x"). La figura 12 muestra la profundidad de cobertura para los oligonucleótidos bloqueantes de ADN no modificados, así como los oligonucleótidos bloqueantes aumentados en T<sub>m</sub>. Para explicar la varianza en las secuencias de índice, es habitual de la técnica utilizar una serie de posiciones de base degeneradas (la totalidad de las 4 bases nucleósidas) o bases universales tales como desoxiinosina. Los valores de profundidad de cobertura se muestran en la Tabla VI. El rendimiento de los oligonucleótidos bloqueantes aumentados en T<sub>m</sub> (que contienen bases universales) muestra un incremento global de la profundidad de cobertura de 20x y 30x en comparación con oligonucleótidos bloqueantes aumentados en T<sub>m</sub> que contienen las mismas bases universales. Las lecturas de secuenciación duplicadas fueron eliminadas antes del análisis.

Tabla VI. Valores de profundidad de cobertura obtenidos con oligonucleótidos bloqueantes.

Código de barras:	Índice 05	Índice 06	Índice 01	Índice 03	Índice 10	Índice 08	Índice 09
Pares de oligos bloqueantes utilizados	P7	P7Comp 6xl	P7Comp Med ANBi 6xl	P7Comp Alta ANBi 6xl	P7 6xl	P7 Med ANBi 6xl	P7 Alto ANBi 6xl
	+	+	+	+	+	+	+

Código de barras:	Índice 05	Índice 06	Índice 01	Índice 03	Índice 10	Índice 08	Índice 09
	P5	P5	P5 Med ANBi	P5 Alta ANBi	P5 Comp	P5Comp Med ANBi	P5Comp Alta ANBi
% Dianas cubiertas							
>1x							
2x	99.08%	98.99%	99.10%	99.11%	99.04%	99.12%	99.04%
10x	98.69%	97.78%	98.71%	98.68%	97.65%	98.84%	98.62%
20x	98.05%	92.12%	98.36%	98.13%	90.76%	98.46%	97.83%
30x	96.74%	81.09%	97.57%	97.10%	78.14%	97.63%	96.06%

Otra métrica de rendimiento es el porcentaje de secuencias que se alinea con la secuencia de interés (porcentaje en la diana). La figura 13 resume los porcentajes en la diana obtenidos con los diversos oligonucleótidos bloqueantes. Se observó una mejora significativa (16-19%) con los bloqueantes aumentados en  $T_m$  con bases universales de desoxiinosina en comparación con los bloqueantes de ADN no modificados con bases universales de desoxiinosina. Se observó una mejora limitada (9%) al comparar los oligonucleótidos bloqueantes que no contenían ANBi con la secuencia de índice específica con los oligonucleótidos bloqueantes que contenían ANBi que contenían desoxiinosinas. Además, no se observó ninguna mejora adicional significativa en 13 de los 20 oligonucleótidos bloqueantes que contenían ANBi (39.86% vs. 38.09%).

Una tercera métrica de rendimiento es el nivel medio de enriquecimiento en las secuencias diana respecto a las secuencias no diana; la figura 14 resume los factores de enriquecimiento. Se observó un enriquecimiento significativo en el factor de enriquecimiento con los oligonucleótidos bloqueantes aumentados en  $T_m$  respecto a los oligonucleótidos bloqueantes no aumentados. El intervalo de enriquecimiento era de entre un mínimo de 197 para los oligonucleótidos bloqueantes no aumentados que contenían desoxiinosinas y un máximo de 580 veces para los oligonucleótidos bloqueantes aumentados en  $T_m$ .

Se obtuvieron tendencias comparables en los experimentos de captura de híbridos utilizando una biblioteca de Illumina que presentaba una secuencia de índice de código de barras de 8 nucleótidos con parejas de oligonucleótidos bloqueantes aumentados en  $T_m$  correspondientes a las SEC ID nº 95 y nº 98 y las SEC ID nº 96 y nº 99. La pareja de oligonucleótidos bloqueantes no modificados correspondiente a las SEC ID nº 94 y nº 97 proporcionó aproximadamente 50% de las lecturas secuenciadas en la diana. En contraste, las parejas de oligonucleótidos bloqueantes aumentados en  $T_m$  correspondientes a las SEC ID nº 95 y nº 98 y a las SEC ID nº 96 y nº 99 proporcionó aproximadamente 70% y aproximadamente 75% de lecturas secuenciadas en la diana, respectivamente. De esta manera, los bloqueantes aumentados en  $T_m$  que presentan 8 grupos que aumentan la  $T_m$  proporcionaron un incremento relativo aproximado respecto a los bloqueantes no modificados de aproximadamente 40%  $(([70\%-50\%]/50\%)\times 100\%)$ . Y los bloqueantes aumentados en  $T_m$  que presentan 17-22 grupos que aumentan la  $T_m$  proporcionaron un incremento relativo aproximado respecto a los bloqueantes no modificados de aproximadamente 50%  $(([75\%-50\%]/50\%)\times 100\%)$ . (Ver la figura 15).

Los ejemplos A-N dados a conocer posteriormente presentan los elementos de una forma de realización para un método de análisis multigénico de una muestra tumoral, que se ilustra mediante el diagrama de flujo proporcionado en la figura 3.

#### **Ejemplo A. Aislamiento de ácidos nucleicos de una muestra tumoral**

Tres secciones de 20  $\mu\text{m}$  de un bloque de parafina se mezclaron con 400  $\mu\text{l}$  de amortiguador FLTL mediante agitación con vórtex y se incubaron a 90°C durante 15 minutos en un tubo de centrífuga de 1.5 ml. Para la incubación, un intervalo de 88-92°C resultaba aceptable. A continuación, se incubó la muestra con 20  $\mu\text{l}$  de proteinasa K a 55°C durante 6 horas y 10  $\mu\text{l}$  de ARNasa (1 mg/ml) a temperatura ambiente durante 5 minutos. A continuación, se añadieron a la muestra 460  $\mu\text{l}$  de amortiguador BL y 500  $\mu\text{l}$  de etanol absoluto. La solución de muestra resultante se mantuvo a temperatura ambiente hasta la utilización posterior.

Para preparar la columna para la unión de ADN, se añadieron 100  $\mu\text{l}$  de amortiguador de equilibrado a una columna MicroElute y la columna se centrifugó a 10,000 x g durante 30 segundos. Se transfirieron 700  $\mu\text{l}$  de la solución de muestra indicada anteriormente a la columna MicroElute y la columna se centrifugó a 10,000 x g durante 1 minuto. Se repitió la etapa de centrifugación en el caso de que el fluido no pasase por completo por la columna MicroElute. La solución de muestra restante se aplicó a la columna MicroElute de la manera indicada anteriormente. A continuación, la columna MicroElute se trató con 500  $\mu\text{l}$  de amortiguador HB y se centrifugó a 10,000 x g durante 1 minuto. Seguidamente, se transfirieron 700  $\mu\text{l}$  del amortiguador de lavado de ADN a la columna MicroElute y la columna se centrifugó a 10,000 x g durante 1 minuto. La columna MicroElute se lavó nuevamente utilizando 700  $\mu\text{l}$  de amortiguador 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. La columna MicroElute se introdujo en un tubo de centrífuga de 1,5 ml estándar con el tapón destapado. Se añadieron a la columna 50-75  $\mu\text{l}$  de amortiguador de

elución precalentado a 70°C y se incubaron a temperatura ambiente durante 3 minutos. La columna se centrifugó en tubo de colección a >13,000 x g durante 1 minuto. Se añadieron otros 50-75 µl de amortiguador de elución precalentado a 70°C a la columna MicroElute y se incubaron a temperatura ambiente durante 3 minutos. La columna se centrifugó nuevamente en tubo de colección a >13,000 x g durante 1 minuto. La solución entera se transfirió a un tubo de centrifuga nuevo de 1,5 ml y se almacenó a -20°C.

En el kit de ADN FFPE E.Z.N.A.™ se proporciona amortiguador FTL, proteinasa K, amortiguador BL, amortiguador de equilibrado, columna MicroElute, amortiguador HB, amortiguador de lavado de ADN y amortiguador de elución (OMEGA bio-tek, Norcross, GA, nº de cat. D3399-00, D3399-01 y D3399-02).

Se dan a conocer métodos adicionales para aislar ácidos nucleicos (por ejemplo, de ADN) a partir de tejidos fijados con formaldehído o paraformaldehído incluidos en parafina (FFPE) en, por ejemplo, Cronin M. et al., Am. J. Pathol. 164(1):35-42, 2004; Masuda N. et al., Nucleic Acids Res. 27(22):4436-4443, 1999; Specht K. et al., Am. J. Pathol. 158(2):419-429, 2001, Ambion RecoverAll™ Total Nucleic Acid Isolation Protocol (Ambion, nº de cat. AM1975, septiembre de 2008), manual técnico del kit de purificación de ADN de LEV Maxwell® 16 FFPE Plus (Promega, literatura nº TM349, febrero de 2011) y manual de tejidos FFPE ADN QIAamp® (Qiagen, nº de cat. 37625, octubre de 2007). El kit de aislamiento de ácidos nucleicos totales RecoverAll™ utiliza xileno a temperaturas elevadas para solubilizar muestras incluidas en parafina y un filtro de fibra de vidrio para capturar los ácidos nucleicos. Se utiliza el kit de purificación de ADN Maxwell® 16 FFPE Plus LEV con el instrumento Maxwell® 16 para la purificación del ADN genómico a partir de secciones de 1 a 10 µm de tejido FFPE. El ADN se purifica utilizando partículas paramagnéticas (PMP) revestidas de sílice y eluyendo en un volumen de elución pequeño. El kit de tejidos FFPE ADN QIAamp® utiliza microtecnología de ADN QIAamp® para la purificación de ADN genómico y mitocondrial.

#### **Ejemplo B.1. Rotura mecánica del ADN**

El instrumento Covaris™ E210 con refrigerante en circulación se reguló a 4°C. El tanque de agua del instrumento se llenó con agua destilada/desionizada hasta el nivel "6" en la línea de llenado. Se abrió el software SonoLab™ y se dejó que el sistema ejecutase la secuencia de "homing" al solicitarlo. El agua en el tanque del instrumento se desgasificó durante como mínimo 45 minutos antes de romper mecánicamente las muestras.

Para preparar las muestras de ADN genómico para la rotura mecánica, las muestras en primer lugar se cuantificaron utilizando un ensayo PicoGreen® (Invitrogen) en un lector de microplacas (Spectramax M2, Molecular Devices). Basándose en la concentración, se utilizaron 120 µl del ADN de entrada deseado (2 ng/µl) con bajo contenido de TE (Tris 10 mM, EDTA 0.2 mM, pH 8.0) para el experimento. Las muestras individuales de 100 µl se pipetearon lentamente en microtubos Covaris (Covaris, nº de cat. 520045) a través del septo en el tapón del tubo. A continuación, los microtubos Covaris en la gradilla de tubos serie E de Covaris. Para la rotura mecánica de 200 pb, los parámetros de configuración eran los siguientes: 10% de ciclo de servicio, intensidad 5, 200 ciclos/ráfaga, tiempo: 180 s y modo de barrido de frecuencias. Tras la rotura mecánica, los microtubos Covaris se sedimentaron brevemente utilizando un adaptador apropiado en una minicentrífuga y las muestras rotas mecánicamente se transfirieron a tubos de microcentrifuga nuevos de 1.5 ml. Cada muestra de ADN roto mecánicamente se purificó utilizando una columna MinElute® de QIAGEN. Brevemente, se añadió 5 x amortiguador PBI QIAGEN a la muestra en un tubo de microcentrifuga de 1.5 ml (por ejemplo, se añadieron 500 µl amortiguador PBI a 100 µl de muestra). Cada muestra se agitó con vórtex, se sedimentó brevemente y se transfirió a una columna de centrifuga MinElute. La columna de centrifuga MinElute se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. Se añadieron 750 µl de amortiguador PE QIAGEN a la columna, se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. La columna de centrifuga se centrifugó nuevamente a 13,000 rpm durante 1 minuto y se transfirió a un tubo de microcentrifuga de 1.5 ml limpio. La columna se secó al aire durante 2-3 minutos. Para la primera elución, se añadieron a cada columna 18 µl de amortiguador de elución de QIAGEN, se incubó durante 2-3 minutos y después se centrifugó a 13,000 rpm durante 1 minuto. Para la segunda elución, se añadieron 15 µl de amortiguador de elución de QIAGEN, se incubó durante 1 min y después se centrifugó a 13,000 rpm durante 1 minuto. Se recolectó el eluyente y se desechó la columna de centrifuga.

Típicamente, se utilizan 200 ng para la rotura mecánica del ADN, aunque la cantidad de ADN puede estar comprendida entre 20 y 200 ng o más.

#### **Ejemplo B.2. Alternativa a la rotura mecánica del ADN**

El presente ejemplo describe un método alternativo de romper mecánicamente el ADN del Ejemplo 2A.

En primer lugar, se desnaturizó un ADN genómico de doble cadena en ADN de cadena sencilla y después se mezcló con cebadores, ADN polimerasa (por ejemplo, ADN polimerasa Exo), dNTP y una pequeña cantidad de ddNTP. La secuencia del cebador puede ser un hexámero aleatorio o un hexámero aleatorio etiquetado con una secuencia adaptadora en el extremo 5'. Los métodos de utilización de la amplificación de hexámeros aleatorios etiquetados para clonar y secuenciar cantidades minúsculas de ADN se describen en, por ejemplo, Wong K.K. et al., Nucleic Acids Res. 24(19):3778-83, 1996. La reacción se incubó bajo condiciones que permitían el apareamiento de cebador-molde y la síntesis del ADN. La síntesis de ADN se termina cuando se incorpora un

ddNTP en la primera cadena recién sintetizada. La longitud de la primera cadena de ADN recién sintetizado puede controlarse con la proporción de dNTP a ddNTP. Por ejemplo, la proporción molar de dNTP a ddNTP es de por lo menos aproximadamente 1000:1, de aproximadamente 5000:1 o de aproximadamente 10000:1. Después de la síntesis de la primera cadena, los fragmentos cortos (tales como cebadores y primera cadena de ADN sintetizada de corta longitud y ddNTP) pueden eliminarse mediante selección por tamaño (por ejemplo, utilizando una columna de centrífuga de selección por tamaño). El ADN de primera cadena resultante se mezcla con cebadores (por ejemplo, hexámeros aleatorios o hexámeros aleatorios etiquetados con una secuencia adaptadora), ADN polimerasa (por ejemplo, ADN polimerasa Exo+) y dNTP. Puede utilizarse una ADN polimerasa Exo+ para eliminar el 3'-ddNTP terminal del ADN de primera cadena o incluso para generar extremos romos en el segundo sitio de cebado. A continuación, se incubó la reacción bajo condiciones que permiten el apareamiento de cebador-molde y la síntesis del ADN. Después de la síntesis de la segunda cadena, los fragmentos de ADN de doble cadena resultantes pueden purificarse y utilizarse directamente en la construcción de bibliotecas. Alternativamente, los fragmentos de ADN de doble cadena pueden amplificarse por PCR utilizando cebadores que contienen secuencias adaptadoras en el caso de que estas secuencias adaptadoras hayan sido incluidas en los cebadores para la síntesis de primera y segunda cadena. Los cebadores para la amplificación por PCR pueden incluir además las secuencias enteras y/o secuencias de código de barras.

### **Ejemplo C. Preparación de bibliotecas**

#### Reacción de reparación de extremos

Se descongelaron reactivos de reparación de extremos (NEB nº E6050L) y se preparó una mezcla maestra de reparación de extremos sobre hielo. Para preparar 70 µl de mezcla maestra para cada muestra, se mezclaron 55 µl de agua libre de nucleasas con 10 µl de amortiguador de reacción de reparación de extremos 10x y 5 µl de mezcla de enzima de reparación de extremos. A continuación, se añadieron 70 µl de mezcla maestra a 30 µl de cada muestra de ADN roto mecánicamente, en una placa de PCR de 96 pocillos sobre hielo. La reacción se incubó en un termociclador a 20°C durante 30 minutos. Cada muestra se purificó utilizando una columna MinElute® de QIAGEN. Brevemente, se añadió 5 x amortiguador PBI QIAGEN a la muestra (por ejemplo, se añadieron 500 µl amortiguador PBI a 100 µl de muestra) en un tubo de microcentrífuga de 1.5 ml. Cada muestra se agitó con vórtex, se sedimentó brevemente y se transfirió a una columna de centrífuga MinElute. La columna de centrífuga MinElute se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. Se añadieron 750 µl de amortiguador PE QIAGEN a la columna, se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. La columna de centrífuga se centrifugó nuevamente a 13,000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga de 1.5 ml limpio. La columna se secó al aire durante 2-3 minutos. Para la primera elución, se añadieron a cada columna 22 µl de amortiguador de elución de QIAGEN (Tris 10 mM, pH 8.5), se incubaron durante 2-3 minutos y después se centrifugó a 13,000 rpm durante 1 minuto. Para la segunda elución, se añadieron 22 µl de amortiguador de elución de QIAGEN, se incubó durante 1 min y después se centrifugó a 13,000 rpm durante 1 minuto. Se recolectó el eluyente y se desechó la columna de centrífuga.

#### Adición de base 3'-A

Se descongelaron reactivos de adición de base A (NEB nº E6053L) sobre hielo y se preparó la mezcla maestra de adición de base A sobre hielo. Para preparar 10 µl de mezcla maestra para cada muestra, se mezclaron 2 µl de agua libre de nucleasas con 5 µl de amortiguador de reacción de adición de colas 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 purificada en una placa de PCR de 96 pocillos sobre hielo. La reacción se incubó en un termociclador a 37°C durante 30 min. Se purificó cada muestra utilizando una columna MinElute® de QIAGEN. Brevemente, se añadió 5 x amortiguador PBI QIAGEN a la muestra (por ejemplo, se añadieron 250 µl amortiguador PBI a 50 µl de muestra) en un tubo de microcentrífuga de 1.5 ml. Cada muestra se agitó con vórtex, se sedimentó brevemente y se transfirió a una columna de centrífuga MinElute. La columna de centrífuga MinElute se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. Se añadieron 750 µl de amortiguador PE QIAGEN a la columna, se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. La columna de centrífuga se centrifugó nuevamente a 13,000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga de 1.5 ml limpio. La columna se secó al aire durante 2-3 min. Para la primera elución, se añadieron a cada columna 13 µl de amortiguador de elución de QIAGEN (Tris 10 mM, pH 8.5), se incubaron durante 2-3 minutos y después se centrifugó a 13,000 rpm durante 1 minuto. Para la segunda elución, se añadieron 13 µl de amortiguador de elución de QIAGEN, se incubó durante 1 min y después se centrifugó a 13,000 rpm durante 1 minuto. Se recolectó el eluyente y se desechó la columna de centrífuga.

#### Ligación de adaptadores multiplex

Se descongelaron reactivos de ligación (NEB nº 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 amortiguador de reacción de ligación Quick 5x a 3,3 µl de adaptador multiplex de Illumina (15 µM, incluido en Illumina, nº de cat. PE-400-1001) (se utilizaron 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 en primer lugar se diluyeron en agua (2 µl de adaptadores más 2 µl de H<sub>2</sub>O), después 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 de T4 Quick a la reacción de ligación.

Para >1 µg de material de partida, se utilizaron >3,3 µl de adaptadores. De esta manera, se añadió menos agua para mantener el volumen total de la mezcla diluida de adaptadores 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. La reacción se incubó en un termociclador a 25°C durante 30 min. Se purificó cada muestra utilizando una columna MinElute® de QIAGEN. Brevemente, se añadió 5 x amortiguador PBI QIAGEN a la muestra (por ejemplo, se añadieron 300 µl amortiguador PBI a 60 µl de muestra) en un tubo de microcentrífuga de 1.5 ml. Cada muestra se agitó con vórtex, se sedimentó brevemente y se transfirió a una columna de centrífuga MinElute. La columna de centrífuga MinElute se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. Se añadieron 750 µl de amortiguador PE QIAGEN a la columna, se centrifugó a 13,000 rpm durante 1 minuto y se desechó el eluido. La columna de centrífuga se centrifugó nuevamente a 13,000 rpm durante 1 minuto y se transfirió a un tubo de microcentrífuga de 1.5 ml limpio. La columna se secó al aire durante 2-3 minutos. Para la primera elución, se añadieron a cada columna 20 µl de amortiguador de elución de QIAGEN (Tris 10 mM, pH 8.5), se incubaron durante 2-3 minutos y después se centrifugó a 13,000 rpm durante 1 minuto. Para la segunda elución, se añadieron 20 µl de amortiguador de elución de QIAGEN, se incubó durante 1 min y después se centrifugó a 13,000 rpm durante 1 minuto. Se recolectó el eluyente y se desechó la columna de centrífuga.

#### Enriquecimiento mediante PCR

Se descongelaron los reactivos de PCR y se preparó una mezcla maestra de PCR sobre hielo. Para 62 µl de mezcla maestra por cada muestra, se utilizaron 50 µl de mezcla maestra de alta fidelidad Phusion 2x con amortiguador HF (Finnzyme, NEB nº de cat. F-531S), 8 µl de agua libre de nucleasas, 2 µl de cebador 1.0 de Illumina (25 µM) y 2 µl de cebador 2.0 de Illumina (0.5 µM). A continuación, se mezclaron 62 µl de mezcla maestra con 2 µl de cebador índice de Illumina (25 µM, incluido en Illumina nº de cat. PE-400-1001) con código de barras apropiado y 36 µl de muestra de ADN ligado en una placa de PCR de 96 pocillos.

La reacción se incubó en un termociclador de la manera siguiente:

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	mantenimiento	

Cada reacción de PCR se seleccionó según tamaño con 1.8x volumen de perlas AMPureXP (Agencourt, Beckman Coulter Genomics, nº de cat. A6388). Brevemente, se añadieron 1,8x de perlas AMPureXP a la muestra (por ejemplo, se añadieron 180 µl de perlas a 100 µl de muestra) en un tubo de microcentrífuga de 1.5 ml, se agitaron con vórtex y se incubaron durante 5 minutos con mezcla de rotación por volteo. Los tubos se colocaron en un soporte magnético hasta clarificar la solución (2 minutos). Se desechó el sobrenadante sin perturbar las perlas capturadas sobre el imán. Se añadieron 600 µl de etanol al 70% recién preparado a las perlas, se incubaron durante 1 min seguido de la eliminación del etanol. Se añadió una segunda alícuota de 600 µl de etanol al 70% recién preparado a las perlas, se incubaron durante 1 minuto y se eliminó el etanol. Los tubos se devolvieron al soporte magnético durante 1-2 minutos para recapturar las perlas. Se eliminó cualquier etanol remanente y las perlas se secaron al aire a temperatura ambiente durante 5-10 minutos. Se añadieron 30 µl de amortiguador de elución de QIAGEN a las perlas, se agitaron con vórtex y se incubaron durante 2 minutos. Los tubos se devolvieron al soporte magnético hasta clarificar la solución (2 minutos). El sobrenadante se transfirió a un tubo de 1.5 ml nuevo y se desecharon las perlas. Las muestras de ADN eluido se cuantificaron utilizando un ensayo de Q-PCR. Estas cuantificaciones permitieron el agrupado equimolar para garantizar una representación igual de cada biblioteca dentro de una selección de captura de híbridos agrupada.

#### **Ejemplo D. Selección de híbridos**

##### Reservorio de bibliotecas indexadas de muestras

Se prepararon reservorios (hasta 12-plex) de bibliotecas que habían sido indexadas, purificadas y cuantificadas mediante Q-PCR sobre hielo. Se prepararon reservorios equimolares en tubos de microcentrífuga de 1.5 ml para garantizar que cada muestra estaba representada uniformemente en el procedimiento de selección de híbridos. La entrada total de ADN en cada uno de dichos reservorios puede encontrarse comprendida entre 2000 ng y 500 ng. Típicamente, la entrada total de ADN es 2000 ng. De esta manera, en el caso de que se agrupen doce muestras, pueden agruparse 166.67 ng de cada para conseguir un total de 2000 ng. El volumen final de un reservorio de bibliotecas de 2000 ng debería ser de 4 µl. Debido a las concentraciones variables de las bibliotecas indexadas, puede prepararse un reservorio con cualquier volumen mayor, aunque en este caso el reservorio debe secarse mediante centrifuga Speedvac (utilizando un bajo nivel de calor) y reconstituirse en 4 µl de agua libre de nucleasas.

A mayor rendimiento en la construcción de una biblioteca, mayor será la complejidad de la biblioteca.

#### Hibridación de las bibliotecas de ADN agrupadas con señuelos de ARN biotinilado

- 5 En el presente experimento se utilizó el kit de extremos apareados de enriquecimiento de dianas Agilent SureSelect (nº G3360A-J). Se descongelaron sobre hielo amortiguador de hibridación nº 3, bloque SureSelect nº 1, bloque SureSelect nº 2, bloque de cebador de extremo apareado 1.0, bloque de cebador índice 1-12, bloque de ARNasa y señuelo de ARN biotinilado.
- 10 Se prepararon las mezclas maestras siguientes:
- a. mezcla de amortiguador de hibridación (13 µl en cada reacción):
    - 15 i. amortiguador de hibridación nº 1 (Agilent) - 25 µl
    - ii. amortiguador de hibridación nº 2 (Agilent) - 1 µl
    - iii. amortiguador de hibridación nº 3 (Agilent) - 10 µl
    - iv. amortiguador de hibridación nº 4 (Agilent) - 13 µl
  - b. mezcla de bloqueo (8 µl en cada reacción):
    - 20 i. bloque SureSelect nº 1 (Agilent) - 2.5 µl
    - ii. bloque SureSelect nº 2 (Agilent) - 2.5 µl
    - iii. bloque cebador de extremo apareado 1.0 (IDT, resuspendido hasta 200 µM con H<sub>2</sub>O) - 1.5 µl
    - 25 iv. bloque cebador índice 1-12 (IDT, resuspendido hasta 200 µM con H<sub>2</sub>O) - 1.5 µl
  - c. Dilución de bloque de ARNasa
    - 30 i. Para señuelos de ARN biotinilado personalizados con territorio <3 Mb: se diluyó 1 µl de bloque ARNasa (Agilent) en 9 µl de agua.
    - ii. Para los señuelos personalizados con un territorio del señuelo >3 Mb: se diluyó 1 µl de bloque ARNasa en 3 µl de agua (todavía 0.5 µl de bloque ARNasa por cada 7 µl de reacción de captura)
  - d. Mezcla de señuelos: (7 µl en cada reacción)
    - 35 i. Señuelos de ARN - 2 µl (para señuelos que presentan un territorio de señuelo >3 Mb, se utilizaron 5 µl de señuelo)
    - 40 ii. Bloque de ARNasa diluida - 5 µl (para señuelos que presentan un territorio del señuelo >3 Mb, se diluyeron 2 µl de bloque ARNasa diluido tal como se ha indicado anteriormente)

Una vez se habían preparado la mezcla de amortiguador de hibridación, mezcla de bloqueo y mezcla o mezclas de señuelos, se agitó con vórtex la mezcla de amortiguador de hibridación, se centrifugó y se calentó a 65°C en el bloque térmico. Se mezclaron 4 µl de cada biblioteca de muestras agrupada para la selección de híbridos, con 8 µl de la mezcla de bloqueo en una placa de PCR de 96 pocillos. La reacción se incubó en un termociclador a 95°C durante 5 minutos y después se mantuvo a 65°C. Tras incubar las bibliotecas de muestras agrupadas/mezcla de bloqueo a 95°C durante 5 min y después a 65°C durante 2.5 minutos, la mezcla de señuelos (=mezcla de señuelo/bloque de ARNasa) se introdujo en el bloque térmico a 65°C durante 2.5 minutos. Los tubos que contenían amortiguador de hibridación se centrifugaron rápidamente y después se devolvieron inmediatamente al bloque térmico a 65°C. Se pipetearon 13 µl de la mezcla de amortiguador de hibridación caliente en cada mezcla de biblioteca de muestras/bloque, mientras que la placa de 96 pocillos se dejó en el termociclador a 65°C. Tras incubar la mezcla de señuelos durante 2.5 minutos a 65°C, se añadieron 7 µl de la mezcla de señuelos a cada mezcla de biblioteca de muestras/bloque/amortiguador de hibridación, mientras que la placa de 96 pocillos permaneció en el termociclador a 65°C. La reacción (el volumen total era de 32 µl) se incubó a 65°C durante 24 horas en un termociclador.

#### Preparación de las perlas magnéticas

60 El amortiguador de lavado SureSelect nº 2 se precalentó a 65°C en el bloque térmico. Las perlas T1 con estreptavidina Dynal MyOne (Invitrogen) se agitaron con vórtex y se resuspendieron. Las perlas se lavaron mediante la adición de 200 µl de amortiguador de unión SureSelect por cada 50 µl de perlas Dynal (por ejemplo, se utilizaron 1200 µl de amortiguador de unión SureSelect para preparar 300 µl de perlas Dynal). Las perlas se agitaron con vórtex durante 5 segundos y se sedimentaron brevemente. Las perlas se introdujeron en un soporte magnético durante aproximadamente 15 segundos o hasta capturar la totalidad de las perlas. Se separó el sobrenadante y se desechó. Se repitió el lavado con amortiguador de unión SureSelect dos veces más hasta un total de tres lavados. Tras el lavado, las perlas se resuspendieron en 200 µl de amortiguador de unión SureSelect

por cada 50 µl de perlas Dynal (por ejemplo, se utilizaron 1200 µl de amortiguador de unión SureSelect para preparar 300 µl de perlas Dynal). Las perlas resuspendidas se agitaron con vórtex y se sedimentaron brevemente. Los 200 µl de perlas resuspendidas se dividieron en alícuotas para tubos de microcentrifuga individuales de 1.5 ml.

#### Selección de ADN de captura de híbridos

Tras 24 horas de incubación, cada muestra hibridada de la placa de PCR en el termociclador a 65°C se pipeteó rápidamente a un tubo que contenía 200 µl de perlas preparadas a temperatura ambiente. Las mezclas de muestras y perlas se agitaron con vórtex durante 5 segundos y se incubaron en un rotador a temperatura ambiente durante 30 minutos, para garantizar una mezcla correcta. A continuación, los tubos se centrifugaron rápidamente. Las perlas se capturaron sobre un imán (durante 2 minutos) y se separó el sobrenadante y se desechó. Las perlas se resuspendieron en 500 µl de amortiguador de lavado SureSelect nº 1 para un lavado de baja astringencia. Las muestras se agitaron con vórtex durante 5 segundos y se incubaron durante 15 min. a temperatura ambiente fuera del imán. Las muestras se agitaron con vórtex durante 5 segundos cada 3-5 minutos. Los tubos se centrifugaron rápidamente. A continuación, las perlas se capturaron sobre un soporte magnético durante 2 minutos y se separó el sobrenadante y se desechó. Para un lavado de alta astringencia para eliminar el material fuera de diana, las perlas se lavaron con amortiguador de lavado SureSelect nº 2 precalentado a 65°C. Brevemente, las perlas se resuspendieron en 500 µl de amortiguador de lavado SureSelect nº 2 precalentado y se mezclaron en un agitador de vórtex durante 5 segundos para resuspender las perlas. Las perlas se centrifugaron brevemente en una centrífuga y se incubaron a 65°C durante 10 min en un bloque térmico con mezcla con vórtex ocasional durante 5 segundos a temperatura ambiente. A continuación, las perlas se sedimentaron brevemente en una centrífuga y se capturaron sobre un imán durante 2 minutos. Se repitió el lavado 2 veces más con amortiguador de lavado SureSelect nº 2 precalentado a 65°C hasta un total de tres lavados. A continuación, el amortiguador de lavado se eliminó por completo y se añadieron 50 µl de amortiguador de elución SureSelect a las perlas tras la agitación con vórtex durante 5 segundos para mezclar las perlas. Las muestras se incubaron durante 10 minutos a temperatura ambiente con mezcla con vórtex ocasional durante 5 segundos. Las perlas se sedimentaron brevemente en una centrífuga y se capturaron sobre un soporte magnético. El sobrenadante que contenía el ADN capturado se pipeteó a un nuevo tubo de microcentrifuga de 1.5 ml. Se añadieron 50 µl de amortiguador de neutralización SureSelect al ADN capturado. Las muestras se agitaron con vórtex durante 5 segundos, se sedimentaron brevemente en una centrífuga y se purificaron utilizando un volumen 1.8x de perlas AMPureXP. Se eluyó el ADN en 40 µl de agua libre de nucleasas.

#### Enriquecimiento mediante 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 cada muestra, se mezclaron 50 µl de mezcla maestra de alta fidelidad Phusion 2x con amortiguador HF (NEB nº F-531S) con 8 µl de agua libre de nucleasas, 1 µl de cebador de QPCR 1.1 (100 µM en H<sub>2</sub>O) y 1 µl de cebador de QPCR nº 2 (100 µM en H<sub>2</sub>O). Las secuencias de cebador para la Q-PCR eran:

Cebador de QPCR 1.1 (purificado mediante HPLC a partir de IDT):

5'AATGATACGGCGACCAACCGAGAT3' (SEC ID nº 79)

Cebador de QPCR 2.1 (purificado mediante HPLC a partir de IDT):

5'CAAGCAGAAGACGGCATACGA3' (SEC ID nº 80)

Se añadieron 60 µl de mezcla maestra a 40 µl de cada muestra de ADN capturado purificada en una placa de PCR de 96 pocillos. La reacción se incubó en un termociclador de la manera siguiente:

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	mantenimiento

Cada 100 µl de reacción de PCR se purificaron con 1.8x el volumen de perlas AMPureXP y se eluyeron en 35 µl de amortiguador de elución (Tris 10 mM, pH 8.5). Las muestras de ADN seleccionadas/capturadas en híbridos se cuantificaron utilizando un ensayo de Q-PCR. El ensayo Q-PCR detectó los adaptadores terminales y las lecturas indicaron cuánto de cada muestra debía cargarse en una celda de flujo de secuenciación para conseguir la densidad de agregado apropiada.

**Ejemplo E. Métodos**

A continuación, se ejemplifican determinadas formas de realización de los métodos y condiciones experimentales utilizadas para identificar las alteraciones según los Ejemplos. Puede llevarse a cabo un cribado adicional de traslocaciones mediante la utilización de, por ejemplo, análisis de qRT-PCR del ADNc preparado a partir de una muestra tumoral preseleccionada.

Se llevó a cabo secuenciaciones de ADN en paralelo masivas en bibliotecas capturadas mediante hibridación, basadas en la ligación de adaptadores, utilizando ADN aislado a partir de tejido fijado e incluido en parafina de archivo. Se utilizó una combinación de herramientas de análisis para analizar los datos y asignar llamadas de alteración de ADN. Se llevó a cabo cribado adicional de traslocaciones mediante la utilización de análisis de qRT-PCR de ADNc preparado a partir de tumores congelados o evaluación de IHC de especímenes FFPE de archivo. Se llevó a cabo secuenciación de ADNc en paralelo masiva para confirmar la expresión de ambas nuevas traslocaciones utilizando ARN aislado a partir de tejido FFPE. El ADN genómico de referencia normal correspondiente procedente de sangre se secuenció para el paciente índice de CPCNP para confirmar el origen somático de la reorganización.

**Secuenciación de ADN genómico**

Se llevó a cabo la secuenciación de 2574 exones de 145 genes de cáncer utilizando ADN procedente de especímenes de tumor fijados en formalina e incluidos en parafina (FFPE) de archivo; 24 procedentes de pacientes de CPCNP. Las bibliotecas de secuenciación se construyeron mediante el método de ligación de adaptadores utilizando 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 llevó a cabo utilizando 36 x 36 lecturas apareadas hasta una profundidad media de 253x. Se llevó a cabo el procesamiento de datos y asignaciones de mutaciones para sustituciones de bases, mutaciones indel, alteraciones del número de copia y reorganizaciones genómicas, utilizando una combinación de herramientas optimizadas para la llamada de mutación de tejido tumoral.

**Secuenciación de ADNc**

Se generó ADNc a partir de ARN total extraído de una única sección de tejido FFPE de 5-10 µm utilizando el kit Roche High Pure y se transcribió inversamente en ADNc con cebadores hexámeros aleatorios mediante el sistema de síntesis de primera cadena SuperScript® III (Invitrogen). Se produjo ADNc de doble cadena con el módulo de síntesis de segunda cadena de ARNm NEBNext® (New England Biolabs) y se utilizó como entrada para la construcción de bibliotecas, captura de híbridos y secuenciación al igual que para las muestras FFPE de ADN. El análisis de los niveles de expresión se llevó a cabo con una combinación de herramientas de análisis.

**Ejemplo F: genes seleccionados ejemplificativos y variantes para el análisis multiplex**

El presente ejemplo proporciona cuatro tablas ejemplificativas que resumen una selección de genes, variantes y tipos de cáncer para el análisis multiplex.

Tabla 1. Lista de genes seleccionados ejemplificativos y variantes, tipos de cáncer asociados y codones de prioridad para el análisis multiplex.

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
ABL1	Prioridad 1	Leucemia (por ejemplo, leucemia mieloide crónica (LMC), leucemia mieloide aguda (LMA), leucemia linfoblástica aguda (LLA))	315
AKT1	Prioridad 1	Cáncer de mama, cáncer colorrectal, cáncer ovárico	
ALK	Prioridad 1	Linfoma (por ejemplo, linfoma no de Hodgkin, linfoma de células grandes anaplásico (LCGA), tumor miofibroblástico inflamatorio	
CPO	Prioridad 1	cáncer colorrectal, meduloblastoma, síndrome de cáncer de reparación de desajustes	1114, 1338, 1450, 1556
RA	Prioridad 1	Cáncer de próstata	
BRAF	Prioridad 1	cáncer de pulmón, linfoma no de Hodgkin, cáncer colorrectal, cáncer de tiroides, melanoma	600
CDKN2A	Prioridad 1	Melanoma, cáncer pancreático, síndrome de Li-Fraumeni, cáncer de pulmón (por ejemplo, cáncer de pulmón de células no pequeñas (CPCNP)), carcinoma de células escamosas, retinoblastoma, astrocitoma	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
CEBPA	Prioridad 1	Leucemia (por ejemplo, leucemia mieloide aguda (LMA), leucemia monoblástica), retinoblastoma	
CTNNB1	Prioridad 1	cáncer colorrectal, cáncer ovárico, cáncer de próstata, cáncer hepático (por ejemplo, hepatoblastoma (HB), carcinoma hepatocelular (HCC)), pilomatixoma, meduloblastoma, adenomas pleiomórficos de las glándulas salivales	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 ovárico, cáncer de pulmón	
ESR1	Prioridad 1	cáncer de mama, cáncer endometrial, adenocarcinoma endometrial, leiomioma, carcinoma ductal mamario	
FGFR1	Prioridad 1	Leucemia, linfoma	
FGFR2	Prioridad 1	cáncer de mama, cáncer prostático	
FGFR3	Prioridad 1	Cáncer de vejiga, cáncer cervical, mieloma múltiple	
FLT3	Prioridad 1	Leucemia (por ejemplo, leucemia mieloide aguda (LMA), leucemia promielocítica aguda, leucemia linfoblástica aguda)	835
HRAS	Prioridad 1	Carcinoma de tiroides de células de Hürthle, cáncer de vejiga, melanoma, cáncer colorrectal	12, 13, 61
JAK2	Prioridad 1	Leucemia (por ejemplo, leucemia linfoblástica crónica (LLC), leucemia linfoblástica aguda (LLA), leucemia mielógena crónica (LMC), leucemia mielógena aguda (LMA))	617
KIT	Prioridad 1	Tumor estromal gastrointestinal (TEGI), tumor testicular, leucemia (por ejemplo, leucemia mieloide aguda (LMA)), tumor de mastocitos, tumor mesenquimal, carcinoma quístico adenoide, cáncer de pulmón (por ejemplo, cáncer de pulmón de células pequeñas), 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 hereditaria (CRPH), cáncer de pulmón (por ejemplo, cáncer de pulmón de células no pequeñas), carcinoma tiroideo papilar, glioma, adenocarcinoma esofágico, osteosarcoma, cáncer endometrial, carcinoma de células escamosas, melanoma, cáncer de mama	
MLL	Prioridad 1	Leucemia (por ejemplo, leucemia linfoblástica aguda (LLA), leucemia mieloide 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 linfoblástica aguda (LLA)), carcinoma medular de tiroides, linfoma (por ejemplo, linfoma tímico, linfoma de células T)	1575, 1601
NPM1	Prioridad 1	Linfoma (por ejemplo, linfoma no de Hodgkin, linfoma de células grandes anaplásico, linfoma anaplásico), leucemia (por ejemplo, leucemia promielocítica aguda, leucemia mielógena aguda (LMA))	
NRAS	Prioridad 1	Leucemia (por ejemplo, leucemia mielomonocítica juvenil (LMMJ), leucemia mieloide aguda (LMA), leucemia linfoblástica aguda (LLA)), melanoma	12, 13, 61
PDGFRA	Prioridad 1	Tumor estromal gastrointestinal (TEGI), leucemia (por ejemplo, leucemia eosinofílica crónica (LEC), leucemia linfocítica aguda (LLA), tumor mesenquimal	
PIK3CA	Prioridad 1	cáncer colorrectal, cáncer de mama, cáncer ovárico, carcinoma hepatocelular, carcinoma de células escamosas de cabeza y cuello (CCECC), carcinoma anaplásico de tiroides, cáncer endometrial, adenocarcinoma de la vesícula biliar, glioblastoma	88, 542, 545, 546, 1047, 1049

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
PTEN	Prioridad 1	carcinoma de células escamosas de cabeza y cuello (CCECC), cáncer endometrial, glioma, cáncer de próstata, glioblastoma	130, 173, 233, 267
RB1	Prioridad 1	Retinoblastoma, cáncer de vejiga, osteosarcoma, cáncer de pulmón (por ejemplo, cáncer de pulmón de células pequeñas, cáncer de pulmón de células no pequeñas), leucemia (por ejemplo, leucemia linfoblástica aguda (LLA))	
RET	Prioridad 1	cáncer colorrectal, carcinoma medular de tiroides, neoplasia múltiple tipo 2B, feocromocitoma, neoplasia múltiple de tipo 2A, carcinoma papilar de tiroides, cáncer de tiroides, retinoblastoma	918
TP53	Prioridad 1	TP53 se encuentra frecuentemente mutado o inactivado en aproximadamente 60% de los cánceres, por ejemplo, el carcinoma de células escamosas esofágico, el síndrome de Li-Fraumeni, los carcinomas de células escamosas de cabeza y cuello (CCECC), cáncer de pulmón, carcinoma adrenocortical hereditario, astrocitoma, carcinoma de células escamosas, cáncer de vejiga, cáncer colorrectal, glioblastoma, retinoblastoma	175, 245, 248, 273, 306
ABL2	Gen de cáncer	Leucemia mieloide aguda (LMA)	
AKT2	Gen de cáncer	Cáncer ovárico, cáncer pancreático	
AKT3	Gen de cáncer	Melanoma, glioma, cáncer uterino, cáncer de próstata, cáncer oral, cáncer ovárico	
ARAF	Gen de cáncer	linfoma de células T angioinmunoblástico, tumor ascítico de Ehrlich	
ARFRP1	Gen de cáncer	Cáncer de mama	
ARID1A	Gen de cáncer	Neuroblastoma, leucemia linfoblástica aguda (LLA), tumor neuroendocrino	
ATM	Gen de cáncer	Leucemia (por ejemplo, leucemia prolinfocítica de células T (LPL-T), linfoma, meduloblastoma, glioma	
ATR	Gen de cáncer	Linfoma asociado a piotórax, linfoma de células T	
AURKA	Gen de cáncer	Carcinoma de células escamosas laríngeo, cáncer ovárico, cáncer de vejiga, carcinoma de células escamosas de cabeza y cuello (CCECC), carcinoma laríngeo, carcinoma de células escamosas esofágico (CCCE), cáncer pancreático	
AURKB	Gen de cáncer	Cáncer colorrectal, astrocitoma, tumor ependimario, glioma, carcinoma de células escamosas esofágico (CCCE), leucemia mieloide aguda (LMA)	
BCL2	Gen de cáncer	Linfoma, adenocarcinoma colorrectal, carcinoma de células escamosas esofágico (CCCE), sarcoma sinovial, leucemia	
BCL2A1	Gen de cáncer	Granuloma pulmonar, adenoma gástrico, linfoma de Burkitt, adenoma de parótida, sarcoma de Kaposi, cáncer gástrico, cáncer de colon	
BCL2L1	Gen de cáncer	carcinoma de células escamosas de cabeza y cuello, glioblastoma, mesotelioma, cáncer pancreático, adenocarcinoma pulmonar	
BCL2L2	Gen de cáncer	Cáncer de cerebro, leucemia, linfoma, adenocarcinoma colorrectal, cáncer colorrectal, adenoma, carcinoma de células escamosas cervical	
BCL6	Gen de cáncer	Linfoma, leucemia	
BRCA1	Gen de cáncer	cáncer de mama, cáncer ovárico	
BRCA2	Gen de cáncer	Cáncer de mama, cáncer ovárico, cáncer pancreático	
CARD 11	Gen de cáncer	linfoma	
CBL	Gen de cáncer	Linfoma, leucemia	
CCND1	Gen de cáncer	Leucemia linfoblástica crónica (LLC), leucemia linfoblástica aguda de células B (LLA-B), cáncer de mama	
CCND2	Gen de cáncer	Retinoblastoma, linfoma de células del manto, leucemia linfoblástica aguda de células T (LLA-T), linfoma de Burkitt, tumor de células germinales testiculares, tumor de células granulosas ováricas, mieloma múltiple	
CCND3	Gen de cáncer	Retinoblastoma, linfoma de células del manto, linfoma anaplásico de células grandes, linfoma (no de Hodgkins),	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
		linfoma de células B, carcinoma de células escamosas laríngeo, linfoma indolente, adenoma de células nulas	
CCNE1	Gen de cáncer	Cáncer de mama, cáncer ovárico, cáncer de vejiga, retinoblastoma	
CDH1	Gen de cáncer	Cáncer gástrico, carcinoma lobular, carcinoma de células escamosas, carcinoma ductal invasivo, carcinoma lobular invasivo	
CDH2	Gen de cáncer	Melanoma, mesotelioma maligno, mesotelioma pleural, melanoma desmoplásico, adenocarcinoma pulmonar, tumor endometrioide, mesotelioma, cáncer de vejiga, carcinoma de células escamosas esofágico (CCCE)	
CDH20	Gen de cáncer	Cáncer de mama	
CDH5	Gen de cáncer	Granuloma, sarcoma epiteliode	
CDK4	Gen de cáncer	Melanoma	
CDK6	Gen de cáncer	Leucemia linfoblástica aguda (LLA)	
CDK8	Gen de cáncer	cáncer de colon, cáncer de pulmón, cáncer rectal, leucemia linfoblástica aguda (LLA)	
CDKN2B	Gen de cáncer	Leucemia, retinoblastoma, carcinoma de células escamosas laríngeo	
CDKN2C	Gen de cáncer	Carcinoma de tiroides, adenoma pituitario, oligodendroglioma, tumor endocrino pancreático, mieloma múltiple, hepatoblastoma, tumor linfoide, neoplasia endocrina múltiple de 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 de células germinales testicular, 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 paratiroides, 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 de cerebro, cáncer endometrial, cáncer ovárico	
EPHB6	Gen de cáncer	Neuroblastoma, melanoma, cáncer de pulmón de células no pequeñas (CPCNP)	
ERBB3	Gen de cáncer	Cáncer de mama, cáncer de pulmón de células no pequeñas (CPCNP), cáncer pancreático, carcinoma ductal invasivo, adenocarcinoma pulmonar, carcinoma endometrioide, astrocitoma pilocítico	
ERBB4	Gen de cáncer	Cáncer de mama, meduloblastoma, carcinoma de células escamosas cervical, 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 de células redondas pequeñas desmoplásico, liposarcoma mixoide, sarcoma de células claras	
ETV4	Gen de cáncer	Cáncer de mama, cáncer ovárico, carcinoma de células escamosas de lengua, sarcoma de Ewing	
ETV5	Gen de cáncer	Ganglioglioma, tumor cerebral	
ETV6	Gen de cáncer	Leucemia, fibrosarcoma congénito, carcinoma secretorio, síndrome mielodisplásico	
EWSR1	Gen de cáncer	Sarcoma de Ewing, sarcoma de células claras, tumor de células redondas pequeñas desmoplásico, condrosarcoma	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
		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 endometrial, leucemia linfoblástica aguda de células T (LLA-T)	
FGFR4	Gen de cáncer	Tumor pituitario, cáncer de próstata, cáncer de pulmón, astrocitoma, rhabdomyosarcoma, adenoma pituitario, fibroadenoma	
FLT1	Gen de cáncer	cáncer de mama, cáncer prostático	
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	Meloma uveal	
GNAS	Gen de cáncer	Adenoma pituitario	
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 uveal, carcinoma adrenocortical, cáncer pancreático	
IGF2R	Gen de cáncer	Tumor gastrointestinal, cáncer hepático	
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 granulosas, tumor estromal del cordón sexual, adenocarcinoma pulmonar, feocromocitoma, tumor de Krukenberg, cáncer ovárico	
IRS2	Gen de cáncer	Hiperinsulinemia, leiomyosarcoma uterino	
JAK1	Gen de cáncer	Leucemia, cáncer ovárico, cáncer de mama	
JAK3	Gen de cáncer	Leucemia linfoblástica aguda (LLA)	
JUN	Gen de cáncer	Cáncer de piel, leucemia	
KDR	Gen de cáncer	cáncer de pulmón de células no pequeñas (CPCNP), 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 pancreático, tumor intestinal	
MAP2K4	Gen de cáncer	Cáncer pancreático, 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	
MITM	Gen de cáncer	Melanoma	
MLH1	Gen de cáncer	cáncer colorrectal, cáncer endometrial, cáncer ovárico, 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 endometrial, cáncer ovárico	
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	cáncer de pulmón de células pequeñas (CPCP)	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
MYCN	Gen de cáncer	Neuroblastoma	
NF2	Gen de cáncer	Meningioma, neuroma acústico, cáncer renal	
NKX2-1	Gen de cáncer	cáncer de pulmón, cáncer de tiroides, adenocarcinoma	
NTRK1	Gen de cáncer	Carcinoma tiroideo papilar:	
NTRK3	Gen de cáncer	Fibrosarcoma congénito, cáncer de mama secretorio	
PAK3	Gen de cáncer	Cáncer de pulmón	
PAX5	Gen de cáncer	Linfoma no de Hodgkin (LNH), leucemia linfoblástica aguda (LLA, por ejemplo LLA de células B)	
PDGFRB	Gen de cáncer	trastorno mieloproliferativo (TMP), leucemia mieloide aguda (LMA), leucemia mieloide crónica (LMC), leucemia mielomonocítica crónica (LMMC)	
PIK3R1	Gen de cáncer	Glioblastoma, cáncer ovárico, cáncer colorrectal	
PKHD1	Gen de cáncer	Cáncer pancreático	
PLCG1	Gen de cáncer	Cáncer de cabeza y cuello, leucemia	
PRKDC	Gen de cáncer	Glioma, glioblastoma, cáncer gástrico, cáncer ovárico	
PTCH1	Gen de cáncer	Células basales de la piel, meduloblastoma	
PTPN11	Gen de cáncer	Leucemia mielomonocítica juvenil (LMMJ), leucemia mieloide aguda (LMA), síndromes mielodisplásicos (SMD)	
PTPRD	Gen de cáncer	cáncer de pulmón, carcinoma de células escamosas cutáneo, 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 prostático	
RUNX1	Gen de cáncer	leucemia mieloide aguda (LMA), leucemia linfoblástica aguda de células pre-B (LLA-preB), leucemia linfoblástica aguda de células T (LLA-T)	
SMAD2	Gen de cáncer	Carcinoma de células escamosas esofágico (CCEE)	
SMAD3	Gen de cáncer	Cáncer de piel, coriocarcinoma	
SMAD4	Gen de cáncer	cáncer pancreático, cáncer de colon	
SMARCA4	Gen de cáncer	cáncer de pulmón de células no pequeñas (CPCNP)	
SMARCB1	Gen de cáncer	Tumor rabdoide maligno	
SMO	Gen de cáncer	Cáncer de piel de células basales	
SOX10	Gen de cáncer	Oligodendroglioma	
SOX2	Gen de cáncer	Carcinoma embrionario, tumor de células germinales	
SRC	Gen de cáncer	sarcoma, cáncer de colon, cáncer de mama	
STK11	Gen de cáncer	cáncer de pulmón de células no pequeñas (CPCNP), cáncer pancreático	
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 mieloide 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	
VHL	Gen de cáncer	Cáncer renal, hemangioma, feocromocitoma	
WT1	Gen de cáncer	Tumor de Wilms, tumor de células redondas pequeñas desmoplásico	
ABCB1	Gen PGx		
ABCC2	Gen PGx		
ABCC4	Gen PGx		
ABCG2	Gen PGx		
Clorf144	Gen PGx		
CYP1B1	Gen PGx		
CYP2C19	Gen PGx		
CYP2C8	Gen PGx		
CYP2D6	Gen PGx		
CYP3A4	Gen PGx		
CYP3A5	Gen PGx		

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios
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 genes o productos génicos seleccionados

“Genes de cáncer” se refiere a genes o productos génicos asociados a cáncer de menor prioridad que Prioridad 1

“Genes Pgx” se refiere a genes que resultan importantes para la farmacogenética y farmacogenómica (PGx)

Tabla 1A. Genes y variantes ejemplificativos adicionales seleccionados, tipos de cáncer asociados, codones de prioridad, categoría de accionabilidad y terapias potenciales

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios	Categoría de accionabilidad	Motivo
ASXL1	Prioridad 1	Mieloma múltiple (MM)		D	Pronóstico (neg SMD)
BACH1	Prioridad 1	Mama		C	Inhibidores de la PARP
BAP1	Prioridad 1	Melanoma uveal, mama, CPCNP		C	Inhibidores de la PARP
BARD1	Prioridad 1	Mama		C	Inhibidores de la PARP
BLM	Prioridad 1	Leucemia, linfoma, células escamosas de la piel, otros cánceres		C	
BRIP1	Prioridad 1	Leucemia mieloide aguda (LMA), leucemia, mama		C	Inhibidores de la PARP
CDKN1B	Prioridad 1	Mama		D	
CREBBP	Prioridad 1	Leucemia linfoblástica aguda (LLA), LMA, LCBGD, linfoma no de Hodgkin de células B (LNH-B)		D	
DDR2	Prioridad 1	CPCNP		C	Dasatinib
EMSY	Prioridad 1	Mama		C	Inhibidores de la 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

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios	Categoría de accionabilidad	Motivo
FANCG	Prioridad 1	LMA, leucemia		C	Inhibidor de PARP
FANCL	Prioridad 1	LMA, leucemia		C	Inhibidor de PARP
HGF	Prioridad 1	MM		C	Resistencias
NFKB1	Prioridad 1	Mama		D	Posible MAL PRONÓSTICO
NOTCH2	Prioridad 1	Linfoma de zona marginal, LCBGD		D	-
PALB2	Prioridad 1	Tumor de Wilms, meduloblastoma, LMA, mama		C	Inhibidores de la PARP
PBRM1	Prioridad 1	Carcinoma renal de células claras, mama		E	Inhibidores de HDAC?
PDK1	Prioridad 1	CPCNP		C	Inhibidores de PDK1
PIK3R2	Prioridad 1	CPCNP		C	INHIBIDORES DE RUTA DE PI3K
RAD50	Prioridad 1	Mama		C	Inhibidores de la PARP
RAD51	Prioridad 1	Mama		C	Inhibidores de la PARP
ROS1	Prioridad 1	Glioblastoma, CPCNP		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	cáncer de pulmón de células no pequeñas (CPCNP), 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	CPCNP, mama		E	
CD79A	Gen de cáncer	linfoma de células B grandes difusas (LCBGD)		E	
CD79B	Gen de cáncer	LCBGD		E	
CDC73	Gen de cáncer	Paratiroides		E	
CDK12	Gen de cáncer	ovárico		E	
CHUK	Gen de cáncer	Colorrectal		E	
CRBN	Gen de cáncer	Tracto aerodigestivo superior		E	
CSF1R	Gen de cáncer	CPCNP		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, mama, pancreático, LMA, LLA, LCBGD		E	
ERCC2	Gen de cáncer	Células basales de la piel, células escamosas de la piel, melanoma		E	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios	Categoría de accionabilidad	Motivo
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 de células granulosas del ovario	134	E	
GATA2	Gen de cáncer	LMA, leucemia mieloide crónica (LMC, transformación blástica)		E	
GATA3	Gen de cáncer	Mama		E	
		LMA, síndrome mielodisplásico			
GRAF	Gen de cáncer	(SMD)		E	
GRIN2A	Gen de cáncer	melanoma maligno		E	
GSK3B	Gen de cáncer	CPCNP		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 linfoblástica aguda de células T (LLA-T)		E	
INSR	Gen de cáncer	CPCNP, glioblastoma, gástrico		E	
IRF4	Gen de cáncer	Mieloma múltiple (MM)		E	
KDM4C	Gen de cáncer	Ovárico, mama		E	
KDM5A	Gen de cáncer	LMA		E	
KDM6A	Gen de cáncer	Renal, carcinoma de células escamosas esofágico (CCE), MM		E	
KEAP1	Gen de cáncer	CPCNP		E	
KLHL6	Gen de cáncer	Leucemia linfocítica crónica (LLC)		E	
LMO1	Gen de cáncer	Leucemia linfoblástica aguda de células T (LLA-T), neuroblastoma		E	
LRP6	Gen de cáncer	CPCNP, melanoma maligno		E	
LRRK2	Gen de cáncer	Ovárico, CPCNP		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	LCBGD de tipo células B activadas (LCBGD-CBA)		E	
MYST3	Gen de cáncer	Mama		E	
NCOR1	Gen de cáncer	Mama		E	
NFE2L2	Gen de cáncer	CPCNP, carcinoma de células escamosas de cabeza y cuello (CCECC)		E	
NFKBIA	Gen de cáncer	Mama		E	
NOTCH3	Gen de cáncer	CPCNP, mama		E	
NOTCH4	Gen de cáncer	CPCNP, mama		E	
NSD1	Gen de cáncer	LMA		E	
NTRK2	Gen de cáncer	Renal, CPCNP		E	
NUP93	Gen de cáncer	Mama		E	

Gen Hugo	Categoría génica	Tipos de cáncer	Codones prioritarios	Categoría de accionabilidad	Motivo
PAK7	Gen de cáncer	CPCNP, melanoma maligno		E	
PHLPP2	Gen de cáncer	Ovário, glioblastoma, CPCNP		E	
PHOX2B	Gen de cáncer	Neuroblastoma		E	
PIK3C2G	Gen de cáncer	CPCNP		E	
PIK3C3	Gen de cáncer	CPCNP		E	
PIK3CG	Gen de cáncer	CPCNP		E	
PNRC1	Gen de cáncer	MM		E	
PRDM1	Gen de cáncer	LCBGD		E	
PRKAR1A	Gen de cáncer	Glándula adrenal, tiroides		E	
PRSS8	Gen de cáncer	Mama		E	
PTCH2	Gen de cáncer	melanoma maligno		E	
PTK2	Gen de cáncer	CPCNP, glioblastoma		E	
PTK2B	Gen de cáncer	CPCNP, mama		E	
REL	Gen de cáncer	Linfoma de Hodgkin		E	
RHEB	Gen de cáncer	CPCNP, colorrectal		E	
ROCK1	Gen de cáncer	Mama		E	
RUNXT1	Gen de cáncer	CPCNP, colorrectal		E	
SETD2	Gen de cáncer	Carcinoma renal de células claras		E	
SH2B3	Gen de cáncer	Síndrome mielodisplásico (SMD)		E	
SOCS1	Gen de cáncer	LCBGD		E	
SPEN	Gen de cáncer	Carcinoma quístico adenoide		E	
STAG2	Gen de cáncer	Glioblastoma		E	
STAT3	Gen de cáncer	Mama		E	
STAT4	Gen de cáncer	Mama		E	
STK12	Gen de cáncer	TNEP, CPCNP		E	
SUFU	Gen de cáncer	Meduloblastoma		E	
TBX23	Gen de cáncer	Mama		E	
TBX3	Gen de cáncer	Mama		E	
TNFAIP3	Gen de cáncer	Linfoma de células B de la zona marginal, linfoma de Hodgkin, linfoma de células B mediastinal primario		E	
TNFRSF14	Gen de cáncer	Linfoma folicular		E	
TNFRSF17	Gen de cáncer	Linfoma intestinal de células T		E	
TNKS	Gen de cáncer	CPCNP		E	
TNKS2	Gen de cáncer	Melanoma, mama		E	
TRRAP	Gen de cáncer	Colorrectal, glioblastoma		E	
TYK2	Gen de cáncer	CPCNP, 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 accionabilidad se clasifican tal como se indica posteriormente. La tabla 1B proporciona un resumen de la aplicación de las diferentes categorías a alteraciones ejemplificativas en diferentes tipos de cáncer.

- 5 Categoría A: alteraciones aprobadas / estándares que predicen sensibilidad o resistencia a terapias aprobadas / estándares *KRAS* G13D en cáncer de colon metastásico  
amplificación de *ERBB2* en cáncer de mama  
*EGFR* L858R en cáncer de pulmón de células no pequeñas
- 10 Categoría B: alteraciones que son criterios de inclusión o exclusión para terapias experimentales específicas  
*KRAS* G13D en cáncer de colon, cáncer de pulmón o cáncer de mama  
*BRAF* V600E en melanoma, cáncer de colon o cáncer de pulmón  
*NRAS* Q61K en melanoma  
*PIK3CA* H1047R en cáncer de mama

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 pancreático

- 5 Categoría C: alteraciones con pruebas limitadas (datos clínicos tempranos, datos clínicos conflictivos, datos preclínicos, teórico) que predicen sensibilidad o resistencia a terapias estándares o experimentales  
*KRAS* Q61H en cáncer de colon (clínico temprano)  
*PIK3CA* H1047R en cáncer de mama (clínica conflictiva)  
*BRAF* V600E en cáncer de colon (clínica conflictiva)
- 10 Mutación o amplificación de *ERBB2* en cáncer de colon (informes de caso)  
*BRAF* D594G en cáncer de pulmón (preclínico)  
 Amplificación de *FGFR1* en cáncer de mama (preclínico)  
 Inactivación bialélica de *ATM* en cáncer de mama (preclínico)  
 inactivación bialélica de *TSC1* en cáncer de colon (preclínico)
- 15 Inactivación bialélica de *ATR* en cáncer de mama (teórica)  
 Mutación *BRAF* V600E en sarcoma (teórica)
- Categoría D: alteraciones con utilidad pronóstica o diagnóstica en un subtipo particular de cáncer  
 inactivación bialélica de *MSH2* en cáncer de colon (pruebas clínicas fuertes)
- 20 *BRAF* V600E en cáncer de colon (pruebas clínicas fuertes)  
*KRAS* G13D en cáncer de pulmón (pruebas clínicas fuertes)  
 inactivación de *BRCA1* en cáncer de mama (pruebas clínicas fuertes)
- Categoría E: alteraciones con claro significado biológico en el cáncer (es decir, mutaciones conductoras) sin claras implicaciones clínicas
- 25 Inactivación bialélica de *APC* en cáncer de colon  
 Inactivación bialélica de *TP53* en cáncer de mama  
 Amplificación de *MITM* en melanoma  
*ARID1A* en cáncer ovárico
- 30 Categoría F: alteraciones sin significado biológico conocido en el cáncer  
 Nuevas alteraciones en genes de cáncer conocidos  
 Dianas terapéuticas  
 ortólogos de genes de cáncer conocidos
- 35

Tabla 1B. Clasificación ejemplificativa 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 1858R	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 pancreático		x	x	x	x
Inactivación bialélica de ATM M	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 renal				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 2. Genes seleccionados ejemplificativos asociados a farmacogenética y farmacogenómica (PGx).

Locus	génico	Mutación	Efecto
ABCB1	chr7:86976581	3853C>T	Mejor supervivencia en LMA asiática tratada con Ida/AraC; supervivencia en pacientes de cáncer de mama tratados con paclitaxel
ABCB1	chr7:86998554	2677G>T/A	Respuesta a taxanos, platinos y toxicidad GI; mejor supervivencia en LMA asiática tratada con Ida/AraC
ABCC2	chr10:101610761		Leucopenia inducida por docetaxel
ABCC4	chr13:94613416		Toxicidad de 6MP
ABCG2	chr4:89252551		MTX
ABCG2	chr4:89271347	q141K	Diarrea después de gefitinib
ABCG2	chr4:89274403		MTX
C1orf144	chr1:16578662		toxicidad de daunorrubicina
CYP1B1	chr2:38151707	CYP1B1*3	toxicidad de daunorrubicina; supervivencia en pacientes de cáncer de mama tratados con paclitaxel
CYP2C19	chr10:96509051	CYP2C19*17	Beneficio mejorado por tamoxifeno
CYP2C19	chr10:96511647	CYP2C19*17	Beneficio mejorado por tamoxifeno
CYP2C8	chr10:96786964	461delV	Metabolización del paclitaxel
CYP2C8	chr10:96788739	K399R	Metabolización del paclitaxel
CYP2C8	chr10:96808096		Metabolización del paclitaxel
CYP2C8	chr10:96808109		Metabolización del paclitaxel
CYP2C8	chr10:96817020		Metabolización del paclitaxel
CYP2D6	chr22:40853554	CYP2D6: 3183 G>A	CYP2D6*29, presente en tanzanos
CYP2D6	chr22:40853749	CYP2D6: 2988 G>A	CYP2D6*41 (IM)
CYP2D6	chr22:40853887	CYP2D6: 2850 C>T	CYP2D6*2 (EM)
CYP2D6	chr22:40854122	CYP2D6: 2613-2615 del AGA	CYP2D6*9 (función no clara?)
CYP2D6	chr22:40854188	CYP2D6: 2549 del A	CYP2D6*3
CYP2D6	chr22:40854891	CYP2D6: 1846 G>A	CYP2D6*4
CYP2D6	chr22:40855030	CYP2D6: 1707 del T	CYP2D6*6
CYP2D6	chr22:40855078	CYP2D6: 1659G>A	CYP2D6*29, presente en tanzanos
CYP2D6	chr22:40855716	CYP2D6: 1023 C>T	Presente en CYP2D6*17
CYP2D6	chr22:40856638	CYP2D6: 100C>T	Presente en CYP2D6*10 (causativo) y *4 (asociado)
CYP3A4	chr7:99196395		
CYP3A4	chr7:99196460		
CYP3A4	chr7:99197606		
CYP3A4	chr7:99204017		
CYP3A4	chr7:99204029	CYP3A4*16B	Metabolización del paclitaxel en japoneses
CYP3A4	chr7:99205328		
CYP3A4	chr7:99205363		
CYP3A4	chr7:99219597		
CYP3A4	chr7:99220032	CYP3A4* 1B	mejor lavado del docetaxel
CYP3A5	chr7:99088330		
CYP3A5	chr7:99100771		
CYP3A5	chr7:99108475		
DPYD	chr1:97688202	DPYD*2A	Toxicidad de 5FU
DPYD	chr1:97753983	DPYD*5	Toxicidad de 5FU
DPYD	chr1:97937679	496A>G	5FU, toxicidad de Xeloda
DPYD	chr1:98121473	DPYD*9A	Toxicidad de 5FU
ERCC2	chr19:50546759	2251A>C	Recaída tras 5FU en asiáticos
ESR1	chr6:152205074		Hipercolesterolemia inducida por tamoxifeno
ESR2	chr14:63769569		Hipercolesterolemia inducida por tamoxifeno
FCGR3A	chr1:159781166	V158F	Respuesta al cetuximab
FGFR4	chr5:176452849	GLY388ARG	

Locus	génico	Mutación	Efecto
GSTP1.	chr11:67109265	1105V	Resistencia a múltiples quimioterapias
GSTP1	chr11:67110155	A114V	No claro, desequilibrio de ligamiento con I105V
ITPA	chr20:3141842		Toxicidad de 6MP
LRP2	chr2:169719231		Asociado a ototoxicidad del cisplatino
MAN1B1	chr9:139102689		toxicidad de daunorrubicina
MTHFR	chr1:11777044		MTX
MTHFR	chr1:11777063		MTX
MTHFR	chr1:11778965	677C>T	MTX
NQO1	chr16:68302646	NQO1*2	Degradación rápida (cisplatino, doxorubicina); baja supervivencia en cáncer de mama tratado con antraciclinas
NRP2	chr2:206360545		toxicidad de daunorrubicina
SLC19A1	chr21:45782222		MTX
SLC22A2	chr6:160590272	Ala270Ser	Nefrotoxicidad reducida del cisplatino
SLCO1B3	chr12:20936961		Leucopenia inducida por docetaxel
SOD2	chr6:160033862	V16A	Supervivencia inferior en cáncer de mama tratado con ciclofosfamida
SULT1A1	chr16:28524986		
SULT1A1	chr16:28525015		
SULT1A1	chr16:28528073		
SULT1A1	chr16:28528301		
TMPT	chr6:18247207	TPMT*3B	Toxicidad de purinas
TPMT	chr6:18238897		Toxicidad de 6MP
TPMT	chr6:18238991		Toxicidad de 6MP
TPMT	chr6:18251934		Toxicidad de 6MP
TYMS	chr18:647646	Repetición en tándem de 28 pb	Toxicidad de 5FU
TYMS	chr18:663451	Delección de 6 pb	Toxicidad de 5FU
UGT1A1	chr2:234255266		Anemia por irinotecán
UGT1A1	chr2:234255709		Trombocitopenia por irinotecán
UGT1A1	chr2:234330398		UGT1A1*60
UGT1A1	chr2:234330521		UGT1A1*93
UGT1A1	chr2:234333620		UGT1A1*28
UGT1A1	chr2:234333883		UGT1A1*6
UGT1A1	chr2:234334358		UGT1A1*27
UMPS	chr3:125939432	Gly213Ala	Toxicidad de 5FU

Tabla 3. Genes seleccionados ejemplificativos asociados a mutaciones por traslocación en tumores sólidos

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
ACSL3	Prioridad 1	ETV1	próstata
ALK	Prioridad 1	NPM1, TPM3, TMG, TPM4, ATIC, CLTC, MSN, ALO17, CARS, EML4	LCGA, CPNCP, Neuroblastoma
BRAF	Prioridad 1	AKAP9, KIAA1549	melanoma, colorrectal, tiroideo papilar, ovárico 'borderline', cáncer de pulmón de células no pequeñas (CPCNP), 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	CPCNP
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

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
ETV5	Prioridad 1	TMPRSS2, SCL45A3	Próstata
FGFR3	Prioridad 1	IGH@, ETV6	Vejiga, MM, linfoma de células 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	Tiroide medular, tiroide papilar, feocromocitoma
ROS1	Prioridad 1	GOPC, ROS1	Glioblastoma, CPCNP
SLC45A3	Prioridad 1	ETV1, ETV5, ELK4, ERG	próstata
TMPRSS2	Prioridad 1	ERG, ETV1, ETV4, ETV5	próstata
AKAP9		BRAF	tiroideo papilar
ASPSCR1		TME3	Sarcoma de tejidos blandos alveolar
ATM1		EWSR1, FUS	Melanoma maligno de tejidos blandos, histiocitoma fibroso angiomatoide
BRD3		NUT	Carcinoma de línea media fatal juvenil
BRD4		NUT	Carcinoma de línea media fatal juvenil
C12orf9		LPP	lipoma
CD74		ROS1	CPCNP
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	Leiomioma uterino
CREB1		EWSR1.	Sarcoma de células claras, histiocitoma fibroso angiomatoide
CREB3L2		FUS	Sarcoma fibromixioide
CRTC3		MAML2.	Mucoepidermoide de glándulas salivales
CTNNB1		PLAG1	Colorrectal, ovárico, hepatoblastoma, otros, adenoma salival pleomórfico
D10S170		RET, PDGFRB	papilar tiroideo, LMC
DDIT3		FUS	Liposarcoma
DUX4		CIC	Sarcoma de tejidos blandos
ELKS		RET	tiroideo papilar
ETV6		NTRK3, RUNX1, PDGFRB, ABL1, MN1, ABL2, FACL6, CHIC2, ARNT, JAK2, EVI1, CDX2, STL, HLXB9, MDS2, PER1, SYK, TTL, FGFR3, PAX5	Fibrosarcoma congénito, leucemia múltiple y linfoma, de mama secretorio, SMD, LLA
EWSR1		FLI1, ERG, ZNF278, NR4A3, FEV, ATM1, ETV1, ETV4, WT1, ZNF384, CREB1, POU5F1, PBX1	Sarcoma de Ewing, tumor de células redondas pequeñas desmoplásico, LLA, sarcoma de células claras, sarcoma, mioepitelioma
FEV		EWSR1, FUS	Sarcoma de Ewing
FLI1		EWSR1	Sarcoma de Ewing
FOXO1A		PAX3	rabdomiosarcomas alveolares
FUS		DDIT3, ERG, FEV, ATM1, CREB3L2	Liposarcoma, LMA, sarcoma de Ewing, histiocitoma fibroso angiomatoide, sarcoma fibromixioide
GOLGA5		RET	tiroideo papilar
HEI10		HMGA2	Leiomioma uterino
HMGA1		?	Adenoma tiroideo microfolicular, diversos tumores mesenquimales benignos
HMGA2		LHFP, RAD51L1, LPP, HEI10, COX6C, CMKOR1, NFIB	
HOOK3		RET	tiroideo papilar
JAZF1		SUZ12	Tumores estromales endometriales
KTN1		RET	tiroideo papilar
LHFP		HMGA2	

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
LIFR		PLAG1	Adenoma salival
LPP		HMGA2, MLL, C12orf9	Lipoma, leucemia
MAML2		MECT1, CRTC3	Mucoepidermoide de glándulas salivales
MECT1		MAML2.	Mucoepidermoide de glándulas salivales
MN1		ETV6	LMA, meningioma
MYB		NFIB	carcinoma quístico adenoide
MYC		IGK, BCL5, BCL7A, BTG1, TRA, IGH	linfoma de Burkitt, amplificado en otros cánceres, LLC-B
NCOA1		PAX3	rabdomiosarcoma alveolar
NCOA4		RET	tiroideo papilar
NFIB		MYB, HGMA2	carcinoma quístico adenoide
NONO		TME3	Cáncer renal papilar
NR4A3		EWSR1	Condrosarcoma mixoide extraesquelético
NTRK1		TPM3, TPR, TMG	tiroideo papilar
NTRK3		ETV6	fibrosarcoma congénito, cáncer secretor de mama
NUT		BRD4, BRD3	Carcinoma de línea media fatal juvenil
OMD		USP6	Quistes óseos aneurismáticos
PAX3		FOXO1A, NCOA1	rabdomiosarcoma alveolar
PAX7		FOXO1A.	rabdomiosarcoma alveolar
PAX8		PPARG	Tiroideo folicular
PBX1		TCF3, EWSR1	LLA-preB, mioepitelioma
PCM1		RET, JAK2	Tiroideo papilar, LMC
PDGFB		COL1A1	DFSP
PDGFRA		FIP1L1	GIST, síndrome hipereosinofílico idiopático
PLAG1		TCEA1, LIFR, CTNNB1, CHCHD7	Adenoma salival
POU5F1		EWSR1.	sarcoma
PPARG		PAX8	Tiroideo folicular
PRCC		TME3	renal papilar
PRKAR1A		RET	tiroideo papilar
PRO1073		TMEB	carcinoma de células renales (epitelioide infantil)
RAD51L1		HMGA2	lipoma, leiomioma uterino
RAF1		SRGAP3.	astrocitoma pilocítico
SFPQ		TME3	cáncer renal papilar
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 estromales endometriales
TAF15		TEC, CHN1, ZNF384	condrosarcomas mixoides extraesqueléticos, LLA
TCEA1		PLAG1	Adenoma salival
TCF12		TEC	Condrosarcoma mixoide extraesquelético
TME3		SFPQ, ASPSCR1, PRCC, NONO, CLTC	renal papilar, sarcoma de tejidos blandos alveolar, renal
TMEB		ALPHA	renal (epitelioide infantil)
TMG		NTRK1, ALK	tiroideo papilar, LCGA, CPCNP
THRAP3		USP6	Quistes óseos aneurismáticos
TPM3		NTRK1, ALK	papilar tiroideo, LCGA
TPR		NTRK1	tiroideo papilar
TRIM27		RET	tiroideo papilar
TRIM33		RET	tiroideo papilar
USP6		COL1A1, CDH11, ZNF9, OMD	Quistes óseos aneurismáticos
ZNF278		EWSR1	Sarcoma de Ewing
ZNF331		?	adenoma tiroideo folicular
ZNF9		USP6	quistes óseos aneurismáticos

Tabla 4. Genes seleccionados ejemplificativos asociados a mutaciones por traslocación en neoplasias malignas hematológicas.

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
ABL1	Prioridad 1	BCR, ETV6, NUP214	LLC, LLA, LLA-T
ALK	Prioridad 1	NPM1, TPM3, TMG, TPM4, ATIC, CLTC, MSN, ALO17, CARS, EML4	LCGA, CPNCP, Neuroblastoma
BCL2	Prioridad 1	IGH	LNH, LLC
BCL6	Prioridad 1	Loci IG, ZNFN1A1, LCP1, PIM1, TMRC, 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	AL, LMA
FGFR1	Prioridad 1	BCR, FOP, ZNF198, CEP1	EMP, LNH
FGFR3	Prioridad 1	IGH, ETV6	Vejiga, MM, linfoma de células T
JAK2	Prioridad 1	ETV6, PCM1, BCR	LLA, LMA, EMP, LMC
MLL	Prioridad 1	MLL, MLLT1, MLLT2, MLLT3, MLLT4, MLLT7, MLLT10, MLLT6, ELL, EPS15, AF1Q, CREBBP, SH3GL1, FBNP1, PNU TL1, MSF, GPHN, GMPS, SSH3BP1, ARHGEF12, GAS7, FOXO3A, LAF4, LCX, SEPT6, LPP, CBFA2T1, GRAF, EP300, PICALM, HEAB	AML, ALL
PDGFRA	Prioridad 1	FIP1L1	GIST, síndrome hipereosinófilo idiopático
RARA	Prioridad 1	PML, ZNF145, TIF1, NUMA1, NPM1	APL
SEPT6		MLL	AML
ABL2		ETV6.	AML
AF15Q14		MLL	AML
AF1Q		MLL	LLA
AF3p21		MLL	LLA
AF5q31		MLL	LLA
ALO17		ALK	LCGA
ARHGEF12		MLL	AML
ARHH		BCL6	LNH
ARNT		ETV6	LMA
ATIC		ALK	LCGA
BCL10		IGH	MALT
BCL11A		IGH	LLC-B
BCL11B		TLX3.	LLA-T
BCL3		IGH	LLC
BCL5		MYC	LLC
BCL7A		MYC	BNHL
BCL9		IGH, IGL	LLA-B
BCR		ABL1, FGFR1, JAK2	LMC, LLA, LMA
BIRC3		MALT1	MALT
BTG1		MYC	LLCB
CARS		ALK	LCGA
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	EMP, LNH
CHIC2		ETV6	LMA
CLTC		ALK, TME3	LCGA, renal
CLTCL1		?	LCGA
DDX10		NUP98	LMA*

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
DDX6		IGH	LNH-B
DEK		NUP214	LMA
EIF4A2		BCL6	LNH
ELF4		ERG	LMA
ELL		MLL	AL
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, PER1, SYK, TTL, FGFR3, PAX5	Fibrosarcoma congénito, leucemia múltiple y linfoma, de mama secretorio, SMD, LLA
EVI1		RUNX1, ETV6, PRDM16, RPN1	LMA, LMC
EWSR1		FLI1, ERG, ZNF278, NR4A3, FEV, ATM1, ETV1, ETV4, WT1, ZNF384, CREB1, POU5F1, PBX1	Sarcoma de Ewing, tumor de células redondas pequeñas desmoplásico, LLA, sarcoma de células claras, sarcoma, mioepitelioma
FACL6		ETV6	LMA, LLA
FCGR2B		?	LLA
FGFR1OP		FGFR1	EMP, LNH
FIP1L1		PDGFRA	síndrome hipereosinófilo idiopático
FNBP1		MLL	LMA
FOXO3A		MLL	AL
FOXP1		PAX5	LLA
FSTL3		CCND1	LLC-B
FUS		DDIT3, ERG, FEV, ATM1, CREB3L2	Liposarcoma, LMA, sarcoma de Ewing, histiocitoma fibroso angiomatoide, sarcoma fibromixioide
FVT1		IGK	LNH-B
GAS7		MLL	LMA*
GMPS		MLL	LMA
GPHN		MLL	AL
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, LCGME
IGK		MYC, FVT1	linfoma de Burkitt, LNH-B
IGL		BCL9, MYC, CCND2	Linfoma de Burkitt
IL2		TNFRSF17	Linfoma intestinal de células T

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
IL21R		BCL6	LNH
IRF4		IGH	MM
IRTA1		IGH	LNH-B
ITK		SYK	Linfoma periférico de células 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	MALT
MDS1		RUNX1	SMD, LMA
MDS2		ETV6	SMD
MHC2TA		BCL6	LNH
MKL1		RBM15	Leucemia megacariocítica aguda
MLF1		NPM1	LMA
MLLT1		MLL	AL
MLLT10		MLL, PICALM, CDK6	AL
MLLT2		MLL	AL
MLLT3		MLL	LLA
MLLT4		MLL	AL
MLLT6		MLL	AL
MLLT7		MLL	AL
MN1		ETV6	LMA, meningioma
MSF		MLL	LMA*
MSI2		HOXA9	LMC
MSN		ALK	LCGA
MTCP1		TRA	Leucemia prolinfocítica de células 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	LCGA
MYST4		CREBBP	LMA
NACA		BCL6	LNH
NCOA2		RUNXBP2.	LMA
NFKB2		IGH	LNH-B
NIN		PDGFRB	EMP
NOTCH1		TRB	LLA-T
NPM1		ALK, RARA, MLF1	LNH, LPA, LMA
NSD1		NUP98	LMA
NUMA1		RARA	APL
NUP214		DEK, SET, ABL1	LMA, LLA-T
NUP98		HOXA9, NSD1, WHSC1L1, DDX10, TOP1, HOXD13, PMX1, HOXA13, HOXD11, HOXA11, RAP1GDS1, HOXC11	LMA
OLIG2		TRA	LLA-T
PAFAH1B2		IGH	LCGME
PAX5		IGH, ETV6, PML, FOXP1, ZNF521, ELN	LNH, LLA, LLA-B
PBX1		TCF3, EWSR1	LLA-preB, mioepitelioma
PCM1		RET, JAK2	Tiroideo papilar, LMC
PCSK7		IGH	LCGME
PDE4DIP		PDGFRB	EMP

Gen Hugo	Categoría génica	Pareja de traslocación	Tipos de cáncer
PDGFRB		ETV6, TRIP11, HIP1, RAB5EP, H4, NIN, HCMOGT-1, PDE4DIP	EMP, LMA, LMMC, LMC
PER1		ETV6	LMA, LMMC
PICALM		MLLT10, MLL	LLAT, 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	AL
SIL		TAL1	LLA-T
SSH3BP1		MLL	LMA
STL		ETV6	LLA-B
SYK		ETV6, ITK	SMD, linfoma periférico de células T
TAF15		TEC, CHN1, ZNF384	condrosarcomas mixoides extraesqueléticos, LLA
TAL1		TRD, SIL	leucemia linfoblástica/bifásica
TAL2		TRB	LLA-T
TCF3		PBX1, HLF, TMPT	LLA-preB
TCL1A		TRA	LLC-T
TCL6		TRA	LLA-T
TMG		NTRK1, ALK	tiroideo papilar, LCGA, CPCNP
TMPT		TCF3	LLA-preB
TMRC		BCL6	LNH
TIF1		RARA	APL
TLX1		TRB, TRD	LLA-T
TLX3		BCL11B.	LLA-T
TNFRSF17		IL2.	Linfoma intestinal de células T
TOP1		NUP98	LMA*
TPM3		NTRK1, ALK	papilar tiroideo, LCGA
TPM4		ALK	LCGA
TRA		ATL,OLIG2, MYC, TCL1A, TCL6, MTCP1, TCL6	LLA-T
TRB		HOX11, LCK, NOTCH1, TAL2, LYL1	LLA-T
TRD		TAL1, HOX11, TLX1, LMO1, LM02, RANBP17	Leucemia de células T
TRIP11		PDGFRB	LMA
TTL		ETV6	LLA
WHSC1		IGH	MM
WHSC1L1		NUP98	LMA
ZNF145		RARA	APL
ZNF198		FGFR1	EMP, LNH
ZNF384		EWSR1, TAF15	LLA
ZNF521		PAX5	LLA
ZNFN1A1		BCL6	LCBGD

**Ejemplo G. Secuencias de señuelo ejemplificativas para la captura de híbridos**

La tabla 7 proporciona señuelos ejemplificativos para tres dianas: SMAD3\_diana\_10, SMAD3\_diana\_11, SMAD3\_diana\_12

Tabla 7. Señuelos ejemplificativos

1. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477013-67477132

CCATTGTGTGTGAGCAAAGGCACCCTGTCCAGTCTAACCTGAATCTCTGTA  
GGAAGAGGCGTGCGGCTCTACTACATCGGAGGGGAGGTCTTCGCAGAGTGCCTC  
AGTGACAGCGCTATT (SEC ID nº 37)

(ID de señuelo: SMAD3\_diana\_10.2)

2. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477037-67477156

CTGTCCAGTCTAACCTGAATCTCTGTAGGAAGAGGCGTGCGGCTCTACTAC  
ATCGGAGGGGAGGTCTTCGCAGAGTGCCTCAGTGACAGCGCTATTTTGTCCAGT  
CTCCCAACTGTAAC (SEC ID nº 38)

(ID de señuelo: SMAD3\_diana\_10.4)

3. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477061-67477180

GTAGGAAGAGGCGTGCGGCTCTACTACATCGGAGGGGAGGTCTTCGCAGA  
GTGCCTCAGTGACAGCGCTATTTTGTCCAGTCTCCCAACTGTAACCAGCGCTAT  
GGCTGGCACCCGGCC (SEC ID nº 39)

(ID de señuelo: SMAD3\_diana\_10.6)

4. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477085-67477204

TACATCGGAGGGGAGGTCTTCGCAGAGTGCCTCAGTGACAGCGCTATTTT  
TGTCCAGTCTCCCAACTGTAACCAGCGCTATGGCTGGCACCCGGCCACCGTCTGC  
AAGATCCCACCAGGT (SEC ID nº 40)

(ID de señuelo: SMAD3\_diana\_10.1)

5. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477109-67477228

GAGTGCCTCAGTGACAGCGCTATTTTGTCCAGTCTCCCAACTGTAACCAG  
CGCTATGGCTGGCACCCGGCCACCGTCTGCAAGATCCCACCAGGTAAACGAGCC  
GCACAGGCACCCCTG (SEC ID nº 41)

(ID de señuelo: SMAD3\_diana\_10.5)

6. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_10	chr15:67477133-67477252

TTTGTCCAGTCTCCCAACTGTAACCAGCGCTATGGCTGGCACCCGGCCACC  
GTCTGCAAGATCCCACCAGGTAAACGAGCCGCACAGGCACCCCTGCCTTGAGGT  
CCCTCTCCGAGTGCA (SEC ID nº 42)

(ID de señuelo: SMAD3\_diana\_10.3)

7. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479655-67479774

GACCTGGCCACTTCCATCCCCACAGCCCTGTTTCTGTGTTTTTGGCAGGAT  
GCAACCTGAAGATCTTCAACAACCAGGAGTTCGCTGCCCTCCTGGCCCAGTCGGT  
CAACCAGGGCTTTG (SEC ID nº 43)

(ID de señuelo: SMAD3\_diana\_11.1)

8. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479679-67479798

GCCCTGTTTCTGTGTTTTTGGCAGGATGCAACCTGAAGATCTTCAACAACC  
AGGAGTTCGCTGCCCTCCTGGCCCAGTCGGTCAACCAGGGCTTTGAGGCTGTCTA  
CCAGTTGACCCGAA (SEC ID nº 44)

(ID de señuelo: SMAD3\_diana\_11.5)

9. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479703-67479822

GATGCAACCTGAAGATCTTCAACAACCAGGAGTTCGCTGCCCTCCTGGCC  
CAGTCGGTCAACCAGGGCTTTGAGGCTGTCTACCAGTTGACCCGAATGTGCACCA  
TCCGCATGAGCTTCG (SEC ID nº 45)

(ID de señuelo: SMAD3\_diana\_11.3)

10. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479727-67479846

ACCAGGAGTTCGCTGCCCTCCTGGCCCAGTCGGTCAACCAGGGCTTTGAG  
GCTGTCTACCAGTTGACCCGAATGTGCACCATCCGCATGAGCTTCGTCAAAGGCT  
GGGGAGCGGAGTACA (SEC ID nº 46)

(ID de señuelo: SMAD3\_diana\_11.4)

11. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479751-67479870

CCCAGTCGGTCAACCAGGGCTTTGAGGCTGTCTACCAGTTGACCCGAATG  
TGCACCATCCGCATGAGCTTCGTCAAAGGCTGGGGAGCGGAGTACAGGTCAGTT  
ATGGGTGCTGCCTACA (SEC ID nº 46)

(ID de señuelo: SMAD3\_diana\_11.2)

12. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_11.	chr15:67479775-67479894

AGGCTGTCTACCAGTTGACCCGAATGTGCACCATCCGCATGAGCTTCGTCA  
AAGGCTGGGGAGCGGAGTACAGGTCAGTTATGGGTGCTGCCTACATCAGGGGAC  
CCAACTCCAGGTGAC (SEC ID nº 48)

(ID de señuelo: SMAD3\_diana\_11.6)

13. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482692-67482811

TGTAACCCCTGGAGATTTTTTAAGTCCCCACCCACCCCTTCCCTATTT  
CTTACAGGAGACAGACTGTGACCAGTACCCCTGCTGGATTGAGCTGCACCTGAA  
TGGGCCTTTCAG (SEC ID nº 49)

(ID de señuelo: SMAD3\_diana\_12.5)

14. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482716-67482835

GTCCCCACCCACCCCTTCCCTATTTCTTACAGGAGACAGACTGTGACC  
AGTACCCCTGCTGGATTGAGCTGCACCTGAATGGGCCTTTCAGTGGCTTGACA  
AGGTCCTCACCCAG (SEC ID nº 50)

(ID de señuelo: SMAD3\_diana\_12.3)

15. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482740-67482859

ATTTCTTACAGGAGACAGACTGTGACCAGTACCCCTGCTGGATTGAGCT  
GCACCTGAATGGGCCTTTCAGTGGCTTGACAAGGTCCTCACCCAGATGGGCTCC  
CCAAGCATCCGCTGT (SEC ID nº 51)

(ID de señuelo: SMAD3\_diana\_12.2)

16. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482764-67482883

ACCAGTACCCCTGCTGGATTGAGCTGCACCTGAATGGGCCTTTCAGTG  
GCTTGACAAGGTCCTCACCCAGATGGGCTCCCCAAGCATCCGCTGTTCCAGTGTG  
TCTTAGAGACATCAA (SEC ID nº 52)

(ID de señuelo: SMAD3\_diana\_12.4)

17. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482788-67482907

CTGCACCTGAATGGGCCTTTCAGTGGCTTGACAAGGTCCTCACCCAGAT  
GGGCTCCCCAAGCATCCGCTGTTCCAGTGTGTCTTAGAGACATCAAGTATGGTAG  
GGGAGGGCAGGCTTG (SEC ID nº 53)

(ID de señuelo: SMAD3\_diana\_12.6)

18. Diana	génica	Localización genómica del señuelo
SMAD3	SMAD3_diana_12.	chr15:67482812-67482931

TGGCTTGACAAGGTCCTCACCCAGATGGGCTCCCCAAGCATCCGCTGTTCC  
AGTGTGTCTTAGAGACATCAAGTATGGTAGGGGAGGGCAGGCTTGGGGAAAATG  
GCCATGCAGGAGGTG (SEC ID nº 54)

(ID de señuelo: SMAD3\_diana\_12.1)

- 5 La tabla 8 proporciona señuelos con secuencias para dos dianas: FLT3\_diana\_24 modificado para reducir la estructura secundaria. FLT4\_diana\_31 presenta algunas secuencias arbitrarias en ambos extremos de los señuelos que es efectivamente similar a un señuelo más corto. Ambos mejoran la cobertura en aproximadamente 4x (mejora ~4x de la cobertura).

10 Tabla 8: señuelos ejemplificativos

1. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674626-28674745

Secuencia original

CGTCGCGCGCCAACGCCGGCATGGCCTCCGGAGCCCCGGGGTCCCCAGGCC  
GCGCCGGCCCCAGCCCTGCGATGCCGCCTGGAGCGGCGCGCCTCGCGCTGCAGGT  
GGCTCTCTTAAGGATG (SEC ID nº 55)

Secuencia modificada

CGTCTCACGCCAACGCAAGCATGTCCTCCGGAGCCCCGGGGTCCCCAGGCC  
GCGCCGGCCCCAGCCCTGCGATGCCGCCTGGAGCGGCGCGCCTCGCACTGCAGAT  
GGCTCTCTTAAGGATG (SEC ID nº 56)

(ID de señuelo: FLT3\_diana\_24.1)

2. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674602-28674721

Secuencia original

TACCGAGCAGCGGCAGCTGGCCGCCGTCGCGCGCCAACGCCGGCATGGCC  
TCCGGAGCCCCGGGGTCCCCAGGCCGCGCCGGCCCCAGCCCTGCGATGCCGCCTGG  
AGCGGCGCGCCTCGCG (SEC ID nº 57)

Secuencia modificada

TACCGAGCAGCGGCAGCTGGCCGCCGTCGCGCGCCAACGCCGGCATGGCC  
TCCGGAGCCCCGGGGTCCCCAGGCCGCGCATGCCAGCCCTGCGATGCCGCCTTGA  
GCAACGCGCCTCACG (SEC ID nº 58)

(ID de señuelo: FLT3\_diana\_24.2)

3. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674578-28674697

Secuencia original

GCTGCGAGCGAGCGAGCGGGGCCTTACCGAGCAGCGGCAGCTGGCCGCC  
GTCGCGCGCCAACGCCGGCATGGCCTCCGGAGCCCCGGGGTCCCCAGGCCGCGCC  
GGCCCCAGCCCTGCGATG (SEC ID nº 59)

Secuencia modificada

GCTTCGAGAGAGCGAGCGGGGCCTTACCGAGCAGCAGCAGCTGGCCGCC  
GTCGCGCGCCAACGCCGGCATGGCCTCCGGAGCCCGGGTCCCCAGGCCGCGCC  
AGCCCAGCCCTGAGATG (SEC ID nº 60)

5 (ID de señuelo: FLT3\_diana\_24.3)

4. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674554-28674673

Secuencia original

GTGGGGGCTGAGGGACCGCGAGGGGCTGCGAGCGAGCGAGCGGGGCCTT  
ACCGAGCAGCGGCAGCTGGCCGCCGTGCGCGCGCCAACGCCGGCATGGCCTCCGG  
AGCCCGGGGTCCCCAGG (SEC ID nº 61)

10

Secuencia modificada

GAGGTGGCTGAGAGACCGCGAGGAGCTGCGAGCGAGCGAGCGGGGCCTT  
ACCGAGCAGCGGCAGCTGGCCGCCGTGCGCGCGCCAACGCAGGCATGGCCTCCGG  
AGCCCAGGGTCCCCAGG (SEC ID nº 62)

15

(ID de señuelo: FLT3\_diana\_24.4)

5. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674506-28674625

Secuencia original

CGAGGCGGCTGGGCCGGAGGAGGCGCGCGCCCGGGTCCACACTGCGGGG  
TGGGGGCTGAGGGACCGCGAGGGGCTGCGAGCGAGCGAGCGGGGCCTTACCGA  
GCAGCGGCAGCTGGCCGC (SEC ID nº 63)

20

Secuencia modificada

CGAGGCGGCTGGGCCGGAGGAGGCGCGCGCCCGGATCCACACTGCGGGG  
TGGGGGCTGAGGGACCGCGAGGGGCTGCGAGCGAGCGAGCGGGGACTTACCGA  
GCAGCGGCAACTGGACGC (SEC ID nº 64)

25

(ID de señuelo: FLT3\_diana\_24.5)

6. Diana	génica	Localización genómica del señuelo
FLT3	FLT3_diana_24.	chr13:28674530-28674649

Secuencia original

GCGCGCCCGGGTCCACACTGCGGGGTGGGGGCTGAGGGACCGCGAGGGG  
CTGCGAGCGAGCGAGCGGGGCCTTACCGAGCAGCGGCAGCTGGCCGCCGTGCGC  
CGCCAACGCCGGCATGG (SEC ID nº 65)

30

Secuencia modificada

35

GCACGCACGGATCCACACTGCGGGGTGGGGGCTGAGGGACCGCGAGGAG  
CTGCGAGCGAGCGAGCGGGGCCTTACCGAGCAGCGGCAGCTGGCAGCCGTCGCG  
CGCCAACGCCGGCATGG (SEC ID nº 66)

(ID de señuelo: FLT3\_diana\_24.6)

7. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076516-180076635

5

Secuencia original

TCGCAGGCACAGCGCGGCGCCCCGCTGCATCTCCGGCCGCTGCGCGTGGG  
TCCGACCCGAGCGGCCGCGGCTCGGGGCTGAAAGTGTCCGCGCGGGCGCCGGCT  
GGCCTGGGGCGGGGCG (SEC ID nº 67)

10 Secuencia modificada

CACACACACAAGCGCGGCGCCCCGCTGCATCTCCGGCCGCTGCGCGTGGG  
TCCGACCCGAGCGGCCGCGGCTCGGGGCTGAAAGTGTCCGCGCGGGCGCCGGCT  
GGCCTGCACACACACA (SEC ID nº 68)

(ID de señuelo: FLT4\_diana\_31.1)

15

8. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076396-180076515

Secuencia original

GGCGGAGCGGTCTCAGCGCCCCGCCAGGTGCGCGGTACCCCTCCCCGG  
CCAGCCCCACGCTCGGGCGGGTGGCCCGTTCCGCCGCGCTACCGTCCAGGAGTCC  
CAGGCAGAGCCACAG (SEC ID nº 69)

20

Secuencia modificada

CACACACACATCTCAGCGCCCCGCCAGGTGCGCGGTACCCCTCCCCGG  
CCAGCCCCACGCTCGGGCGGGTGGCCCGTTCCGCCGCGCTACCGTCCAGGAGTCC  
CAGGCCACACACACA (SEC ID nº 70)

25

(ID de señuelo: FLT4\_diana\_31.2)

9. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076420-180076539

Secuencia original

CCAGGTGCGCGGTACCCCTCCCCGGCCAGCCCCACGCTCGGGCGGGTGG  
CCCGTTCCGCCGCGCTACCGTCCAGGAGTCCAGGCAGAGCCACAGTCGCAGGC  
ACAGCGCGGCGCCCCG (SEC ID nº 71)

30

Secuencia modificada

CACACACACAGGTACCCCTCCCCGGCCAGCCCCACGCTCGGGCGGGTGG  
CCCGTTCCGCCGCGCTACCGTCCAGGAGTCCAGGCAGAGCCACAGTCGCAGGC  
ACAGCGCACACACACA (SEC ID nº 72)

35

(ID de señuelo: FLT4\_diana\_31.3)

10. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076468-180076587

Secuencia original

5

GGCCCGTTTCGCCGCGCTCACCGTCCAGGAGTCCCAGGCAGAGCCACAGTC  
GCAGGCACAGCGCGGCGCCCCGCTGCATCTCCGGCCGCTGCGCGTGGGTCCGAC  
CCGAGCGGCCGCGGCT (SEC ID nº 73)

Secuencia modificada

10

CACACACACACCGCGCTCACCGTCCAGGAGTCCCAGGCAGAGCCACAGTC  
GCAGGCACAGCGCGGCGCCCCGCTGCATCTCCGGCCGCTGCGCGTGGGTCCGAC  
CCGAGCCACACACACA (SEC ID nº 74)

(ID de señuelo: FLT4\_diana\_31(4))

11. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076444-180076563

15

Secuencia original

GGCCAGCCCCACGCTCGGGCGGGTGGCCCGTTTCGCCGCGCTCACCGTCCA  
GGAGTCCCAGGCAGAGCCACAGTCGCAGGCACAGCGCGGCGCCCCGCTGCATCT  
CCGGCCGCTGCGCGTG (SEC ID nº 75)

Secuencia modificada

20

CACACACACAACGCTCGGGCGGGTGGCCCGTTTCGCCGCGCTCACCGTCCA  
GGAGTCCCAGGCAGAGCCACAGTCGCAGGCACAGCGCGGCGCCCCGCTGCATCT  
CCGGCCACACACACA (SEC ID nº 76)

(ID de señuelo: FLT4\_diana\_31(5))

12. Diana	génica	Localización genómica del señuelo
FLT4	FLT4_diana_31.	chr5:180076492-180076611

25

Secuencia original

CAGGAGTCCCAGGCAGAGCCACAGTCGCAGGCACAGCGCGGCGCCCCGC  
TGCATCTCCGGCCGCTGCGCGTGGGTCCGACCCGAGCGGCCGCGGCTCGGGGCT  
GAAAGTGTCCGCGCGGG (SEC ID nº 77)

30

Secuencia modificada

CACACACACAAGGCAGAGCCACAGTCGCAGGCACAGCGCGGCGCCCCGC  
TGCATCTCCGGCCGCTGCGCGTGGGTCCGACCCGAGCGGCCGCGGCTCGGGGCT  
GAAAGTGACACACACA (SEC ID nº 78)

(ID de señuelo: FLT4\_diana\_31(6))

35

**Ejemplo H. Un enfoque bayesiano para la detección sensible de alteraciones genómicas somáticas en secuenciación de nueva generación de especímenes clínicos de cáncer**

El enfoque bayesiano indicado en la presente memoria se implementó en los ejemplos siguientes.

La utilidad de este enfoque se ilustra mediante los cálculos de potencia que describen el impacto de previos controlados por los datos sobre la detección de sustituciones en el intervalo inferior de frecuencias de mutación relevantes en el contexto clínico. Tal como se muestra en la figura 4, los valores de expectativa previa (por ejemplo,  $1e-6$  o 10% previo) y la frecuencia de mutación (por ejemplo, 1%, 5% o 15% de mutaciones) corresponde a los valores indicados en (i) y (ii) de "Un enfoque bayesiano para la detección sensible de alteraciones genómicas somáticas en secuenciación de nueva generación de especímenes clínicos de cáncer", respectivamente. La figura 4 muestra que la incorporación de expectativas previas puede mejorar el poder de detección de mutaciones más raras, por ejemplo mediante la reducción de la profundidad de cobertura requerida en los sitios mutados, o el incremento del poder (sensibilidad) estimado para detectar mutaciones.

#### **Ejemplo I. Un enfoque bayesiano: aplicación a una muestra multiclonal de baja pureza construida**

Con el fin de demostrar adicionalmente estos beneficios del enfoque bayesiano dado a conocer en la presente memoria, se construyó una muestra de "tumor" multiclonal de baja pureza artificial mediante la mezcla igual de ADN procedente de 10 participante en el proyecto 1000 Genomes, creando de esta manera un reservorio de ADN que contenía un gran número de variantes de secuencia presente en ~5% o 10% del ADN total (surgido de SNP heterocigóticos privados). La mezcla se sometió a selección de híbridos para exones de 182 genes relacionados con cáncer y se secuenciaron en la plataforma HiSeq2000 de Illumina, proporcionando una cobertura media de aproximadamente 350x en todo el panel génico. De manera similar, se procesó individualmente cada muestra constituyente con el fin de determinar el genotipo en todos los sitios de SNP. De aproximadamente 260 ~5% "mutaciones" presentes en el reservorio, se detectaron 89% con alta confianza utilizando un previo de  $10^{-6}$ , mientras que eran detectables 94% y 95% utilizando un previo de 1% y 10% (cobertura media de sitios faltantes ~125x), respectivamente, respaldando las conclusiones teóricas, anteriormente. De las 102 "mutaciones" 10% presentes en el reservorio, se detectaron 98% con alta confianza utilizando un previo de  $10^{-6}$ , mientras que eran detectables 99% y 99% utilizando un previo de 1% y 10% (cobertura media de sitios faltantes 13x).

#### **Ejemplo J. Un enfoque bayesiano: aplicación a muestras de tumor de pulmón y colon**

Las expectativas previas de frecuencia de mutaciones relevantes en varios tipos de cáncer de la base de datos COSMIC (en la web en [sanger.ac.uk/genetics/CGP/cosmic](http://sanger.ac.uk/genetics/CGP/cosmic)) se derivaron y analizaron más de 80 muestras de cáncer de pulmón y colon extraídas de especímenes clínicos rutinarios. Se observaron mutaciones conocidas en más de 20 genes diferentes, incluyendo un 1% de mutación PIK3CA p.H1047R en un cáncer de colon que sólo pudo detectarse mediante la incorporación del previo 3% para esta mutación en este tipo de cáncer. Estos resultados muestran que la incorporación juiciosa de expectativas previas respecto a los espectros de mutación específicos de tipo tumoral puede resultar beneficioso en la traducción del análisis del genoma tumoral basado en NGS al contexto clínico.

#### **Ejemplo K. Un enfoque bayesiano: aplicación de muestras de cáncer de mama**

Se llevó a cabo la llamada de mutaciones por sustitución en exones de 182 genes relacionados con cáncer secuenciados a ~260x para muestras FFPE de cáncer de mama. El número de sitios con >2 copias de un alelo alternativo era de 1,793. El número de sitios con >99% creencia posterior en presencia de mutación era de 402. El número de sitios que quedó después de los filtros fue de 188, que es aproximadamente el número esperado de sitios variantes. El número de sitios que no se encuentran en dsSNP era de 14, que es aproximadamente el número esperado de sitios no en dbSNP ya que las capturas de dbSNP >90% de variación. El número de sitios no sinónimos era 5. El número de sitios en COSMIC era 2 (PIK3CA p.H1047R y P53 p.F113S).

#### **Ejemplo L. Un enfoque bayesiano: detección de mutaciones infrecuentes**

Muchos especímenes clínicos rutinarios contienen mutaciones raras relevantes. La figura 5 muestra las frecuencias de mutación en más de 100 muestras clínicas de cáncer. Las muestras eran biopsias FFPE, resecciones quirúrgicas o aspirados con aguja fina de predominantemente cánceres de colon y pulmón. El espectro de frecuencias de mutaciones conocidas observadas en una serie de muestras clínicas se muestra en la tabla 12.

Tabla 12. Espectro de frecuencias de mutaciones conocidas observadas en una serie de muestras clínicas.

Espectro de frecuencias de mutaciones conocidas observadas en una serie de muestras clínicas				
Fracción de mutación <5%	Fracción de mutación <10%	Fracción de mutación <25%	Fracción de mutación <50%	Fracción de mutación <100%
7%*	17%	50%	85%	100%
*probablemente subestimado				

### **Ejemplo M.1. Selección de dianas basada en solución de alto rendimiento, utilizando sondas de captura oligonucleótidas sintetizadas individualmente**

La disponibilidad de técnicas de selección de dianas genómicas basadas en solución ha permitido el rápido desarrollo de aplicaciones de secuenciación dirigida, algunas de las cuales han conducido a la introducción de pruebas de secuenciación clínica. Los reactivos de captura de hibridación comercializados se basan en oligonucleótidos sintetizados en matriz, los cuales se convierten en sondas de ADN o ARN biotinilado ("señuelos"). Sin embargo, los métodos para generar estos reservorios complejos de sondas se enfrentan a retos de rendimiento, por ejemplo la captura de dianas de elevado contenido de GC.

En la presente memoria se describe un enfoque alternativo que utiliza oligonucleótidos 5'-biotinilados sintetizados individualmente ("señuelos oligo") para capturar una región diana de ~130 kb que representa 57 genes relacionados con cáncer clínicamente relevantes y accionables. Las bibliotecas de secuenciación indexadas seleccionadas utilizando dichos señuelos oligo con un procedimiento de hibridación de 24 horas proporcionaron un enriquecimiento en la diana de 5,000 veces. 50M lecturas 49 x 49 de extremos apareados generaron una cobertura media de diana de 2100x con una desviación estándar de 568x (27%). Todas las dianas se cubrieron con éxito, con 99.95% de las bases diana cubiertas a >500x. Además, la cobertura de dianas no presentaba virtualmente ningún sesgo de GC. Las dianas con contenido de GC >70% presentaban de media una cobertura de 1,975x y las dianas con contenido de GC <35% presentaban una cobertura media de 1,996x.

Se retuvo el alto rendimiento incluso al utilizar tiempos de hibridación más cortos: 99.3% de las bases diana se cubrieron a >500x tras una hibridación de 2.5 horas.

La utilización de SSPE (esperma de salmón, PE)/solución de Denhardt rindió mejor que los amortiguadores de hib./lavado que contenían TEACI, TMACI y/o sulfato de dextrano.

Los señuelos oligo pueden añadirse a reservorios de señuelos derivados de matriz a fin de incrementar la cobertura de regiones que de otro modo resultarían difíciles de capturar (por ejemplo, de alto % de GC) o para añadir rápidamente nuevo contenido génico. Este enfoque ofrece un método altamente eficaz y escalable para desarrollar pruebas de secuenciación clínica dirigida de alto rendimiento.

### **Ejemplo M.2. Método de optimización de señuelos de captura**

Se sometieron a ensayo tres señuelos. Los resultados se resumen en la figura 7. Los conjuntos de señuelos fueron los siguientes:

el conjunto de señuelos nº 1 consistía en únicamente señuelos oligonucleótidos de ADN sintetizados individualmente.

El conjunto de señuelos nº 2 incluía señuelos de ADN derivados de matriz biotinilados con adición de señuelos oligonucleótidos de ADN sintetizados individualmente 5'-biotinilados.

El conjunto de señuelos nº 3 consistía en únicamente señuelos de ARN derivados de matriz biotinilados.

Todos los oligonucleótidos de ADN sintetizados individualmente 5'-biotinilados presentaban 120 bases con una biotina en 5'.

La figura 7 es un histograma de coberturas que compara la uniformidad de cobertura detectada con el conjunto de señuelos nº 1 y el conjunto de señuelos nº 2, en comparación con el conjunto de señuelos nº 3. Los conjuntos de señuelos se muestran como nº 1, 2 y 3 en la figura 7. Se encontraban presentes varios huecos en la cobertura utilizando el conjunto de señuelos nº 3, correspondientes a un alto % de GC, mientras que las regiones correspondientes se cubrieron profundamente utilizando los conjuntos de señuelos nº 1 y nº 2, tal como se ilustra en la figura 7. En la figura 7, el panel a la izquierda, etiquetado "GC\_densidad\_diana..." indica el contenido local de GC dentro de la diana. La línea representa un contenido de GC de 65%, en el que cualesquiera valores sobre la línea representan un contenido de GC más elevado. Tal como se muestra en el histograma, la cobertura es más baja para el conjunto de señuelos nº 3, en zonas de elevado contenido de GC. El panel inferior en la figura 7 etiquetado "IDT\_señuelos..." indica la localización de los oligos que cubren la diana mostrada.

Una representación gráfica de los cambios en el número de dianas y cobertura con los conjuntos de señuelos derivados de matriz solos o con adición de señuelos sintetizados individualmente se ilustra en la figura 6. Más específicamente, la figura 6 es una representación lineal de un histograma de coberturas. El número de dianas (eje y) se ilustra como función de la cobertura (eje x). La línea nº 1 representa la cobertura utilizando un conjunto de señuelos que incluye señuelos oligonucleótidos de ARN derivados de matriz 5'-biotinilados con adición de señuelos oligonucleótidos de ADN sintetizados individualmente 5'-biotinilados (denominados en la figura 6, "Conjunto de señuelos nº 1"). La línea nº 2 representa la cobertura utilizando un conjunto de señuelos que incluye señuelos oligonucleótidos de ARN derivados de matriz biotinilados únicamente (denominados en la figura 6,

“Conjunto de señuelos nº 2”). La cobertura media global utilizando el conjunto de señuelos nº 2 era de 924, mientras que la cobertura en zonas de elevado contenido de GC (aproximadamente 68%) utilizando el conjunto de señuelos nº 2 era de 73. En contraste, al utilizar el conjunto de señuelos nº 1, la cobertura global era similar al conjunto de señuelos nº 1, aproximadamente 918, pero la cobertura había mejorado hasta 183 en zonas de elevado contenido de GC.

### Ejemplo M.3. Condiciones experimentales ejemplificativas para evaluar los conjuntos de señuelos

El conjunto de señuelos A consistía en únicamente señuelos oligonucleótidos de ADN sintetizados individualmente 5'-biotinilados. El conjunto original presentaba 1000 oligos, cubriendo 133 kb de territorio diana (denominado en la presente memoria “el conjunto grande”, “conjunto de señuelos A” o “señuelos oligo de ADN”).

Para los experimentos “de adición”, se añadió el conjunto de 1000 oligos de ADN original (“el conjunto grande”) a un conjunto de señuelos que consistía en señuelos oligonucleótidos de ARN derivados de matriz biotinilados (denominados en el presente ejemplo “Conjunto de señuelos B” o “señuelos de ARN”). Se mezclaron diferentes proporciones de señuelos oligo de ADN del conjunto de señuelos A con señuelos de ARN procedentes del conjunto de señuelos B. En particular, se utilizó una proporción de señuelos oligo de ADN: señuelos de ARN de 1:10 (10 ng de señuelos oligo de ADN total para 100 ng de señuelos de ARN total). Las condiciones de hibridación y lavado se hicieron corresponder a las más ideales para los señuelos de ARN.

Con densidades de embaldosado bajas, se detectaron periodicidades fuertes en la cobertura al utilizar señuelos oligo de ADN que correspondían a la localización de los señuelos. Además, las densidades de embaldosado bajas pueden dificultar la captura de alelos con mutaciones indel. Por lo tanto, se diseñaron conjuntos de señuelos para MAP3K1 con las diferentes densidades de embaldosado ilustradas en la Tabla 13. En las mezclas posteriores, se añadió la Mezcla 1, que contenía señuelos oligo de ADN individualmente sintetizados 5'-biotinilados diseñados para capturar los exones de seis genes relevantes al cáncer (DAXX, TRRAP, CREBBP, GRIN2A, SPOP, GNA11) a los señuelos oligonucleótidos de ARN derivados de matriz únicamente (conjunto de señuelos B). DAXX, TRRAP, CREBBP, GRIN2A y SPOP no se encontraban presentes en el conjunto de señuelos de ARN. Las Mezclas 2 a 4 se añadieron al conjunto de señuelos A (el conjunto grande de señuelos oligo de ADN) a fin de someter a ensayo las diferentes densidades de embaldosado (siendo la Mezcla 2 la más densa) de señuelos de captura para los exones de MAP3K1. El conjunto de señuelos de ARN por sí solo cubrió aproximadamente 1 MB de secuencia.

Tabla 13. Mezclas para métodos que utilizaban sondas de captura

Categoría	Número
Mezcla 1	369 oligos de genes de melanoma
Mezcla 2	91 oligos densidad de baldosas de 60 de MAP3K1
Mezcla 3	57 oligos densidad de baldosas de 100 de MAP3K1
Mezcla 4	40 oligos densidad de baldosas de 150 de toMAP3K1
Mezcla 5	3 oligos de STK11 exón 3

La cantidad introducida en la captura fue de 2 µg de bibliotecas de ADN de línea celular agrupadas. Se mezclaron 2 µg de biblioteca con mezcla de bloqueo (tabla 14), se secaron y se resuspendieron en 9 µl de agua. A continuación, dicha mezcla se añadió a una placa, se transfirió a un ciclador y se sometió a 98°C durante 5 minutos, seguido de 68°C durante 2 minutos. Seguidamente se abrió la placa y se añadieron 11 µl de mezcla de señuelo de ADN/amortiguador de hib. a 68°C. La mezcla de señuelo de ADN/hib. A 68°C = 10 µl de amortiguador de hib. + 1 µl de señuelo (que contenía 10 ng, 50 ng o 100 ng de señuelo).

Para las capturas con señuelos de ADN solamente (por ejemplo, el conjunto de señuelos A), la hibridación se llevó a cabo a 68°C y se llevaron a cabo lavados. Los señuelos se sometieron a ensayo a 5 ng, 10 ng, 100 ng, 1000 ng y 2000 ng (por cada 2 µg de biblioteca de entrada). Para las hib. de 24 h, se sometieron a ensayo las condiciones de 5-10 ng y condiciones hasta 100 ng.

Para las capturas con el conjunto de señuelos de ADN grande (100 kb) con adición al conjunto de señuelos de matriz de ARN (B) para rescatar regiones de bajo rendimiento/elevado contenido de GC, la hibridación se llevó a cabo a 68°C y los lavados se llevaron a cabo a 70°C. Los conjuntos de señuelos se sometieron a ensayo a una proporción de oligo de ADN: señuelos de ARN de 1:10 (es decir, 10 ng de masa total de señuelos oligo y 100 ng de masa total de señuelos de ARN).

Para capturas con el conjunto pequeño de señuelos de ADN enfocado en un gen añadido al conjunto de señuelos de ARN, la hibridación se llevó a cabo a 68°C y se sometió a ensayo un abanico de temperaturas de lavado (62°C, 64°C, 66°C, 68°C, 70°C y 72°C).

La Mezcla 1 (adición de 6 genes nuevos) se sometió a ensayo en las proporciones siguientes: 1:5, 1:10 y 1:20 de masa total de señuelo oligo de ADN: masa de señuelo de ARN (es decir, 20 ng: 100 ng, 10 ng:100 ng y 5 ng:100 ng).

- 5 La Mezcla 5 (3 oligos que representaban el exón 3 de STK11, baja cobertura) se sometieron a ensayo a 1:500, 1:1000 y 1:2000 oligo de ADN:oligo de ARN. Se utilizaron 100 ng de señuelos de ARN total. Se sometió a ensayo STK11, ya que representa una importante diana de cáncer con bajo rendimiento de detección al capturarlo con los señuelos de ARN solamente. La adición de oligos de ADN del exón 3 de STK11 potencia la cobertura de una media de 70x a 300x.

10

Tabla 14. Amortiguadores para métodos que utilizaban sondas de captura

<u>Señuelos (oligos de IDT agrupados)</u>	39600 (g/mol)	100 nmoles =	0.0039600 gramos = 3,960,000 nanogramos
resuspendidos en amortiguador 'low TE'	25 ml	250 µl de Tris, 5 µl de EDTA	
<u>Mezcla de bloqueo</u>	<i>[Solución madre]</i>	<i>[Solución de trabajo]</i>	14.5 µl/reacción
Cot1	1 µg/µl	1 µg/µl	10
Esperma de salmón	10 µg/µl	10.0 µg/µl	1
PE 1.0	800 µM	800 µM	1,75
Índice universal	800 µM	800 µM	1,75
<u>2 x amortiguador de hib.</u>	<i>[Solución madre]</i>	<i>[Final]</i>	en 10 ml (10 µl/reacción)
SSPE	20x	10x	5 ml
Solución de Denhardt	50x	10x	2 ml
EDTA	0.5M	0.01 M	200 µl
SDS	10%	0.20%	200 µl
Agua			2.6 ml
<u>Lavado de perlas</u>	<i>[Solución madre]</i>	<i>[Final]</i>	en 50 ml (200 µl/lavado)
NaCl	5 M	1 M	10 ml
Tris	1 M	10 mM	500 µl
EDTA	0.5 M	1 mM	100 µl
Agua			39.4 ml
<u>Amortiguador de lavado 1</u>	<i>[Solución madre]</i>	<i>[Final]</i>	en 50 ml (150 µl/lavado)
SSC	20x	1x	2.5 ml
SDS	10%	0.10%	500 µl
Agua			47 ml
<u>Amortiguador de lavado 2</u>	<i>[Solución madre]</i>	<i>[Final]</i>	en 50 ml (150 µl/lavado)
SSC	20x	0.1x	250 µl
SDS	10%	0.10%	500 µl
Agua			49.25 ml

15

#### Ejemplo N. Reducción de la unión fuera de diana de ácidos nucleicos de miembros de biblioteca

Las interacciones de ácidos nucleicos fuera de diana pueden limitar la eficiencia de la selección de los ácidos nucleicos diana mediante hibridación (por ejemplo, hibridación en solución o en fase sólida) con una sonda de captura, por ejemplo un señuelo oligonucleótido. La selección fuera de diana típicamente se incrementa al reducirse las condiciones de astringencia de la selección de híbridos, por ejemplo al seleccionar un dúplex de diana:captura que presenta una temperatura de fusión de los ácidos nucleicos más baja (por ejemplo, la  $T_m$  de los dúplex de ADN:ADN comparado con la de los dúplex de ARN:ADN). De esta manera, la captura de una secuencia fuera de diana puede resultar más problemática en las hibridaciones de ADN:ADN. La selección de secuencias fuera de diana puede resultar, por ejemplo, en uno o más de: rendimientos reducidos de captura de hibridación y/o la captura de híbridos artefactos que, a su vez, conduce a ineficiencias en etapas posteriores, por ejemplo, en la secuenciación.

Los miembros de biblioteca pueden incluir una inserción de biblioteca (que, si está en la diana, forma un dúplex con la sonda de captura, por ejemplo, un señuelo) y una o más secuencias no diana (por ejemplo, una o más secuencias adaptadoras, cebadores o etiquetas de amplificación y etiquetas de código de barras). Típicamente, un señuelo se hibrida con la inserción de biblioteca, por ejemplo, un ADN diana. Sin embargo, la inserción de biblioteca puede presentar adaptadores universales, los cuales se encuentran típicamente presentes en cada fragmento en la biblioteca. La secuencia no diana del miembro de biblioteca hibridado con sonda de captura puede,

30

mediante formación de dúplex con otras secuencias en la mezcla de reacción (por ejemplo, mediante unión a secuencias adaptadoras), conducir a la selección de secuencias no deseadas, por ejemplo miembros de una biblioteca de secuencias fuera de diana.

Aunque sin respaldo teórico, la concatenación de miembros de una biblioteca de secuencias fuera de diana que ha formado un dúplex con la sonda de captura y secuencias fuera de diana puede resultar en la selección de secuencias fuera de diana. La figura 6 ilustra en forma de diagrama una configuración ejemplificativa de concatámeros no diana de los miembros de biblioteca. Las regiones no diana (por ejemplo, los adaptadores ilustrados como "P5" y "P7") se muestran como hibridantes con sus cadenas no diana complementarias (ilustradas como "rcP5" y "rcP7", respectivamente). Se muestra un señuelo etiquetado con biotina que se hibrida con una región complementaria de la inserción de diana del miembro de biblioteca. La unión fuera de diana puede conducir a una concatenación de miembros de biblioteca, conduciendo de esta manera a una reducción de la especificidad de unión a la diana (también denominada en la presente memoria, selección fuera de diana incrementada).

En los dúplex en la diana:captura que implican dúplex de ADN (miembro de biblioteca):ARN (señuelo), la concatenación entre un miembro de biblioteca en la diana que ha formado un dúplex con la sonda de captura y secuencias fuera de diana puede romperse durante los lavados de alta astringencia típicamente realizadas a 65-70°C. Típicamente, los lavados que implican una menor fusión de los dúplex de ADN:ADN se llevan a cabo a temperaturas más bajas que los dúplex de ARN:ADN. La incapacidad de romper la concatenación mantiene el porcentaje de captura de dianas relativamente bajo al utilizar señuelos de ADN (45-50%). Los oligos bloqueantes disponibles comercialmente complementarios a adaptadores se añaden para minimizar la concatenación, aunque típicamente no inhiben adecuadamente la formación de cadenas, particularmente en las hibridaciones ADN:ADN.

Se dan a conocer en la presente memoria métodos y composiciones que reducen la selección de secuencias no diana, por ejemplo la selección mediada por adaptadores. En determinadas formas de realización, se dan a conocer oligonucleótidos bloqueantes que son complementarios o pueden formar un dúplex con la secuencia de ácidos nucleicos no diana del miembro de biblioteca (por ejemplo, una secuencia adaptadora) y presentan un valor de un parámetro relacionado con la interacción de unión entre el oligonucleótido bloqueante y la secuencia de ácidos nucleicos no diana del miembro de biblioteca que es más elevado que el valor de la secuencia de ácidos nucleicos no diana respecto a un ácido nucleico de fondo, por ejemplo otras secuencias de ácidos nucleicos no diana complementarias. Entre los oligonucleótidos bloqueantes ejemplificativos que presentan una interacción de unión incrementada se incluyen oligonucleótidos que presentan una longitud de bloqueante extendida, por ejemplo una complementariedad extendida respecto a un ácido nucleico no diana; los oligonucleótidos bloqueantes que presentan uno o más nucleótidos no naturales y los oligonucleótidos bloqueantes que incluyen (o están compuestos sustancialmente) de oligonucleótidos, en lugar de desoxirribonucleótidos.

#### **Ejemplo O. Longitud de bloqueante extendida**

El presente ejemplo demuestra que el porcentaje de selección en la diana puede mejorarse mediante la extensión de la longitud del oligonucleótido bloqueante.

Se añaden oligonucleótidos bloqueantes específicos de adaptador a la reacción de hibridación llevada a cabo tal como se indica en la presente memoria para evitar la contaminación cruzada de unión fuera de diana de ácidos nucleicos, tal como se indica en el Ejemplo 14. En las condiciones experimentales indicadas en el Ejemplo 4, se llevan a cabo lavados de alta astringencia, que es probable que desnaturalicen la unión fuera de diana. Sin embargo, las condiciones de hibridación y lavado óptimas para las interacciones ADN:ADN reducen las temperaturas de los lavados, tal como se indica en los Ejemplos 13A-13C, incrementando de esta manera la unión fuera de diana.

Pueden diseñarse oligos bloqueantes complementarios a los adaptadores, por ejemplo los adaptadores multiplex de Illumina indicados en el Ejemplo C, a fin de incrementar el grado de complementariedad entre el adaptador y el oligo bloqueante. Por ejemplo, el oligo bloqueante P5 presenta una longitud de 58 pb pero el bloqueante sólo presenta 46 bases. La longitud del oligo bloqueante P5 se extendió en 19 bases. La extensión de la longitud del oligo bloqueante en 19 bases incrementó la eficiencia de selección en aproximadamente 5% (tal como se muestra en la figura 9). La figura 9 es un gráfico de columnas que ilustra el porcentaje de selección de dianas utilizando oligos de bloqueo estándares y ampliados. Se muestran los datos de cuatro experimentos representativos. La figura 10 ilustra un histograma de cobertura de exones que muestra los resultados de captura utilizando bloqueantes estándares o extendidos.

Puede conseguirse un bloqueo mejorado mediante la extensión de la longitud de la región complementaria entre el adaptador y el oligo bloqueante, incrementando de esta manera la temperatura de fusión.

La terminología utilizada en la presente memoria tiene el propósito de describir formas de realización particulares únicamente, y que no pretender ser limitativa. Con respecto a la utilización de sustancialmente, cualesquiera términos plurales y/o singulares en la presente memoria, el experto en la materia podrá traducirlos del plural según

resulte apropiado al contexto y/o aplicación. Las diversas permutaciones singulares/plurales pueden indicarse expresamente en la presente memoria en aras de la claridad.

- 5 Se pretende que la presente invención no se limite a las formas de realización particulares o ejemplos dados a conocer, sino que la presente invención incluye la totalidad de las formas de realización comprendidas dentro del alcance según las reivindicaciones adjuntas.

#### Listado de secuencias

- 10 <110> Integrated DNA Technologies, Inc. Foundation Medicine, Inc.
- <120> OLIGONUCLEÓTIDOS DE BLOQUEO AUMENTADOS EN TM Y SEÑUELOS PARA UN ENRIQUECIMIENTO DE DIANA MEJORADO Y UNA SELECCIÓN FUERA DE DIANA REDUCIDA
- 15 <130> IDT01-001-PCT
- <150> US 61/667,919
- <151> 2012-07-03
- 20 <150> US 61/745,435
- <151> 2012-12-21
- <160> 99
- 25 <170> PatentIn versión 3.5
- <210> 1
- <211> 58
- <212> ADN
- 30 <213> Secuencia artificial
- <220>
- <223> Oligonucleótido sintético
- 35 <400> 1
- agatcggaag agcgtcgtgt agggaaagag ttagatctc ggtggtcgcc gtatcatt 58
- <210> 2
- <211> 58
- 40 <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Oligonucleótido sintético
- 45 <220>
- <221> misc\_feature
- <222> (16)..(16)
- <223> iANB-meC en posiciones 16.
- 50 <220>
- <221> misc\_feature
- <222> (40)..(40)
- <223> iANB-meC en posiciones 40.
- 55 <400> 2
- agatcggaag agcgtcgtgt agggaaagag ttagatctc ggtggtcgcc gtatcatt 58
- <210> 3
- <211> 58
- 60 <212> ADN
- <213> Secuencia artificial
- <220>
- 65 <223> Oligonucleótido sintético

5 <220>  
 <221> misc\_feature  
 <222> (5)..(5)  
 <223> iANB-meC en posición 5.

10 <220>  
 <221> misc\_feature  
 <222> (16)..(16)  
 <223> iANB-meC en posición 16.

15 <220>  
 <221> misc\_feature  
 <222> (40)..(40)  
 <223> iANB-meC en posición 40.

20 <220>  
 <221> misc\_feature  
 <222> (50)..(50)  
 <223> iANB-meC en posición 50.

25 <400> 3  
 agatcgggaag agcgcgtgt agggaaagag ttagatctc ggtggtcgcc gtatcatt 58

30 <210> 4  
 <211> 58  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <223> Oligonucleótido sintético

40 <220>  
 <221> misc\_feature  
 <222> (5)..(5)  
 <223> iANB-meC en posición 5.

45 <220>  
 <221> misc\_feature  
 <222> (13)..(13)  
 <223> iANB-meC en posición 13.

50 <220>  
 <221> misc\_feature  
 <222> (16)..(16)  
 <223> iANB-meC en posición 16

55 <220>  
 <221> misc\_feature  
 <222> (40)..(40)  
 <223> iANB-meC en posición 40.

60 <220>  
 <221> misc\_feature  
 <222> (16)..(16)  
 <223> iANB-meC en posición 16

65 <220>  
 <221> misc\_feature  
 <222> (47)..(47)  
 <223> iANB-meC en posición 47.

70 <220>  
 <221> misc\_feature  
 <222> (50)..(50)  
 <223> iANB-meC en posición 50.

75 <220>  
 <221> misc\_feature  
 <222> (55)..(55)  
 <223> iANB-meC en posición 55.

	<400> 4 agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58
5	<210> 5 <211> 58 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
15	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
20	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC está en posición 13.	
25	<220> <221> misc_feature <222> (16)..(16) <223> iANB-meC está en posición 16.	
30	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC está en posición 38.	
35	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.	
40	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.	
45	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC está en posición 50.	
50	<220> <221> misc_feature <222> (55)..(55) <223> iANB-meC está en posición 55.	
55	<400> 5 agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58
60	<210> 6 <211> 58 <212> ADN <213> Secuencia artificial	
65	<220> <223> Oligonucleótido sintético	
	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	

5 <220>  
<221> misc\_feature  
<222> (9)..(9)  
<223> iANB-A está en posición 9.

10 <220>  
<221> misc\_feature  
<222> (13)..(13)  
<223> iANB-meC está en posición 13.

15 <220>  
<221> misc\_feature  
<222> (16)..(16)  
<223> iANB-meC está en posición 16.

20 <220>  
<221> misc\_feature  
<222> (21)..(21)  
<223> iANB-A está en posición 21.

25 <220>  
<221> misc\_feature  
<222> (25)..(25)  
<223> iANB-A está en posición 25.

30 <220>  
<221> misc\_feature  
<222> (38)..(38)  
<223> iANB-meC está en posición 38.

35 <220>  
<221> misc\_feature  
<222> (40)..(40)  
<223> iANB-meC está en posición 40.

40 <220>  
<221> misc\_feature  
<222> (47)..(47)  
<223> iANB-meC está en posición 47.

45 <220>  
<221> misc\_feature  
<222> (49)..(49)  
<223> iANB-meC está en posición 49.

50 <220>  
<221> misc\_feature  
<222> (50)..(50)  
<223> iANB-meC está en posición 50.

55 <220>  
<221> misc\_feature  
<222> (55)..(55)  
<223> iANB-meC está en posición 55.

60 <400> 6  
agatcgggaag agcgcgtgtg agggaaagag tgtagatctc ggtggtcgcc gtatcatt 58

65 <210> 7  
<211> 58  
<212> ADN  
<213> Secuencia artificial

65 <220>  
<223> Oligonucleótido sintético

5	<220> <221> misc_feature <222> (3)..(3) <223> iANB-A está en posición 3.
10	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.
15	<220> <221> misc_feature <222> (9)..(9) <223> iANB-A está en posición 9.
20	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC está en posición 13.
25	<220> <221> misc_feature <222> (16)..(6) <223> iANB-meC está en posición 16.
30	<220> <221> misc_feature <222> (21)..(21) <223> iANB-A está en posición 21.
35	<220> <221> misc_feature <222> (25)..(24) <223> iANB-A está en posición 25.
40	<220> <221> misc_feature <222> (26)..(26) <223> iANB-A está en posición 26.
45	<220> <221> misc_feature <222> (27)..(27) <223> iANB-A está en posición 27.
50	<220> <221> misc_feature <222> (36)..(36) <223> iANB-A está en posición 36.
55	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC está en posición 38.
60	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.
65	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.

5	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (49)..(49)          &lt;223&gt; iANB-meC está en posición 49.</p>	
10	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (50)..(50)          &lt;223&gt; iANB-meC está en posición 50.</p>	
15	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (55)..(55)          &lt;223&gt; iANB-meC está en posición 55.</p>	
	<p>&lt;400&gt; 7          agatcggaag agcgctgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt</p>	58
20	<p>&lt;210&gt; 8          &lt;211&gt; 58          &lt;212&gt; ADN          &lt;213&gt; Secuencia artificial</p>	
25	<p>&lt;220&gt;          &lt;223&gt; Oligonucleótido sintético</p>	
30	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (3)..(3)          &lt;223&gt; iANB-A está en posición 3.</p>	
35	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (5)..(5)          &lt;223&gt; iANB-meC está en posición 5.</p>	
40	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (8)..(8)          &lt;223&gt; iANB-A está en posición 8.</p>	
45	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (9)..(9)          &lt;223&gt; iANB-A está en posición 9.</p>	
50	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (11)..(11)          &lt;223&gt; iANB-A está en posición 11.</p>	
55	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (13)..(13)          &lt;223&gt; iANB-meC está en posición 13.</p>	
60	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (16)..(16)          &lt;223&gt; iANB-meC está en posición 16.</p>	
65	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (21)..(21)          &lt;223&gt; iANB-A está en posición 21.</p>	

	<220>
	<221> misc_feature
	<222> (25)..(25)
5	<223> iANB-A está en posición 25.
	<220>
	<221> misc_feature
	<222> (26)..(26)
10	<223> iANB-A está en posición 26.
	<220>
	<221> misc_feature
	<222> (27)..(27)
15	<223> iANB-A está en posición 27.
	<220>
	<221> misc_feature
	<222> (29)..(29)
20	<223> iANB-A está en posición 29.
	<220>
	<221> misc_feature
	<222> (34)..(34)
25	<223> iANB-A está en posición 34.
	<220>
	<221> misc_feature
	<222> (36)..(36)
30	<223> iANB-A está en posición 36.
	<220>
	<221> misc_feature
	<222> (38)..(38)
35	<223> iANB-meC está en posición 38.
	<220>
	<221> misc_feature
	<222> (40)..(40)
40	<223> iANB-meC está en posición 40.
	<220>
	<221> misc_feature
	<222> (47)..(47)
45	<223> iANB-meC está en posición 47.
	<220>
	<221> misc_feature
	<222> (49)..(49)
50	<223> iANB-meC está en posición 49.
	<220>
	<221> misc_feature
	<222> (50)..(50)
55	<223> iANB-meC está en posición 50.
	<220>
	<221> misc_feature
	<222> (53)..(53)
60	<223> iANB-A está en posición 53.
	<220>
	<221> misc_feature
	<222> (55)..(55)
65	<223> iANB-meC está en posición 55.
	<220>

	<221> misc_feature	
	<222> (56)..(56)	
	<223> iANB-A está en posición 56.	
5	<400> 8	
	agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggcgcc gtatcatt	58
	<210> 9	
	<211> 65	
10	<212> ADN	
	<213> Secuencia artificial	
	<220>	
15	<223> Oligonucleótido sintético	
	<220>	
	<221> modified_base	
	<222> (34)..(42)	
20	<223> I	
	<220>	
	<221> misc_feature	
	<222> (34)..(41)	
25	<223> n es a, c, g o t	
	<400> 9	
	gacggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgat gccgtcttct	60
	gcttg	65
	<210> 10	
30	<211> 65	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
35	<223> Oligonucleótido sintético	
	<220>	
	<221> misc_feature	
	<222> (31)..(31)	
40	<223> iANB-meC está en posición 31.	
	<220>	
	<221> modified_base	
	<222> (34)..(41)	
45	<223> I	
	<220>	
	<221> misc_feature	
	<222> (34)..(41)	
50	<223> n es a, c, g o t	
	<220>	
	<221> misc_feature	
	<222> (46)..(46)	
55	<223> iANB-meC está en posición 46.	
	<400> 10	
	gacggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgat gccgtcttct	60
	gcttg	65
60	<210> 11	
	<211> 65	
	<212> ADN	

<213> Secuencia artificial  
 <220>  
 <223> Oligonucleótido sintético  
 5  
 <220>  
 <221> misc\_feature  
 <222> (24)..(24)  
 <223> iANB-meC está en posición 24.  
 10  
 <220>  
 <221> misc\_feature  
 <222> (31)..(31)  
 <223> iANB-meC está en posición 31.  
 15  
 <220>  
 <221> modified\_base  
 <222> (34)..(41)  
 <223> I  
 20  
 <220>  
 <221> misc\_feature  
 <222> (34)..(41)  
 <223> n es a, c, g o t  
 25  
 <220>  
 <221> misc\_feature  
 <222> (46)..(46)  
 <223> iANB-meC está en posición 46.  
 30  
 <220>  
 <221> misc\_feature  
 <222> (53)..(53)  
 <223> iANB-meC está en posición 53.  
 35  
 <400> 11  
 gatcggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct 60  
 gcttg 65  
 40  
 <210> 12  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 45  
 <220>  
 <223> Oligonucleótido sintético  
 <220>  
 <221> misc\_feature  
 <222> (12)..(12)  
 <223> iANB-meC está en posición 12.  
 50  
 <220>  
 <221> misc\_feature  
 <222> (24)..(24)  
 <223> iANB-meC está en posición 24.  
 55  
 <220>  
 <221> misc\_feature  
 <222> (31)..(31)  
 <223> iANB-meC está en posición 31.  
 60  
 <220>  
 <221> modified\_base  
 <222> (34)..(41)

	<223> I	
	<220>	
5	<221> misc_feature	
	<222> (34)..(41)	
	<223> n es a, c, g o t	
	<220>	
10	<221> misc_feature	
	<222> (46)..(46)	
	<223> iANB-meC está en posición 46.	
	<220>	
15	<221> misc_feature	
	<222> (53)..(53)	
	<223> iANB-meC está en posición 53.	
	<220>	
20	<221> misc_feature	
	<222> (62)..(62)	
	<223> iANB-meC está en posición 62.	
	<400> 12	
	gatcgggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct	60
	gcttg	65
25		
	<210> 13	
	<211> 65	
	<212> ADN	
	<213> Secuencia artificial	
30		
	<220>	
	<223> Oligonucleótido sintético	
	<220>	
35	<221> misc_feature	
	<222> (4)..(4)	
	<223> iANB-meC está en posición 4.	
	<220>	
40	<221> misc_feature	
	<222> (12)..(12)	
	<223> iANB-meC está en posición 12.	
	<220>	
45	<221> misc_feature	
	<222> (24)..(24)	
	<223> iANB-meC está en posición 24.	
	<220>	
50	<221> misc_feature	
	<222> (31)..(31)	
	<223> iANB-meC está en posición 31.	
	<220>	
55	<221> misc_feature	
	<222> (33)..(33)	
	<223> iANB-meC está en posición 33.	
	<220>	
60	<221> modified_base	
	<222> (34)..(41)	
	<223> I	
	<220>	

	<221> misc_feature	
	<222> (34)..(41)	
	<223> n es a, c, g o t	
5	<220>	
	<221> misc_feature	
	<222> (46)..(46)	
	<223> iANB-meC está en posición 46.	
10	<220>	
	<221> misc_feature	
	<222> (53)..(53)	
	<223> iANB-meC está en posición 53.	
15	<220>	
	<221> misc_feature	
	<222> (62)..(62)	
	<223> iANB-meC está en posición 62.	
20	<400> 13	
	gacggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgat gccgtcttct	60
	gcttg	65
	<210> 14	
	<211> 65	
25	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
30	<220>	
	<221> misc_feature	
	<222> (4)..(4)	
	<223> iANB-meC está en posición 4.	
35	<220>	
	<221> misc_feature	
	<222> (12)..(12)	
	<223> iANB-meC está en posición 12.	
40	<220>	
	<221> misc_feature	
	<222> (16)..(16)	
	<223> iANB-meC está en posición 16.	
45	<220>	
	<221> misc_feature	
	<222> (19)..(19)	
	<223> iANB-meC está en posición 19.	
50	<220>	
	<221> misc_feature	
	<222> (24)..(24)	
	<223> iANB-meC está en posición 24.	
55	<220>	
	<221> misc_feature	
	<222> (27)..(27)	
	<223> iANB-meC está en posición 27.	
60	<220>	
	<221> misc_feature	
	<222> (31)..(31)	
	<223> iANB-meC está en posición 31.	

5 <220>  
 <221> modified\_base  
 <222> (34)..(41)  
 <223> |

10 <220>  
 <221> misc\_feature  
 <222> (34)..(41)  
 <223> n es a, c, g o t

15 <220>  
 <221> misc\_feature  
 <222> (46)..(46)  
 <223> iANB-meC está en posición 46.

20 <220>  
 <221> misc\_feature  
 <222> (53)..(53)  
 <223> iANB-meC está en posición 53.

25 <220>  
 <221> misc\_feature  
 <222> (56)..(56)  
 <223> iANB-meC está en posición 56.

30 <220>  
 <221> misc\_feature  
 <222> (59)..(59)  
 <223> iANB-meC está en posición 59.

35 <220>  
 <221> misc\_feature  
 <222> (62)..(62)  
 <223> iANB-meC está en posición 62.

<400> 14  
 gatcgggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct 60  
 gcttg 65

40 <210> 15  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> Oligonucleótido sintético

50 <220>  
 <221> misc\_feature  
 <222> (4)..(4)  
 <223> iANB-meC está en posición 4.

55 <220>  
 <221> misc\_feature  
 <222> (12)..(12)  
 <223> iANB-meC está en posición 12.

60 <220>  
 <221> misc\_feature  
 <222> (14)..(14)  
 <223> iANB-meC está en posición 14.

<220>  
 <221> misc\_feature

<222> (16)..(16)  
 <223> iANB-meC está en posición 16.

5      <220>  
       <221> misc\_feature  
       <222> (19)..(19)  
       <223> iANB-meC está en posición 19.

10     <220>  
       <221> misc\_feature  
       <222> (24)..(24)  
       <223> iANB-meC está en posición 24.

15     <220>  
       <221> misc\_feature  
       <222> (26)..(26)  
       <223> iANB-meC está en posición 26.

20     <220>  
       <221> misc\_feature  
       <222> (27)..(27)  
       <223> iANB-meC está en posición 27.

25     <220>  
       <221> misc\_feature  
       <222> (31)..(31)  
       <223> iANB-meC está en posición 31.

30     <220>  
       <221> misc\_feature  
       <222> (33)..(33)  
       <223> iANB-meC está en posición 33.

35     <220>  
       <221> modified\_base  
       <222> (34)..(41)  
       <223> l

40     <220>  
       <221> misc\_feature  
       <222> (34)..(41)  
       <223> n es a, c, g o t

45     <220>  
       <221> misc\_feature  
       <222> (44)..(44)  
       <223> iANB-meC está en posición 44.

50     <220>  
       <221> misc\_feature  
       <222> (46)..(46)  
       <223> iANB-meC está en posición 46.

55     <220>  
       <221> misc\_feature  
       <222> (52)..(52)  
       <223> iANB-meC está en posición 52.

60     <220>  
       <221> misc\_feature  
       <222> (53)..(53)  
       <223> iANB-meC está en posición 53.

65     <220>  
       <221> misc\_feature  
       <222> (56)..(56)

<223> iANB-meC está en posición 56.

<220>  
<221> misc\_feature  
5 <222> (59)..(59)  
<223> iANB-meC está en posición 59.

<220>  
10 <221> misc\_feature  
<222> (62)..(62)  
<223> iANB-meC está en posición 62.

<400> 15  
gatcgggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct 60  
gcttg 65

15 <210> 16  
<211> 65  
<212> ADN  
<213> Secuencia artificial

20 <220>  
<223> Oligonucleótido sintético

25 <220>  
<221> misc\_feature  
<222> (4)..(4)  
<223> iANB-meC está en posición 4.

30 <220>  
<221> misc\_feature  
<222> (7)..(7)  
<223> iANB-A está en posición 7.

35 <220>  
<221> misc\_feature  
<222> (8)..(8)  
<223> iANB-A está en posición 8.

40 <220>  
<221> misc\_feature  
<222> (12)..(12)  
<223> iANB-meC está en posición 12.

45 <220>  
<221> misc\_feature  
<222> (14)..(14)  
<223> iANB-meC está en posición 14.

50 <220>  
<221> misc\_feature  
<222> (16)..(16)  
<223> iANB-meC está en posición 16.

55 <220>  
<221> misc\_feature  
<222> (19)..(19)  
<223> iANB-meC está en posición 19.

60 <220>  
<221> misc\_feature  
<222> (22)..(22)  
<223> iANB-A está en posición 22.

<220>

<221> misc\_feature  
 <222> (24)..(24)  
 <223> iANB-meC está en posición 24.

5      <220>  
 <221> misc\_feature  
 <222> (26)..(26)  
 <223> iANB-meC está en posición 26.

10     <220>  
 <221> misc\_feature  
 <222> (27)..(27)  
 <223> iANB-meC está en posición 27.

15     <220>  
 <221> misc\_feature  
 <222> (31)..(31)  
 <223> iANB-meC está en posición 31.

20     <220>  
 <221> misc\_feature  
 <222> (33)..(33)  
 <223> iANB-meC está en posición 33.

25     <220>  
 <221> modified\_base  
 <222> (34)..(41)  
 <223> I

30     <220>  
 <221> misc\_feature  
 <222> (34)..(41)  
 <223> n es a, c, g o t

35     <220>  
 <221> misc\_feature  
 <222> (42)..(42)  
 <223> iANB-A está en posición 42.

40     <220>  
 <221> misc\_feature  
 <222> (44)..(44)  
 <223> iANB-meC está en posición 44.

45     <220>  
 <221> misc\_feature  
 <222> (46)..(46)  
 <223> iANB-meC está en posición 46.

50     <220>  
 <221> misc\_feature  
 <222> (49)..(49)  
 <223> iANB-A está en posición 49.

55     <220>  
 <221> misc\_feature  
 <222> (52)..(52)  
 <223> iANB-meC está en posición 52.

60     <220>  
 <221> misc\_feature  
 <222> (53)..(53)  
 <223> iANB-meC está en posición 53.

65     <220>  
 <221> misc\_feature

	<222> (56)..(56)	
	<223> iANB-meC está en posición 56.	
5	<220> <221> misc_feature <222> (59)..(59) <223> iANB-meC está en posición 59.	
10	<220> <221> misc_feature <222> (62)..(62) <223> iANB-meC está en posición 62.	
	<400> 16 gatcggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgat gccgtcttct	60
15	gcttg	65
20	<210> 17 <211> 64 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
25	<220> <221> misc_feature <222> (25)..(30) <223> N = A, G, C, T	
30	<400> 17 caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg	60
	atct	64
35	<210> 18 <211> 64 <212> ADN <213> Secuencia artificial	
40	<220> <223> Oligonucleótido sintético	
45	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
50	<220> <221> misc_feature <222> (12)..(12) <223> iANB-meC está en posición 12.	
55	<220> <221> misc_feature <222> (15)..(15) <223> iANB-meC está en posición 15.	
60	<220> <221> misc_feature <222> (19)..(19) <223> iANB-meC está en posición 19.	
	<220> <221> misc_feature	

	<222> (25)..(30)	
	<223> N = A G, C, T	
5	<220> <221> misc_feature <222> (35)..(35) <223> iANB-meC está en posición 35.	
10	<220> <221> misc_feature <222> (43)..(43) <223> iANB-meC está en posición 43.	
15	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.	
20	<220> <221> misc_feature <222> (53)..(53) <223> iANB-meC está en posición 53.	
25	<220> <221> misc_feature <222> (58)..(58) <223> iANB-meC está en posición 58.	
30	<220> <221> misc_feature <222> (59)..(59) <223> iANB-meC está en posición 59.	
	<400> 18 caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg	60
35	atct	64
	<210> 19 <211> 64 <212> ADN <213> Secuencia artificial	
40	<220> <223> Oligonucleótido sintético	
45	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
50	<220> <221> misc_feature <222> (12)..(12) <223> iANB-meC está en posición 12.	
55	<220> <221> misc_feature <222> (15)..(15) <223> iANB-meC está en posición 15.	
60	<220> <221> misc_feature <222> (17)..(17) <223> iANB-T está en posición 17.	

5  
 <220>  
 <221> misc\_feature  
 <222> (19)..(19)  
 <223> iANB-meC está en posición 19.

10  
 <220>  
 <221> misc\_feature  
 <222> (24)..(24)  
 <223> iANB-T está en posición 24.

15  
 <220>  
 <221> misc\_feature  
 <222> (25)..(30)  
 <223> N = A, G, C, T

20  
 <220>  
 <221> misc\_feature  
 <222> (32)..(32)  
 <223> iANB-T está en posición 32.

25  
 <220>  
 <221> misc\_feature  
 <222> (35)..(35)  
 <223> iANB-meC está en posición 35.

30  
 <220>  
 <221> misc\_feature  
 <222> (36)..(36)  
 <223> iANB-T está en posición 36.

35  
 <220>  
 <221> misc\_feature  
 <222> (41)..(41)  
 <223> iANB-T está en posición 41.

40  
 <220>  
 <221> misc\_feature  
 <222> (42)..(42)  
 <223> iANB-T está en posición 42.

45  
 <220>  
 <221> misc\_feature  
 <222> (43)..(43)  
 <223> iANB-meC está en posición 43.

50  
 <220>  
 <221> misc\_feature  
 <222> (47)..(47)  
 <223> iANB-meC está en posición 47.

55  
 <220>  
 <221> misc\_feature  
 <222> (49)..(49)  
 <223> iANB-T está en posición 49.

60  
 <220>  
 <221> misc\_feature  
 <222> (51)..(51)  
 <223> iANB-T está en posición 51.

65  
 <220>  
 <221> misc\_feature  
 <222> (53)..(53)  
 <223> iANB-meC está en posición 53.

<220>

	<221> misc_feature	
	<222> (56)..(56)	
	<223> iANB-T está en posición 56.	
5	<220>	
	<221> misc_feature	
	<222> (58)..(58)	
	<223> iANB-meC está en posición 58.	
10	<220>	
	<221> misc_feature	
	<222> (59)..(59)	
	<223> iANB-meC está en posición 59.	
15	<220>	
	<221> misc_feature	
	<222> (62)..(62)	
	<223> iANB-T está en posición 62.	
20	<400> 19	
	caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg	60
	atct	64
	<210> 20	
	<211> 64	
25	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
30	<220>	
	<221> misc_feature	
	<222> (25)..(30)	
	<223> N = A, G, C, T	
35	<220>	
	<221> misc_feature	
	<222> (35)..(40)	
	<223> n es a, c, g o t	
40	<400> 20	
	agatcggaag agcacacgtc tgaactccag tcacnnnnnn atctcgtatg ccgtcttctg	60
	cttg	64
	<210> 21	
	<211> 64	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
	<220>	
	<221> misc_feature	
	<222> (5)..(5)	
55	<223> iANB-meC está en posición 5.	
	<220>	
	<221> misc_feature	
	<222> (13)..(13)	
60	<223> iANB-meC está en posición 13.	
	<220>	

	<221> misc_feature	
	<222> (17)..(17)	
	<223> iANB-meC está en posición 17.	
5	<220>	
	<221> misc_feature	
	<222> (25)..(25)	
	<223> iANB-meC está en posición 25.	
10	<220>	
	<221> misc_feature	
	<222> (32)..(32)	
	<223> iANB-meC está en posición 32.	
15	<220>	
	<221> misc_feature	
	<222> (34)..(34)	
	<223> iANB-meC está en posición 34.	
20	<220>	
	<221> misc_feature	
	<222> (35)..(40)	
	<223> N = A, G, C, T	
25	<220>	
	<221> misc_feature	
	<222> (45)..(45)	
	<223> iANB-meC está en posición 45.	
30	<220>	
	<221> misc_feature	
	<222> (52)..(52)	
	<223> iANB-meC está en posición 52.	
35	<220>	
	<221> misc_feature	
	<222> (55)..(55)	
	<223> iANB-meC está en posición 55.	
40	<220>	
	<221> misc_feature	
	<222> (61)..(61)	
	<223> iANB-meC está en posición 61.	
45	<400> 21	
	agatcggaag agcacacgtc tgaactccag tcacnnnnnn atctcgatg ccgtcttctg	60
	cttg	64
	<210> 22	
	<211> 64	
50	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
55	<220>	
	<221> misc_feature	
	<222> (4)..(4)	
	<223> iANB-T está en posición 4.	
60	<220>	
	<221> misc_feature	
	<222> (5)..(5)	
	<223> iANB-meC está en posición 5.	

5	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC está en posición 13.
10	<220> <221> misc_feature <222> (15)..(5) <223> iANB-meC está en posición 15.
15	<220> <221> misc_feature <222> (17)..(17) <223> iANB-meC está en posición 17.
20	<220> <221> misc_feature <222> (20)..(20) <223> iANB-meC está en posición 20.
25	<220> <221> misc_feature <222> (25)..(25) <223> iANB-meC está en posición 25.
30	<220> <221> misc_feature <222> (27)..(27) <223> iANB-meC está en posición 27.
35	<220> <221> misc_feature <222> (28)..(28) <223> iANB-meC está en posición 28.
40	<220> <221> misc_feature <222> (32)..(32) <223> iANB-meC está en posición 32.
45	<220> <221> misc_feature <222> (34)..(34) <223> iANB-meC está en posición 34.
50	<220> <221> misc_feature <222> (35)..(40) <223> N = A,, G, C, T
55	<220> <221> misc_feature <222> (43)..(43) <223> iANB-meC está en posición 43.
60	<220> <221> misc_feature <222> (45)..(45) <223> iANB-meC está en posición 45.
65	<220> <221> misc_feature <222> (47)..(47) <223> iANB-T está en posición 47.

5	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (49)..(49)          &lt;223&gt; iANB-T está en posición 49.</p>	
10	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (51)..(51)          &lt;223&gt; iANB-meC está en posición 51.</p>	
15	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (52)..(52)          &lt;223&gt; iANB-meC está en posición 52.</p>	
20	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (55)..(55)          &lt;223&gt; iANB-meC está en posición 55.</p>	
25	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (58)..(58)          &lt;223&gt; iANB-meC está en posición 58.</p>	
30	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (61)..(61)          &lt;223&gt; iANB-meC está en posición 61.</p>	
	<p>&lt;400&gt; 22          agatcggaag agcacacgtc tgaactccag tcacnnnnnn atctcgatg ccgtcttctg          cttg</p>	<p>60          64</p>
35	<p>&lt;210&gt; 23          &lt;211&gt; 58          &lt;212&gt; ADN          &lt;213&gt; Secuencia artificial</p>	
40	<p>&lt;220&gt;          &lt;223&gt; Oligonucleótido sintético</p>	
	<p>&lt;400&gt; 23          aatgatacgg cgaccaccga gatctacact cttccctac acgacgctct tccgatct</p>	58
45	<p>&lt;210&gt; 24          &lt;211&gt; 58          &lt;212&gt; ADN          &lt;213&gt; Secuencia artificial</p>	
50	<p>&lt;220&gt;          &lt;223&gt; Oligonucleótido sintético</p>	
55	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (8)..(8)          &lt;223&gt; iANB-meC está en posición 8.</p>	
60	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (11)..(11)          &lt;223&gt; iANB-meC está en posición 11.</p>	
	<p>&lt;220&gt;          &lt;221&gt; misc_feature</p>	

	<222> (14)..(14)	
	<223> iANB-meC está en posición 14.	
5	<220> <221> misc_feature <222> (17)..(17) <223> iANB-meC está en posición 17.	
10	<220> <221> misc_feature <222> (24)..(24) <223> iANB-meC está en posición 24.	
15	<220> <221> misc_feature <222> (29)..(29) <223> iANB-meC está en posición 29.	
20	<220> <221> misc_feature <222> (36)..(36) <223> iANB-meC está en posición 36.	
25	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC está en posición 50.	
30	<220> <221> misc_feature <222> (54)..(54) <223> iANB-meC está en posición 54.	
35	<220> <221> misc_feature <222> (58)..(58) <223> iANB-meC está en posición 58.	
40	<400> 24 aatgatacgg cgaccaccga gatctacact ctttcctac acgacgctct tccgatct	58
45	<210> 25 <211> 58 <212> ADN <213> Secuencia artificial	
50	<220> <223> oligonucleótido sintético	
55	<220> <221> misc_feature <222> (8)..(8) <223> iANB-meC está en posición 8.	
60	<220> <221> misc_feature <222> (11)..(11) <223> iANB-meC está en posición 11.	
65	<220> <221> misc_feature	

	<222> (17)..(17)
	<223> iANB-meC está en posición 17.
5	<220> <221> misc_feature <222> (24)..(24) <223> iANB-meC está en posición 24.
10	<220> <221> misc_feature <222> (27)..(27) <223> iANB-meC está en posición 27.
15	<220> <221> misc_feature <222> (29)..(29) <223> iANB-meC está en posición 29.
20	<220> <221> misc_feature <222> (31)..(31) <223> iANB-meC está en posición 31.
25	<220> <221> misc_feature <222> (35)..(35) <223> iANB-meC está en posición 35.
30	<220> <221> misc_feature <222> (37)..(37) <223> iANB-meC está en posición 37.
35	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.
40	<220> <221> misc_feature <222> (42)..(42) <223> iANB-meC está en posición 42.
45	<220> <221> misc_feature <222> (45)..(45) <223> iANB-meC está en posición 45.
50	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.
55	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC está en posición 49.
60	<220> <221> misc_feature <222> (53)..(53) <223> iANB-meC está en posición 53.
65	<220> <221> misc_feature <222> (57)..(57)

	<223> iANB-meC está en posición 57.	
5	<400> 25 aatgatacgg cgaccaccga gatctacact cttccctac acgacgctct tccgatct	58
	<210> 26 <211> 58 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
15	<400> 26 agatcggaag agcgcgtgt agggaaagag tgtagatctc ggtggtgcc gtatcatt	58
	<210> 27 <211> 58 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
25	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
30	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC está en posición 13.	
35	<220> <221> misc_feature <222> (16)..(16) <223> iANB-meC está en posición 16.	
40	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC está en posición 38.	
45	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.	
50	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.	
55	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC está en posición 49.	
60	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC está en posición 50.	
65	<220> <221> misc_feature	

	<222> (55)..(55)	
	<223> iANB-meC está en posición 55.	
5	<400> 27 agatcggaag agcgtcgtgt agggaaagag ttagatctc ggtggcgcc gtatcatt	58
	<210> 28	
	<211> 58	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
15	<220>	
	<221> misc_feature	
	<222> (3)..(3)	
	<223> iANB-A está en posición 3.	
20	<220>	
	<221> misc_feature	
	<222> (5)..(5)	
	<223> iANB-meC está en posición 5.	
25	<220>	
	<221> misc_feature	
	<222> (8)..(8)	
	<223> iANB-A está en posición 8.	
30	<220>	
	<221> misc_feature	
	<222> (11)..(11)	
	<223> iANB-A está en posición 11.	
35	<220>	
	<221> misc_feature	
	<222> (13)..(13)	
	<223> iANB-meC está en posición 13.	
40	<220>	
	<221> misc_feature	
	<222> (16)..(16)	
	<223> iANB-meC está en posición 16.	
45	<220>	
	<221> misc_feature	
	<222> (21)..(21)	
	<223> iANB-A está en posición 21.	
50	<220>	
	<221> misc_feature	
	<222> (25)..(25)	
	<223> iANB-A está en posición 25.	
55	<220>	
	<221> misc_feature	
	<222> (27)..(27)	
	<223> iANB-A está en posición 27.	
60	<220>	
	<221> misc_feature	
	<222> (29)..(29)	
	<223> iANB-A está en posición 29.	
65	<220>	
	<221> misc_feature	

	<222> (30)..(30)	
	<223> iANB-meC está en posición 30.	
5	<220> <221> misc_feature <222> (34)..(34) <223> iANB-A está en posición 34.	
10	<220> <221> misc_feature <222> (36)..(36) <223> iANB-A está en posición 36.	
15	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.	
20	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.	
25	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC está en posición 49.	
30	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC está en posición 50.	
35	<220> <221> misc_feature <222> (53)..(53) <223> iANB-A está en posición 53.	
40	<220> <221> misc_feature <222> (55)..(55) <223> iANB-meC está en posición 55.	
45	<220> <221> misc_feature <222> (56)..(56) <223> iANB-A está en posición 56.	
50	<400> 28 agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58
	<210> 29 <211> 41 <212> ADN	
55	<213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
60	<400> 29 ccactacgcc tccgctttcc tctctatggg cagtcggtga t	41
	<210> 30 <211> 41 <212> ADN	
65	<213> Secuencia artificial	

	<220>		
	<223>	Oligonucleótido sintético	
5	<220>		
	<221>	misc_feature	
	<222>	(2)..(2)	
	<223>	iANB-meC está en posición 2.	
10	<220>		
	<221>	misc_feature	
	<222>	(4)..(4)	
	<223>	iANB-meC está en posición 4.	
15	<220>		
	<221>	misc_feature	
	<222>	(7)..(7)	
	<223>	iANB-meC está en posición 7.	
20	<220>		
	<221>	misc_feature	
	<222>	(10)..(10)	
	<223>	iANB-meC está en posición 10.	
25	<220>		
	<221>	misc_feature	
	<222>	(13)..(13)	
	<223>	iANB-meC está en posición 13.	
30	<220>		
	<221>	misc_feature	
	<222>	(15)..(15)	
	<223>	iANB-meC está en posición 15.	
35	<220>		
	<221>	misc_feature	
	<222>	(20)..(20)	
	<223>	iANB-meC está en posición 20.	
40	<220>		
	<221>	misc_feature	
	<222>	(22)..(22)	
	<223>	iANB-meC está en posición 22.	
45	<220>		
	<221>	misc_feature	
	<222>	(24)..(24)	
	<223>	iANB-meC está en posición 24.	
50	<220>		
	<221>	misc_feature	
	<222>	(31)..(31)	
	<223>	iANB-meC está en posición 31.	
55	<220>		
	<221>	misc_feature	
	<222>	(35)..(35)	
	<223>	iANB-meC está en posición 35.	
60	<400>	30	
		ccactacgcc tccgctttcc tctctatggg cagtcggtga t	41
	<210>	31	
	<211>	43	
65	<212>	ADN	
	<213>	Secuencia artificial	

	<220>	
	<223> Oligonucleótido sintético	
5	<400> 31 atcaccgact gcccatagag aggaaagcgg aggcgtagtg gtt	43
	<210> 32	
	<211> 43	
10	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
15	<220>	
	<221> misc_feature	
	<222> (3)..(3)	
	<223> iANB-meC está en posición 3.	
20	<220>	
	<221> misc_feature	
	<222> (6)..(6)	
	<223> iANB-meC está en posición 6.	
25	<220>	
	<221> misc_feature	
	<222> (9)..(9)	
	<223> iANB-meC está en posición 9.	
30	<220>	
	<221> misc_feature	
	<222> (13)..(13)	
	<223> iANB-meC está en posición 13.	
35	<220>	
	<221> misc_feature	
	<222> (15)..(15)	
	<223> iANB-A está en posición 15.	
40	<220>	
	<221> misc_feature	
	<222> (17)..(17)	
	<223> iANB-A está en posición 17.	
45	<220>	
	<221> misc_feature	
	<222> (19)..(19)	
	<223> iANB-A está en posición 19.	
50	<220>	
	<221> misc_feature	
	<222> (21)..(21)	
	<223> iANB-A está en posición 21.	
55	<220>	
	<221> misc_feature	
	<222> (25)..(25)	
	<223> iANB-A está en posición 25.	
60	<220>	
	<221> misc_feature	
	<222> (26)..(26)	
	<223> iANB-A está en posición 26.	
65	<220>	

<221> misc\_feature  
 <222> (28)..(28)  
 <223> iANB-meC está en posición 28.  
 5 <220>  
 <221> misc\_feature  
 <222> (31)..(31)  
 <223> iANB-A está en posición 31.  
 10 <220>  
 <221> misc\_feature  
 <222> (34)..(34)  
 <223> iANB-meC está en posición 34.  
 15 <220>  
 <221> misc\_feature  
 <222> (37)..(37)  
 <223> iANB-A está en posición 37.  
 20 <400> 32  
 atcaccgact gcccatagag aggaaagcgg aggcgtagtg gtt 43  
 <210> 33  
 <211> 30  
 25 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Oligonucleótido sintético  
 30 <400> 33  
 ccatctcatc cctgcgtgtc tccgactcag 30  
 <210> 34  
 <211> 30  
 35 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Oligonucleótido sintético  
 40 <220>  
 <221> misc\_feature  
 <222> (2)..(2)  
 45 <223> iANB-meC está en posición 2.  
 <220>  
 <221> misc\_feature  
 <222> (5)..(5)  
 50 <223> iANB-meC está en posición 5.  
 <220>  
 <221> misc\_feature  
 <222> (7)..(7)  
 55 <223> iANB-meC está en posición 7.  
 <220>  
 <221> misc\_feature  
 <222> (10)..(10)  
 60 <223> iANB-meC está en posición 10.  
 <220>  
 <221> misc\_feature  
 <222> (11)..(11)  
 65 <223> iANB-meC está en posición 11.

5 <220>  
 <221> misc\_feature  
 <222> (15)..(15)  
 <223> iANB-meC está en posición 15.

10 <220>  
 <221> misc\_feature  
 <222> (20)..(20)  
 <223> iANB-meC está en posición 20.

15 <220>  
 <221> misc\_feature  
 <222> (22)..(22)  
 <223> iANB-meC está en posición 22.

20 <220>  
 <221> misc\_feature  
 <222> (23)..(23)  
 <223> iANB-meC está en posición 23.

25 <220>  
 <221> misc\_feature  
 <222> (26)..(26)  
 <223> iANB-meC está en posición 26.

30 <220>  
 <221> misc\_feature  
 <222> (28)..(28)  
 <223> iANB-meC está en posición 28.

<400> 34  
 ccatctcatc cctgcgtgtc tccgactcag 30

35 <210> 35  
 <211> 32  
 <212> ADN  
 <213> Secuencia artificial

40 <220>  
 <223> Oligonucleótido sintético

<400> 35  
 ctgagtcgga gacacgcagg gatgagatgg tt 32

45 <210> 36  
 <211> 32  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <223> Oligonucleótido sintético

55 <220>  
 <221> misc\_feature  
 <222> (1)..(1)  
 <223> iANB-meC está en posición 1.

60 <220>  
 <221> misc\_feature  
 <222> (4)..(4)  
 <223> iANB-A está en posición 4.

65 <220>  
 <221> misc\_feature  
 <222> (7)..(7)  
 <223> iANB-meC está en posición 7.

5 <220>  
 <221> misc\_feature  
 <222> (10)..(10)  
 <223> iANB-A está en posición 10.

10 <220>  
 <221> misc\_feature  
 <222> (13)..(13)  
 <223> iANB-meC está en posición 13.

15 <220>  
 <221> misc\_feature  
 <222> (17)..(17)  
 <223> iANB-meC está en posición 17.

20 <220>  
 <221> misc\_feature  
 <222> (22)..(22)  
 <223> iANB-A está en posición 22.

25 <220>  
 <221> misc\_feature  
 <222> (25)..(25)  
 <223> iANB-A está en posición 25.

30 <220>  
 <221> misc\_feature  
 <222> (27)..(27)  
 <223> iANB-A está en posición 27.

35 <400> 36  
 ctgagtcgga gacacgcagg gatgagatgg tt 32

40 <210> 37  
 <211> 120  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> Oligonucleótido sintético

50 <400> 37  
 ccattgtgtg tgagcaaagg caccctgtcc agtctaacct gaatctctgt aggaagaggc 60  
 gtgcggctct actacatcgg aggggaggtc ttgcgcagagt gcctcagtga cagcgctatt 120

55 <210> 38  
 <211> 120  
 <212> ADN  
 <213> Secuencia artificial

60 <220>  
 <223> Oligonucleótido sintético

65 <400> 38  
 ctgtccagtc taacctgaat ctctgtagga agaggcgtgc ggctctacta catcggaggg 60  
 gaggtcttcg cagagtgcct cagtgcacgc gctatttttg tccagtctcc caactgtaac 120

70 <210> 39  
 <211> 120  
 <212> ADN  
 <213> Secuencia artificial

75 <220>

	<223> Oligonucleótido sintético	
	<400> 39	
	gtaggaagag gcgtagcggt ctactacatc ggaggggagg tcttcgcaga gtgcctcagt	60
5	gacagcgcta tttttgtcca gtctcccaac tgtaaccagc gctatggctg gcacccggcc	120
	<210> 40	
	<211> 120	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
	<400> 40	
	tacatcggag gggaggtctt cgcagagtgc ctcaagtaca gcgctatattt tgtccagtct	60
15	cccaactgta accagcgcta tggctggcac ccggccaccg tctgcaagat cccaccaggt	120
	<210> 41	
	<211> 120	
	<212> ADN	
20	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
25	<400> 41	
	gagtgcctca gtgacagcgc tatttttgtc cagtctccca actgtaacca gcgctatggc	60
	tggcaccgag ccaccgtctg caagatccca ccaggtaaac gagccgcaca ggcaccctg	120
30	<210> 42	
	<211> 120	
	<212> ADN	
	<213> Secuencia artificial	
35	<220>	
	<223> Oligonucleótido sintético	
	<400> 42	
	tttgtccagt ctcccaactg taaccagcgc tatggctggc acccgccac cgtctgcaag	60
40	atcccaccag gtaaacgagc cgcacaggca cccctgcctt gaggtccctc tccgagtga	120
	<210> 43	
	<211> 120	
	<212> ADN	
45	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	
	<400> 43	
	gacctggcca cttccatccc cacagccctg tttctgtgtt tttggcagga tgcaacctga	60
50	agatcttcaa caaccaggag ttcgctgcc tcttgccca gtcggtcaac cagggtttg	120
	<210> 44	
	<211> 120	
	<212> ADN	
55	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido sintético	

# ES 2 769 796 T3

	<400> 44 gccctgtttc tgtgtttttg gcaggatgca acctgaagat cttcaacaac caggagttcg	60
	ctgcctcct ggcccagtcg gtcaaccagg gctttgagcg tgtctaccag ttgacccgaa	120
5	<210> 45 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 45 gatgcaacct gaagatcttc aacaaccagg agttcgctgc cctcctggcc cagtcgggtca	60
15	accagggctt tgaggctgtc taccagttga cccgaatgtg caccatccgc atgagcttcg	120
	<210> 46 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 46 accaggagtt cgctgccctc ctggcccagt cgggtcaacca gggctttgag gctgtctacc	60
25	agttgacccg aatgtgcacc atccgcatga gcttcgtcaa aggctgggga gcggagtaca	120
	<210> 47 <211> 120 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
35	<400> 47 cccagtcggt caaccagggc tttgaggctg tctaccagtt gacccgaatg tgcaccatcc	60
	gcatgagctt cgtcaaaggc tggggagcgg agtacaggtc agttatgggt gctgcctaca	120
	<210> 48 <211> 120 <212> ADN <213> Secuencia artificial	
40	<220> <223> Oligonucleótido sintético	
45	<400> 48 aggctgtcta ccagttgacc cgaatgtgca ccatccgcat gagcttcgtc aaaggctggg	60
	gagcggagta caggtcagtt atgggtgctg cctacatcag gggacccaac tccaggtgac	120
	<210> 49 <211> 120 <212> ADN <213> Secuencia artificial	
50	<220> <223> Oligonucleótido sintético	
55		

# ES 2 769 796 T3

	<400> 49	
	tgtaaccccc tggagatddd ttaagtcccc caccaccccc ctttccctat ttcttacagg	60
	agacagactg tgaccagtac cccctgctgg attgagctgc acctgaatgg gcctttgcag	120
5	<210> 50 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 50	
	gtccccacc ccacccttt ccctatttct tacaggagac agactgtgac cagtaccccc	60
15	tgctggattg agctgcacct gaatgggcct ttgcagtggc ttgacaaggt cctcaccag	120
20	<210> 51 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
25	<400> 51	
	atttcttaca ggagacagac tgtgaccagt accccctgct ggattgagct gcacctgaat	60
	gggcctttgc agtggcttga caaggtcctc acccagatgg gctccccaag catccgctgt	120
30	<210> 52 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
35	<400> 52	
	accagtaccc cctgctggat tgagctgcac ctgaatgggc ctttgcagtg gcttgacaag	60
	gtcctcacc agatgggctc cccaagcatc cgctgttcca gtgtgtotta gagacatcaa	120
40	<210> 53 <211> 120 <212> ADN <213> Secuencia artificial	
45	<220> <223> Oligonucleótido sintético	
	<400> 53	
	ctgcacctga atgggccttt gcagtggctt gacaagggtc tcaccagat gggtcccca	60
	agcatccgct gttccagtgt gtcttagaga catcaagtat ggtaggggag ggcaggcttg	120
50	<210> 54 <211> 120 <212> ADN <213> Secuencia artificial	
55	<220> <223> Oligonucleótido sintético	
	<400> 54	

# ES 2 769 796 T3

	tggcttgaca aggtcctcac ccagatgggc tccccaaagca tccgctgttc cagtgtgtct	60
	tagagacatc aagtatggta ggggagggca ggcttgggga aaatggccat gcaggaggtg	120
5	<210> 55 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 55 cgtcgcgcgc caacgcgcgc atggcctccg gagcccgggg tccccaggcc gcgccggccc	60
	agccctgcga tgccgcctgg agcggcgcg ctcgcgtgc aggtggctct cttaaggatg	120
15	<210> 56 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 56 cgtctcacgc caacgcaagc atgtcctccg gagcccgggg tccccaggcc gcgccggccc	60
	agccctgcga tgccgcctgg agcggcgcg ctcgcactgc agatggctct cttaaggatg	120
25	<210> 57 <211> 120 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
	<400> 57 taccgagcag cggcagctgg ccgccgtcgc gcgccaacgc cggcatggcc tccggagccc	60
35	ggggtcccca ggccgcgcgc gccagccct gcgatgccgc ctggagcggc gcgcctcgcg	120
40	<210> 58 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 58 taccgagcag cggcagctgg ccgccgtcgc gcgccaacgc cggcatggcc tccggagccc	60
45	ggggtcccca ggccgcgcgc gccagccct gcgatgccgc cttgagcaac gcgcctcacg	120
50	<210> 59 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
55	<400> 59 gctgcgagcg agcgcgcggg gccttaccga gcagcggcag ctggccgcgc tcgcgcgcga	60

	acgccggcat ggcctccgga gcccggggtc cccaggccgc gccggcccag ccctgcatg	120
5	<210> 60 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 60 gcttcgagag agcgagcggg gccttaccga gcagcagcag ctggccgccg tcgcgcgcca	60
	acgccggcat ggcctccgga gcccggggtc cccaggccgc gccggcccag ccctgagatg	120
15	<210> 61 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 61 gtgggggctg agggaccgcg aggggctgcg agcgagcgcg cggggcctta ccgagcagcg	60
25	gcagctggcc gccgtcgcgc gccaacgcgc gcatggcctc cggagcccag ggtcccag	120
30	<210> 62 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 62 gaggtggctg agagaccgcg aggagctgcg agcgagcgcg cggggcctta ccgagcagcg	60
35	gcagctggcc gccgtcgcgc gccaacgcag gcatggcctc cggagcccag ggtcccag	120
40	<210> 63 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
45	<400> 63 cgaggcggct gggccggagg aggcgcgcgc ccgggtccac actgcggggt gggggctgag	60
	ggaccgcgag gggctgcgag cgagcgagcg gggccttacc gagcagcggc agctggccgc	120
50	<210> 64 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
55	<400> 64	

# ES 2 769 796 T3

	cgaggcggct gggccggagg aggcgcgcgc ccgcatccac actgcggggt gggggctgag	60
	ggaccgcgag gggctgcgag cgagcgagcg gggacttacc gagcagcggc aactggacgc	120
5	<210> 65 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 65 gcgcgcccg gtccacactg cggggtgggg gctgagggac cgcgaggggc tgcgagcgag	60
	cgagcggggc cttaccgagc agcggcagct ggccgccgtc gcgcgccaac gccggcatgg	120
15	<210> 66 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 66 gcacgcacgg atccacactg cggggtgggg gctgagggac cgcgaggagc tgcgagcgag	60
	cgagcggggc cttaccgagc agcggcagct ggccgccgtc gcgcgccaac gccggcatgg	120
25	<210> 67 <211> 120 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
	<400> 67 tcgcaggcac agcgcggcgc cccgctgcat ctccggccgc tgcgcgtggg tccgaccga	60
35	gcggccgcgg ctccgggctg aaagtgtccg cgcgggcgcc ggctggcctg gggcggggcg	120
40	<210> 68 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
45	<400> 68 cacacacaca agcgcggcgc cccgctgcat ctccggccgc tgcgcgtggg tccgaccga	60
	gcggccgcgg ctccgggctg aaagtgtccg cgcgggcgcc ggctggcctg cacacacaca	120
50	<210> 69 <211> 120 <212> ADN <213> Secuencia artificial	
55	<220> <223> Oligonucleótido sintético	
	<400> 69	

# ES 2 769 796 T3

	ggcggagcgg tctcagcgcc cgccccaggt gcgcggtacc ccctccccgg ccagccccac	60
	gctcggggcgg gtggcccgtt cgccgcgctc accgtccagg agtcccaggc agagccacag	120
5	<210> 70 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 70 cacacacaca tctcagcgcc cgccccaggt gcgcggtacc ccctccccgg ccagccccac	60
	gctcggggcgg gtggcccgtt cgccgcgctc accgtccagg agtcccaggc cacacacaca	120
15	<210> 71 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 71 ccaggtgcgc ggtacccct ccccgccag cccacgctc gggcgggtgg cccgttcgcc	60
	gcgctcaccg tccaggagtc ccaggcagag ccacagtcgc aggcacagcg cggcgccccg	120
25	<210> 72 <211> 120 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
	<400> 72 cacacacaca ggtacccct ccccgccag cccacgctc gggcgggtgg cccgttcgcc	60
35	gcgctcaccg tccaggagtc ccaggcagag ccacagtcgc aggcacagcg cacacacaca	120
40	<210> 73 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 73 ggcccgttcg ccgcgctcac cgtccaggag tcccaggcag agccacagtc gcaggcacag	60
45	cgccggcgccc cgctgcatct ccggccgctg cgcgtgggtc cgacccgagc ggccgcggct	120
50	<210> 74 <211> 120 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
55	<400> 74	

# ES 2 769 796 T3

	cacacacaca ccgcgctcac cgtccaggag tcccaggcag agccacagtc gcaggcacag	60
	cgcgggcgccc cgctgcatct ccggccgctg cgcgtgggtc cgacccgagc cacacacaca	120
5	<210> 75 <211> 120 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 75 ggccagcccc acgctcgggc gggtggcccg ttcgccgcgc tcaccgtcca ggagtcccag	60
	gcagagccac agtcgcaggc acagcgcggc gcccgcgtgc atctccggcc gctgcgcgtg	120
15	<210> 76 <211> 120 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 76 cacacacaca acgctcgggc gggtggcccg ttcgccgcgc tcaccgtcca ggagtcccag	60
	gcagagccac agtcgcaggc acagcgcggc gcccgcgtgc atctccggcc cacacacaca	120
25	<210> 77 <211> 120 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
35	<400> 77 caggagtccc aggcagagcc acagtcgcag gcacagcgcg gcgccccgct gcatctccgg	60
	ccgctgcgcg tgggtccgac ccgagcggcc gcggctcggg gctgaaagtg tccgcgcggg	120
40	<210> 78 <211> 120 <212> ADN <213> Secuencia artificial	
45	<220> <223> Oligonucleótido sintético	
	<400> 78 cacacacaca aggcagagcc acagtcgcag gcacagcgcg gcgccccgct gcatctccgg	60
	ccgctgcgcg tgggtccgac ccgagcggcc gcggctcggg gctgaaagtg cacacacaca	120
50	<210> 79 <211> 23 <212> ADN <213> Secuencia artificial	
55	<220> <223> Oligonucleótido sintético	
	<400> 79 aatgatacgg cgaccaccga gat	23

5 <210> 80  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
  
 10 <400> 80  
 caagcagaag acggcatacg a 21  
  
 15 <210> 81  
 <211> 63  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
  
 20 <400> 81  
 gatcgggaaga gcacacgtct gaactccagt cacacagtga tctcgtatgc cgtcttctgc 60  
 ttg 63  
  
 25 <210> 82  
 <211> 64  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
  
 30 <220>  
 <221> modified\_base  
 <222> (25)..(30)  
 35 <223> I  
  
 <220>  
 <221> misc\_feature  
 <222> (25)..(30)  
 40 <223> n es a, c, g o t  
  
 <220>  
 <221> misc\_feature  
 <222> (64)..(64)  
 45 <223> 3SpC3 group modifica el extremo 3' de la posición 64.  
  
 <400> 82  
 caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg 60  
 atct 64  
  
 50 <210> 83  
 <211> 58  
 <212> ADN  
 <213> Secuencia artificial  
  
 55 <220>  
 <223> Oligonucleótido sintético  
  
 <220>  
 <221> misc\_feature  
 60 <222> (58)..(58)  
 <223> 3spc3 modifica el extremo 3' de la posición 58.

	<400> 83 aatgatacgg cgaccaccga gatctacact ctttcctac acgacgtct tccgatct	58
5	<210> 84 <211> 64 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
15	<220> <221> misc_feature <222> (1)..(1) <223> 5ANB-meC está en posición 1.	
20	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
25	<220> <221> misc_feature <222> (12)..(12) <223> iANB-meC está en posición 12.	
30	<220> <221> misc_feature <222> (15)..(15) <223> iANB-meC está en posición 15.	
35	<220> <221> misc_feature <222> (19)..(19) <223> iANB-meC está en posición 19.	
40	<220> <221> modified_base <222> (25)..(30) <223> I	
45	<220> <221> misc_feature <222> (25)..(30) <223> n es a, c, g, o t	
50	<220> <221> misc_feature <222> (35)..(35) <223> iANB-meC está en posición 35.	
55	<220> <221> misc_feature <222> (43)..(43) <223> iANB-meC está en posición 43.	
60	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.	
65	<220> <221> misc_feature <222> (53)..(53) <223> iANB-meC está en posición 53.	

	<220>		
	<221> misc_feature		
	<222> (55)..(55)		
5	<223> iANB-meC está en posición 55.		
	<220>		
	<221> misc_feature		
	<222> (58)..(58)		
10	<223> iANB-meC está en posición 58.		
	<220>		
	<221> misc_feature		
	<222> (59)..(59)		
15	<223> iANB-meC está en posición 59.		
	<220>		
	<221> misc_feature		
	<222> (63)..(63)		
20	<223> iANB-meC está en posición 63.		
	<220>		
	<221> misc_feature		
	<222> (64)..(64)		
25	<223> 3SpC3 está en posición 64.		
	<400> 84		
	caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg	60	
	atct	64	
30	<210> 85		
	<211> 58		
	<212> ADN		
	<213> Secuencia artificial		
35	<220>		
	<223> Oligonucleótido sintético		
	<220>		
	<221> misc_feature		
	<222> (8)..(8)		
40	<223> iANB-meC está en posición 8.		
	<220>		
	<221> misc_feature		
	<222> (11)..(11)		
45	<223> iANB-meC está en posición 11.		
	<220>		
	<221> misc_feature		
	<222> (14)..(14)		
50	<223> iANB-meC está en posición 14.		
	<220>		
	<221> misc_feature		
	<222> (18)..(18)		
55	<223> iANB-meC está en posición 18.		
	<220>		
	<221> misc_feature		
	<222> (24)..(24)		
60	<223> iANB-meC está en posición 24.		
	<220>		
	<221> misc_feature		
	<222> (27)..(27)		

	<223> iANB-meC está en posición 27.	
5	<220> <221> misc_feature <222> (31)..(31) <223> iANB-meC está en posición 31.	
10	<220> <221> misc_feature <222> (35)..(35) <223> iANB-meC está en posición 35.	
15	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.	
20	<220> <221> misc_feature <222> (45)..(45) <223> iANB-meC está en posición 45.	
25	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC está en posición 49.	
30	<220> <221> misc_feature <222> (53)..(53) <223> iANB-meC está en posición 53.	
35	<220> <221> misc_feature <222> (57)..(57) <223> iANB-meC está en posición 57.	
40	<220> <221> misc_feature <222> (58)..(58) <223> 3SpC3 está en posición 58.	
45	<400> 85 aatgatacgg cgaccaccga gatctacact ctttcctac acgacgctct tccgatct	58
50	<210> 86 <211> 64 <212> ADN <213> Secuencia artificial	
55	<220> <223> Oligonucleótido sintético	
60	<220> <221> misc_feature <222> (1)..(1) <223> 5ANB-meC está en posición 1.	
65	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC está en posición 5.	
	<220> <221> misc_feature <222> (9)..(9)	

<223> iANB-A está en posición 9.  
 <220>  
 <221> misc\_feature  
 5 <222> (12)..(12)  
 <223> iANB-meC está en posición 12.  
 <220>  
 <221> misc\_feature  
 10 <222> (15)..(15)  
 <223> iANB-meC está en posición 15.  
 <220>  
 <221> misc\_feature  
 15 <222> (19)..(19)  
 <223> iANB-meC está en posición 19.  
 <220>  
 <221> misc\_feature  
 20 <222> (21)..(21)  
 <223> iANB-A está en posición 21.  
 <220>  
 <221> misc\_feature  
 25 <222> (23)..(23)  
 <223> iANB-A está en posición 23.  
 <220>  
 <221> modified\_base  
 30 <222> (25)..(30)  
 <223> I  
 <220>  
 <221> misc\_feature  
 35 <222> (25)..(30)  
 <223> n es a, c, g o t  
 <220>  
 <221> misc\_feature  
 40 <222> (34)..(34)  
 <223> iANB-A está en posición 34.  
 <220>  
 <221> misc\_feature  
 45 <222> (35)..(35)  
 <223> iANB-meC está en posición 35.  
 <220>  
 <221> misc\_feature  
 50 <222> (39)..(39)  
 <223> iANB-A está en posición 39.  
 <220>  
 <221> misc\_feature  
 55 <222> (43)..(43)  
 <223> iANB-meC está en posición 43.  
 <220>  
 <221> misc\_feature  
 60 <222> (46)..(46)  
 <223> iANB-A está en posición 46.  
 <220>  
 <221> misc\_feature  
 65 <222> (47)..(47)  
 <223> iANB-meC está en posición 47.

5 <220>  
 <221> misc\_feature  
 <222> (53)..(53)  
 <223> iANB-meC está en posición 53.

10 <220>  
 <221> misc\_feature  
 <222> (55)..(55)  
 <223> iANB-meC está en posición 55.

15 <220>  
 <221> misc\_feature  
 <222> (58)..(58)  
 <223> iANB-meC está en posición 58.

20 <220>  
 <221> misc\_feature  
 <222> (59)..(59)  
 <223> iANB-meC está en posición 59.

25 <220>  
 <221> misc\_feature  
 <222> (61)..(61)  
 <223> iANB-A está en posición 61.

30 <220>  
 <221> misc\_feature  
 <222> (63)..(63)  
 <223> iANB-meC está en posición 63.

35 <220>  
 <221> misc\_feature  
 <222> (64)..(64)  
 <223> 3SpC3 modifica el extremo 3' en posición 64.

<400> 86  
 caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg 60  
 atct 64

40 <210> 87  
 <211> 58  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> Oligonucleótido sintético

50 <220>  
 <221> misc\_feature  
 <222> (8)..(8)  
 <223> iANB-meC está en posición 8.

55 <220>  
 <221> misc\_feature  
 <222> (11)..(11)  
 <223> iANB-meC está en posición 11.

60 <220>  
 <221> misc\_feature  
 <222> (14)..(14)  
 <223> iANB-meC está en posición 14.

<220>  
 <221> misc\_feature

	<222> (18)..(18)
	<223> iANB-meC está en posición 18.
5	<220> <221> misc_feature <222> (24)..(24) <223> iANB-meC está en posición 24.
10	<220> <221> misc_feature <222> (27)..(27) <223> iANB-meC está en posición 27.
15	<220> <221> misc_feature <222> (29)..(29) <223> iANB-meC está en posición 29.
20	<220> <221> misc_feature <222> (31)..(31) <223> iANB-meC está en posición 31.
25	<220> <221> misc_feature <222> (35)..(35) <223> iANB-meC está en posición 35.
30	<220> <221> misc_feature <222> (37)..(37) <223> iANB-meC está en posición 37.
35	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC está en posición 40.
40	<220> <221> misc_feature <222> (42)..(42) <223> iANB-meC está en posición 42.
45	<220> <221> misc_feature <222> (45)..(45) <223> iANB-meC está en posición 45.
50	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC está en posición 47.
55	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC está en posición 49.
60	<220> <221> misc_feature <222> (52)..(52) <223> iANB-meC está en posición 52.
65	<220> <221> misc_feature <222> (53)..(53)

<223> iANB-meC está en posición 53.

<220>  
<221> misc\_feature  
5 <222> (57)..(57)  
<223> iANB-meC está en posición 57.

<220>  
10 <221> misc\_feature  
<222> (58)..(58)  
<223> 3SpC3 modifica el extremo 3' en posición 58.

<400> 87  
15 aatgatacgg cgaccaccga gatctacact cttccctac acgacgctct tccgatct 58

<210> 88  
<211> 63  
<212> ADN  
20 <213> Secuencia artificial

<220>  
<223> Oligonucleótido sintético

<220>  
25 <221> modified\_base  
<222> (34)..(39)  
<223> I

<220>  
30 <221> misc\_feature  
<222> (34)..(39)  
<223> n es a, c, g o t

<220>  
35 <221> misc\_feature  
<222> (63)..(63)  
<223> 3SpC3 modifica el extremo 3' en posición 63.

<400> 88  
gatcggaaga gcacacgtct gaactccagt cacnnnnna tctcgtatgc cgtcttctgc 60

40 ttg 63

<210> 89  
<211> 58  
<212> ADN  
45 <213> Secuencia artificial

<220>  
<223> Oligonucleótido sintético

50 <220>  
<221> misc\_feature  
<222> (58)..(58)  
<223> 3SpC3 modifica el extremo 3' en posición 58.

55 <400> 89  
agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt 58

<210> 90  
<211> 63  
60 <212> ADN  
<213> Secuencia artificial

<220>  
<223> Oligonucleótido sintético

5	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (4)..(4)          &lt;223&gt; iANB-meC está en posición 4.</p>
10	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (12)..(12)          &lt;223&gt; iANB-meC está en posición 12.</p>
15	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (19)..(19)          &lt;223&gt; iANB-meC está en posición 19.</p>
20	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (24)..(24)          &lt;223&gt; iANB-meC está en posición 24.</p>
25	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (31)..(31)          &lt;223&gt; iANB-meC está en posición 31.</p>
30	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (33)..(33)          &lt;223&gt; iANB-meC está en posición 33.</p>
35	<p>&lt;220&gt;          &lt;221&gt; modified_base          &lt;222&gt; (34)..(39)          &lt;223&gt; I</p>
40	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (34)..(39)          &lt;223&gt; n es a, c, g o t</p>
45	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (42)..(42)          &lt;223&gt; iANB-meC está en posición 42.</p>
50	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (44)..(44)          &lt;223&gt; iANB-meC está en posición 44.</p>
55	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (51)..(51)          &lt;223&gt; iANB-meC está en posición 51.</p>
60	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (60)..(50)          &lt;223&gt; iANB-meC está en posición 60.</p>
65	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (63)..(63)          &lt;223&gt; 3SpC3 modifica el extremo 3' en posición 63.</p>

	<400> 90 gatcggaaga gcacacgtct gaactccagt cacnnnnna tctcgatatgc cgtttctgc	60
	ttg	63
5	<210> 91 <211> 58 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
15	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC localizado en posición 5.	
20	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC localizado en posición 13.	
25	<220> <221> misc_feature <222> (16)..(16) <223> iANB-meC localizado en posición 16.	
30	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC localizado en posición 38.	
35	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC localizado en posición 40.	
40	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC localizado en posición 47.	
45	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC localizado en posición 49.	
50	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC localizado en posición 50.	
55	<220> <221> misc_feature <222> (55)..(55) <223> iANB-meC localizado en posición 55.	
60	<220> <221> misc_feature <222> (58)..(58) <223> 3SpC3 modifica el extremo 3' en posición 58.	
	<400> 91 agatcggaag agcgctgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58

5	<210> 92
	<211> 63
	<212> ADN
	<213> Secuencia artificial
10	<220>
	<223> Oligonucleótido sintético
15	<220>
	<221> misc_feature
	<222> (4)..(4)
	<223> iANB-meC localizado en posición 4.
20	<220>
	<221> misc_feature
	<222> (12)..(12)
	<223> iANB-meC localizado en posición 12.
25	<220>
	<221> misc_feature
	<222> (14)..(14)
	<223> iANB-meC localizado en posición 14.
30	<220>
	<221> misc_feature
	<222> (16)..(16)
	<223> iANB-meC localizado en posición 16.
35	<220>
	<221> misc_feature
	<222> (19)..(19)
	<223> iANB-meC localizado en posición 19.
40	<220>
	<221> misc_feature
	<222> (24)..(24)
	<223> iANB-meC localizado en posición 24.
45	<220>
	<221> misc_feature
	<222> (26)..(26)
	<223> iANB-meC localizado en posición 26.
50	<220>
	<221> misc_feature
	<222> (27)..(27)
	<223> iANB-meC localizado en posición 27.
55	<220>
	<221> misc_feature
	<222> (31)..(31)
	<223> iANB-meC localizado en posición 31.
60	<220>
	<221> misc_feature
	<222> (33)..(33)
	<223> iANB-meC localizado en posición 33.
65	<220>
	<221> modified_base
	<222> (34)..(39)
	<223> I
	<220>
	<221> misc_feature
	<222> (34)..(39)

	<223> n es a, c, g o t	
	<220>	
5	<221> misc_feature	
	<222> (42)..(42)	
	<223> iANB-meC localizado en posición 42.	
	<220>	
10	<221> misc_feature	
	<222> (44)..(44)	
	<223> iANB-meC localizado en posición 44.	
	<220>	
15	<221> misc_feature	
	<222> (50)..(50)	
	<223> iANB-meC localizado en posición 50.	
	<220>	
20	<221> misc_feature	
	<222> (51)..(51)	
	<223> iANB-meC localizado en posición 51.	
	<220>	
25	<221> misc_feature	
	<222> (54)..(54)	
	<223> iANB-meC localizado en posición 54.	
	<220>	
30	<221> misc_feature	
	<222> (57)..(57)	
	<223> iANB-meC localizado en posición 57.	
	<220>	
35	<221> misc_feature	
	<222> (60)..(60)	
	<223> iANB-meC localizado en posición 60.	
	<400> 92	
	gatcggaaga gcacacgtct gaactccagt cacnnnnnna tctcgtatgc cgtcttctgc	60
	ttg	63
40		
	<210> 93	
	<211> 58	
	<212> ADN	
	<213> Secuencia artificial	
45		
	<220>	
	<223> Oligonucleótido sintético	
	<220>	
50	<221> misc_feature	
	<222> (3)..(3)	
	<223> iANB-A localizado en posición 3.	
	<220>	
55	<221> misc_feature	
	<222> (5)..(5)	
	<223> iANB-meC localizado en posición 5.	
	<220>	
60	<221> misc_feature	
	<222> (8)..(8)	
	<223> iANB-A localizado en posición 8.	
	<220>	

	<221> misc_feature <222> (9)..(9) <223> iANB-A localizado en posición 9.
5	<220> <221> misc_feature <222> (11)..(11) <223> iANB-A localizado en posición 11.
10	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC localizado en posición 13.
15	<220> <221> misc_feature <222> (16)..(16) <223> iANB-meC localizado en posición 16.
20	<220> <221> misc_feature <222> (21)..(21) <223> iANB-A localizado en posición 21.
25	<220> <221> misc_feature <222> (25)..(25) <223> iANB-A localizado en posición 25.
30	<220> <221> misc_feature <222> (26)..(26) <223> iANB-A localizado en posición 26.
35	<220> <221> misc_feature <222> (27)..(27) <223> iANB-A localizado en posición 27.
40	<220> <221> misc_feature <222> (29)..(29) <223> iANB-A localizado en posición 29.
45	<220> <221> misc_feature <222> (34)..(34) <223> iANB-A localizado en posición 34.
50	<220> <221> misc_feature <222> (36)..(36) <223> iANB-A localizado en posición 36.
55	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC localizado en posición 38.
60	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC localizado en posición 40.
65	<220> <221> misc_feature

	<222> (47)..(47)	
	<223> iANB-meC localizado en posición 47.	
5	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC localizado en posición 49.	
10	<220> <221> misc_feature <222> (50)..(50) <223> iANB-meC localizado en posición 50.	
15	<220> <221> misc_feature <222> (53)..(53) <223> iANB-A localizado en posición 53.	
20	<220> <221> misc_feature <222> (55)..(55) <223> iANB-meC localizado en posición 55.	
25	<220> <221> misc_feature <222> (56)..(56) <223> iANB-A localizado en posición 56.	
30	<220> <221> misc_feature <222> (58)..(58) <223> 3SpC3 modifica el extremo 3' en posición 58	
35	<400> 93 agatcggaag agcgcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58
40	<210> 94 <211> 58 <212> ADN <213> Secuencia artificial	
45	<220> <223> Oligonucleótido sintético	
50	<400> 94 agatcggaag agcgcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt	58
55	<210> 95 <211> 58 <212> ADN <213> Secuencia artificial	
60	<220> <223> Oligonucleótido sintético	
65	<220> <221> misc_feature <222> (5)..(5) <223> iANB-meC localizado en posición 5.	
	<220> <221> misc_feature <222> (13)..(13) <223> iANB-meC localizado en posición 13.	
	<220>	

<221> misc\_feature  
 <222> (16)..(16)  
 <223> iANB-meC localizado en posición 16.

5      <220>  
 <221> misc\_feature  
 <222> (38)..(38)  
 <223> iANB-meC localizado en posición 38.

10     <220>  
 <221> misc\_feature  
 <222> (40)..(40)  
 <223> iANB-meC localizado en posición 40.

15     <220>  
 <221> misc\_feature  
 <222> (47)..(47)  
 <223> iANB-meC localizado en posición 47.

20     <220>  
 <221> misc\_feature  
 <222> (50)..(50)  
 <223> iANB-meC localizado en posición 50.

25     <220>  
 <221> misc\_feature  
 <222> (55)..(55)  
 <223> iANB-meC localizado en posición 55.

30     <400> 95  
 agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggcgccc gtatcatt      58

<210> 96  
 <211> 58  
 <212> ADN  
 <213> Secuencia artificial

35     <220>  
 <223> Oligonucleótido sintético

40     <220>  
 <221> misc\_feature  
 <222> (3)..(3)  
 <223> iANB-A localizado en posición 3.

45     <220>  
 <221> misc\_feature  
 <222> (5)..(5)  
 <223> iANB-meC localizado en posición 5.

50     <220>  
 <221> misc\_feature  
 <222> (8)..(8)  
 <223> iANB-A localizado en posición 8.

55     <220>  
 <221> misc\_feature  
 <222> (9)..(9)  
 <223> iANB-A localizado en posición 9.

60     <220>  
 <221> misc\_feature  
 <222> (11)..(11)  
 <223> iANB-A localizado en posición 11.

65     <220>

	<221> misc_feature <222> (13)..(13) <223> iANB-meC localizado en posición 13.
5	<220> <221> misc_feature <222> (16)..(16) <223> iANB-meC localizado en posición 16.
10	<220> <221> misc_feature <222> (21)..(21) <223> iANB-A localizado en posición 21.
15	<220> <221> misc_feature <222> (25)..(25) <223> iANB-A localizado en posición 25.
20	<220> <221> misc_feature <222> (26)..(26) <223> iANB-A localizado en posición 26.
25	<220> <221> misc_feature <222> (27)..(27) <223> iANB-A localizado en posición 27.
30	<220> <221> misc_feature <222> (29)..(29) <223> iANB-A localizado en posición 29.
35	<220> <221> misc_feature <222> (34)..(34) <223> iANB-A localizado en posición 34.
40	<220> <221> misc_feature <222> (36)..(36) <223> iANB-A localizado en posición 36.
45	<220> <221> misc_feature <222> (38)..(38) <223> iANB-meC localizado en posición 38.
50	<220> <221> misc_feature <222> (40)..(40) <223> iANB-meC localizado en posición 40.
55	<220> <221> misc_feature <222> (47)..(47) <223> iANB-meC localizado en posición 47.
60	<220> <221> misc_feature <222> (49)..(49) <223> iANB-meC localizado en posición 49.
65	<220> <221> misc_feature

<222> (50)..(50)  
 <223> iANB-meC localizado en posición 50.

5      <220>  
       <221> misc\_feature  
       <222> (53)..(53)  
       <223> iANB-A localizado en posición 53.

10     <220>  
       <221> misc\_feature  
       <222> (55)..(55)  
       <223> iANB-meC localizado en posición 55.

15     <220>  
       <221> misc\_feature  
       <222> (56)..(56)  
       <223> iANB-A localizado en posición 56.

20     <400> 96  
       agatcggaag agcgtcgtgt agggaaagag tgtagatctc ggtggtcgcc gtatcatt      58

25     <210> 97  
       <211> 65  
       <212> ADN  
       <213> Secuencia artificial

30     <220>  
       <221> modified\_base  
       <222> (34)..(51)  
       <223> I

35     <220>  
       <221> misc\_feature  
       <222> (34)..(41)  
       <223> n es a, c, g, or t

40     <400> 97  
       gatcggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct      60  
       gcttg      65

45     <210> 98  
       <211> 65  
       <212> ADN  
       <213> Secuencia artificial

50     <220>  
       <223> Oligonucleótido sintético

55     <220>  
       <221> misc\_feature  
       <222> (4)..(4)  
       <223> iANB-meC localizado en posición 4.

60     <220>  
       <221> misc\_feature  
       <222> (12)..(12)  
       <223> iANB-meC localizado en posición 12.

60     <220>  
       <221> misc\_feature  
       <222> (24)..(24)  
       <223> iANB-meC localizado en posición 24.

5	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (31)..(31)          &lt;223&gt; iANB-meC localizado en posición 31.</p>	
10	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (33)..(33)          &lt;223&gt; iANB-meC localizado en posición 33.</p>	
15	<p>&lt;220&gt;          &lt;221&gt; modified_base          &lt;222&gt; (34)..(41)          &lt;223&gt; l</p>	
20	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (34)..(41)          &lt;223&gt; n es a, c, g o t</p>	
25	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (46)..(46)          &lt;223&gt; iANB-meC localizado en posición 46.</p>	
30	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (53)..(53)          &lt;223&gt; iANB-meC localizado en posición 53.</p>	
35	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (62)..(62)          &lt;223&gt; iANB-meC localizado en posición 62.</p>	
	<p>&lt;400&gt; 98          gatcgggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct</p>	60
	<p>gcttg</p>	65
40	<p>&lt;210&gt; 99          &lt;211&gt; 65          &lt;212&gt; ADN          &lt;213&gt; Secuencia artificial</p>	
45	<p>&lt;220&gt;          &lt;223&gt; Oligonucleótido sintético</p>	
50	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (4)..(4)          &lt;223&gt; iANB-meC localizado en posición 4.</p>	
55	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (12)..(12)          &lt;223&gt; iANB-meC localizado en posición 12.</p>	
60	<p>&lt;220&gt;          &lt;221&gt; misc_feature          &lt;222&gt; (14)..(14)          &lt;223&gt; iANB-meC localizado en posición 14.</p>	
	<p>&lt;220&gt;          &lt;221&gt; misc_feature</p>	

	<222> (16)..(16)
	<223> iANB-meC localizado en posición 16.
5	<220> <221> misc_feature <222> (19)..(19) <223> iANB-meC localizado en posición 19.
10	<220> <221> misc_feature <222> (24)..(24) <223> iANB-meC localizado en posición 24.
15	<220> <221> misc_feature <222> (26)..(26) <223> iANB-meC localizado en posición 26.
20	<220> <221> misc_feature <222> (27)..(27) <223> iANB-meC localizado en posición 27.
25	<220> <221> misc_feature <222> (31)..(31) <223> iANB-meC localizado en posición 31.
30	<220> <221> misc_feature <222> (33)..(33) <223> iANB-meC localizado en posición 33.
35	<220> <221> modified_base <222> (34)..(41) <223> l
40	<220> <221> misc_feature <222> (34)..(41) <223> n es a, c, g o t
45	<220> <221> misc_feature <222> (44)..(44) <223> iANB-meC localizado en posición 44.
50	<220> <221> misc_feature <222> (46)..(46) <223> iANB-meC localizado en posición 46.
55	<220> <221> misc_feature <222> (52)..(52) <223> iANB-meC localizado en posición 52.
60	<220> <221> misc_feature <222> (53)..(53) <223> iANB-meC localizado en posición 53.
65	<220> <221> misc_feature <222> (56)..(56)

	<223> iANB-meC localizado en posición 56.	
	<220>	
	<221> misc_feature	
5	<222> (59)..(59)	
	<223> iANB-meC localizado en posición 59.	
	<220>	
	<221> misc_feature	
10	<222> (62)..(62)	
	<223> iANB-meC localizado en posición 62.	
	<400> 99	
	gacggaaga gcacacgtct gaactccagt cacnnnnnnn natctcgtat gccgtcttct	60
	gcttg	65
15		

## REIVINDICACIONES

1. Método de selección de un ácido nucleico molde deseado a partir de una población de ácidos nucleicos molde, que comprende:

(a) poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  como un bloqueante para formar una mezcla; y

(b) aislar el ácido nucleico molde deseado a partir de la mezcla,

en el que la etapa de poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  comprende incubar la mezcla a una temperatura que presenta un valor de  $T_m$  del oligonucleótido aumentado en  $T_m$ ,

en el que la etapa de aislar el ácido nucleico molde deseado comprende:

(i) formar un complejo híbrido entre el ácido nucleico deseado y un segundo oligonucleótido como un señuelo; y

(ii) separar el complejo híbrido de la mezcla

en el que

el oligonucleótido aumentado en  $T_m$  comprende una pluralidad de grupos de residuos seleccionados de entre un grupo que consiste en un grupo de ácidos nucleicos bloqueados, un grupo de ácidos nucleicos bicíclicos, una pirimidina modificada en C5, un grupo de ácidos nucleicos peptídicos o combinaciones de los mismos,

el oligonucleótido aumentado en  $T_m$  es complementario a un ácido nucleico no diana y forma un dúplex con dicha secuencia de ácido nucleico no diana seleccionada de entre secuencias adaptadoras y de código de barras, y

la etapa de poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  da como resultado la inhibición de la formación de complejo entre el ácido nucleico molde deseado y los ácidos nucleicos molde no deseados.

2. Método según la reivindicación 1, en el que el oligonucleótido aumentado en  $T_m$  comprende desde aproximadamente 2 a aproximadamente 25 grupos de residuos seleccionados de entre un grupo que consiste en un grupo de ácidos nucleicos bloqueados, un grupo de ácidos nucleicos bicíclicos, una pirimidina modificada en C5, un grupo de ácidos nucleicos peptídicos o combinaciones de los mismos.

3. Método según la reivindicación 1, en el que el oligonucleótido aumentado en  $T_m$  comprende una pluralidad de grupos de residuos que comprenden unos grupos de ácidos nucleicos bloqueados o unos grupos de ácidos nucleicos bicíclicos.

4. Método según la reivindicación 4, en el que los grupos de ácidos nucleicos bloqueados o los grupos de ácidos nucleicos bicíclicos comprenden unas nucleobases seleccionadas de entre un grupo que consiste en citosina, adenina y timina.

5. Método según la reivindicación 1, en el que el oligonucleótido aumentado en  $T_m$  comprende por lo menos un miembro seleccionado de entre un grupo que consiste en SEC ID nº 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, 32, 34 y 36.

6. Método según la reivindicación 1, en el que el bloqueante presenta una complementariedad de secuencia a por lo menos una secuencia en un extremo de cada miembro de la población de ácidos nucleicos molde.

7. Método según la reivindicación 1, en el que el bloqueante comprende además un dominio de código de barras que presenta una pluralidad de nucleótidos.

8. Método según la reivindicación 1, en el que la pluralidad de nucleótidos comprende desde aproximadamente 5 a aproximadamente 12 nucleótidos dispuestos contiguos.

9. Método según la reivindicación 7, en el que el dominio de código de barras comprende unos nucleótidos que presentan como nucleobases por lo menos un miembro seleccionado de entre un grupo seleccionado de entre inosina, 3-nitropirrol, 5-nitroindol y combinaciones de los mismos.

10. Método según la reivindicación 1, en el que el segundo oligonucleótido comprende un señuelo.

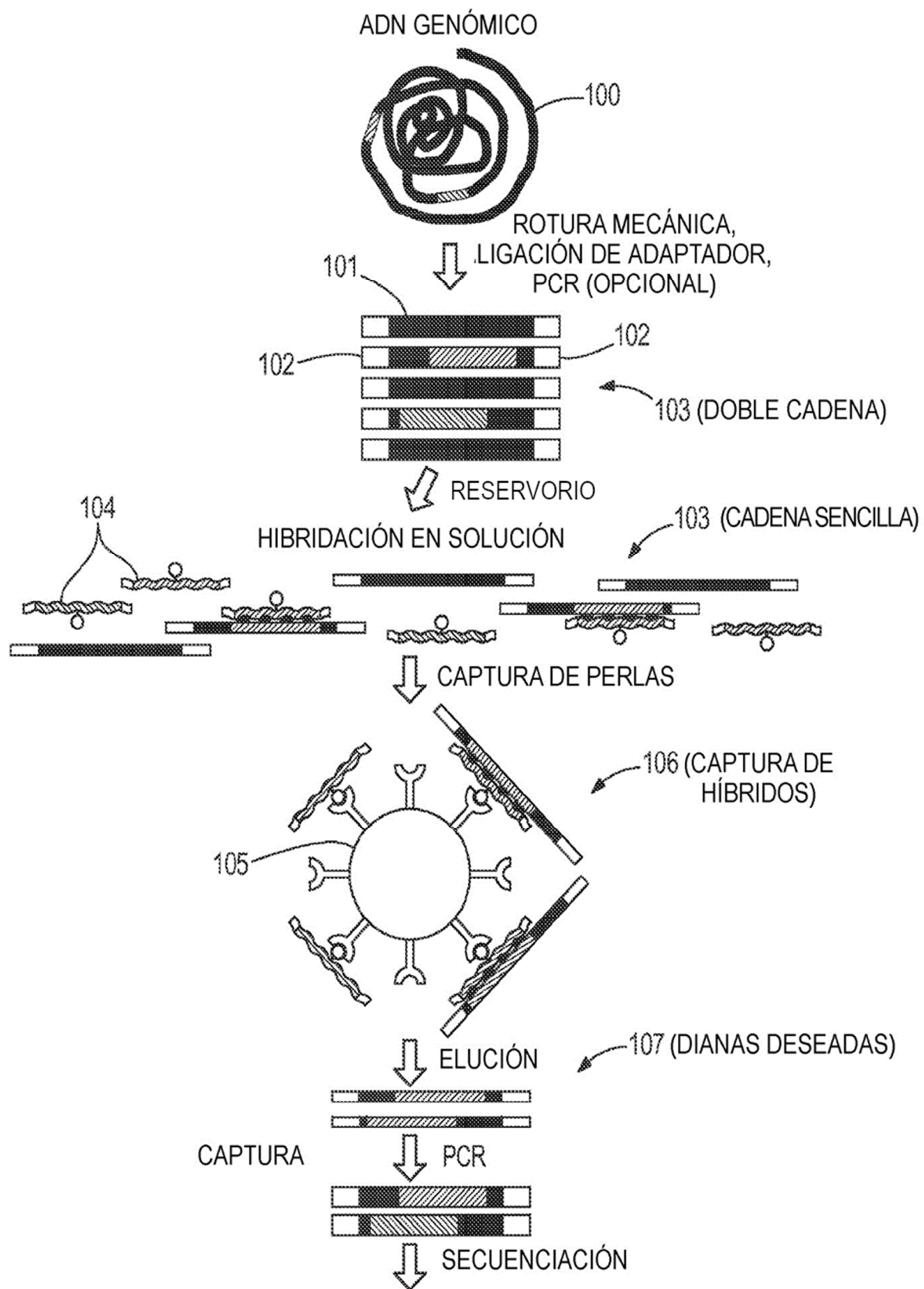
11. Método según la reivindicación 10, en el que el señuelo comprende una secuencia que presenta una complementariedad de secuencia a una secuencia dentro del ácido nucleico molde deseado.
- 5 12. Método según la reivindicación 10, en el que el señuelo comprende una pluralidad de grupos de residuos seleccionados de entre un grupo que consiste en un grupo de ácidos nucleicos bloqueados, un grupo de ácidos nucleicos bicíclicos, una pirimidina modificada en C5, un grupo de ácidos nucleicos peptídicos o combinaciones de los mismos.
- 10 13. Método según la reivindicación 10, en el que el señuelo incluye una modificación covalente para permitir la selección del complejo híbrido.
14. Método según la reivindicación 13, en el que la modificación covalente es un grupo biotinilado.
- 15 15. Método según la reivindicación 14, en el que el complejo híbrido se pone en contacto con un soporte sólido inmovilizado con avidina o estreptavidina.
16. Método para realizar una secuenciación en paralelo masiva, que comprende:
- 20 (a) preparar una población de biblioteca de ácidos nucleicos molde;
- (b) poner en contacto la población de biblioteca de ácidos nucleicos molde con por lo menos un oligonucleótido aumentado en  $T_m$  como bloqueante, una pluralidad de oligonucleótidos como señuelos y ADN  $C_{ot}$ -1 para formar una mezcla;
- 25 (c) aislar una pluralidad de ácidos nucleicos molde deseados a partir de la mezcla; y
- (d) secuenciar la pluralidad de ácidos nucleicos molde deseados,
- 30 en el que por lo menos un miembro de la pluralidad de oligonucleótidos como señuelos presenta una complementariedad de secuencia a una secuencia dentro de por lo menos un miembro de la pluralidad de ácidos nucleicos molde deseados
- en el que la etapa de poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  como un bloqueante comprende incubar la mezcla a una temperatura que presenta un valor de  $T_m$  del oligonucleótido aumentado en  $T_m$ ,
- 35 el oligonucleótido aumentado en  $T_m$  comprende una pluralidad de grupos de residuos seleccionados de entre un grupo que consiste en un grupo de ácidos nucleicos bloqueados, un grupo de ácidos nucleicos bicíclicos, una pirimidina modificada en C5, un grupo de ácidos nucleicos peptídicos o combinaciones de los mismos,
- 40 el oligonucleótido aumentado en  $T_m$  es complementario a un ácido nucleico no diana y forma un dúplex con dicha secuencia de ácido nucleico no diana seleccionada de entre secuencias adaptadoras y de código de barras,
- y
- la etapa de poner en contacto la población de ácidos nucleicos molde con un primer oligonucleótido que comprende un oligonucleótido aumentado en  $T_m$  da como resultado la inhibición de la formación de complejo entre el ácido
- 45 nucleico molde deseado y los ácidos nucleicos molde no deseados.
17. Método según la reivindicación 16, en el que cada uno de los miembros de la población de biblioteca de ácidos nucleicos molde incluye por lo menos una secuencia adaptadora terminal idéntica que presenta un intervalo de tamaño desde aproximadamente 15 nucleótidos a aproximadamente 75 nucleótidos.
- 50 18. Método según la reivindicación 17, en el que el bloqueante presenta una complementariedad de secuencia a por lo menos una secuencia adaptadora terminal idéntica de la población de biblioteca de ácidos nucleicos molde.
19. Método según la reivindicación 17, en el que la por lo menos una secuencia adaptadora terminal idéntica incluye un dominio de código de barras.
- 55 20. Método según la reivindicación 19, en el que el bloqueante presenta complementariedad de secuencia a por lo menos una secuencia adaptadora terminal idéntica.
21. Método según la reivindicación 18, en el que la etapa de poner en contacto (b) comprende incubar la mezcla a una temperatura que presenta un valor de  $T_m$  de por lo menos un oligonucleótido aumentado en  $T_m$ .
- 60 22. Método según la reivindicación 16, en el que el por lo menos un oligonucleótido aumentado en  $T_m$  como bloqueante comprende por lo menos un miembro seleccionado de entre el grupo que consiste en SEC ID nº 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, 32, 34 y 36.
- 65

23. Método según la reivindicación 16, en el que la etapa de aislar una pluralidad de ácidos nucleicos molde deseados a partir de la mezcla comprende:

- 5           (i) formar una pluralidad de complejos híbridos entre la pluralidad de ácidos nucleicos molde deseados y una pluralidad de oligonucleótidos como señuelos; y  
          (ii) separar la pluralidad de complejos híbridos de la mezcla.

10       24. Método según la reivindicación 23, en el que la pluralidad de oligonucleótidos como señuelos comprende una pluralidad de grupos de residuos seleccionados de entre un grupo que consiste en un grupo de ácidos nucleicos bloqueados, un grupo de ácidos nucleicos bicíclicos, una pirimidina modificada en C5, un grupo de ácidos nucleicos peptídicos o combinaciones de los mismos.

15       25. Método según la reivindicación 23, en el que cada señuelo incluye una modificación covalente para permitir la selección del complejo híbrido que incluye un señuelo.



**FIG. 1**

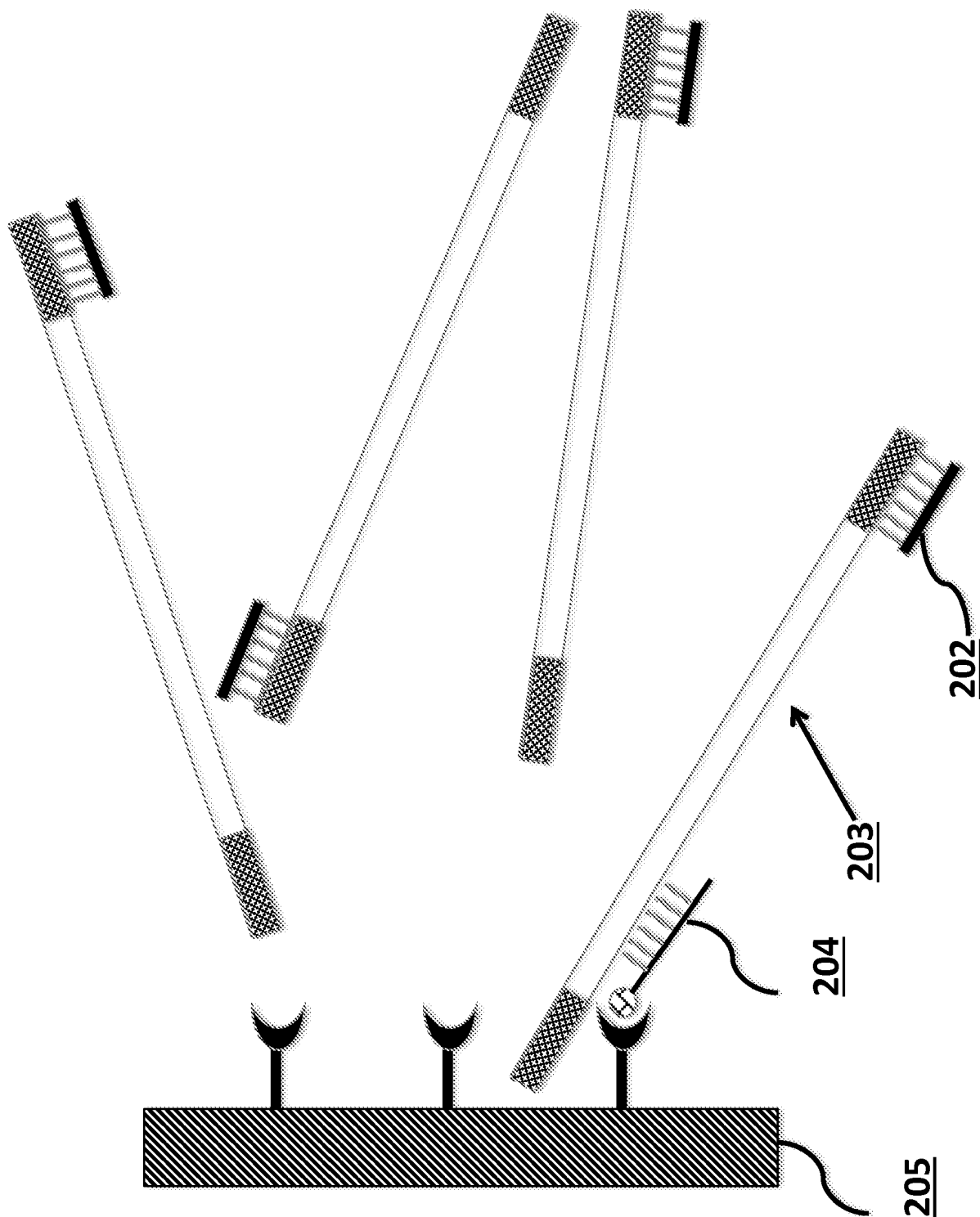


FIG. 2A

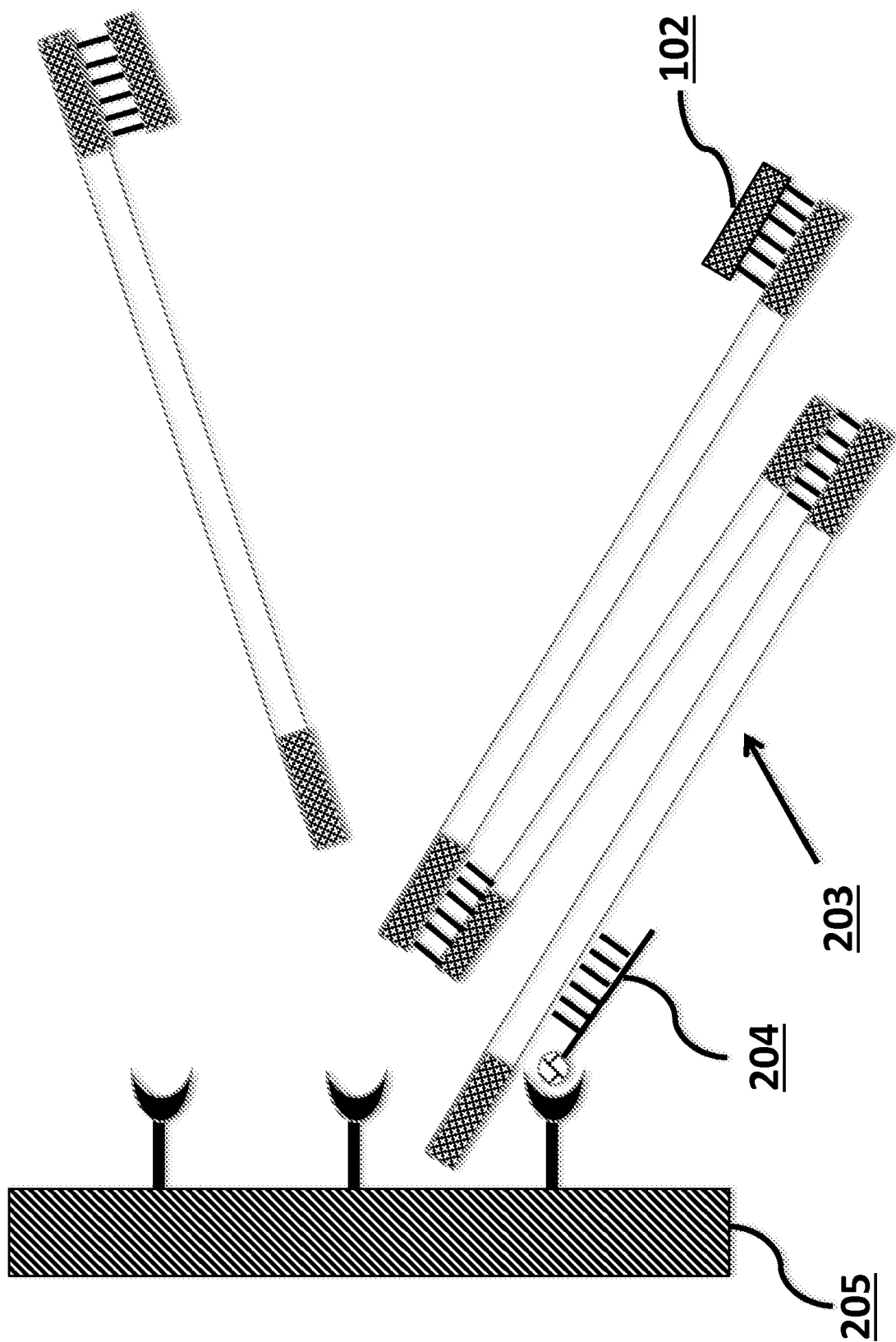


FIG. 2B

# Recepción de muestras y control de calidad; Aislamiento del ADN

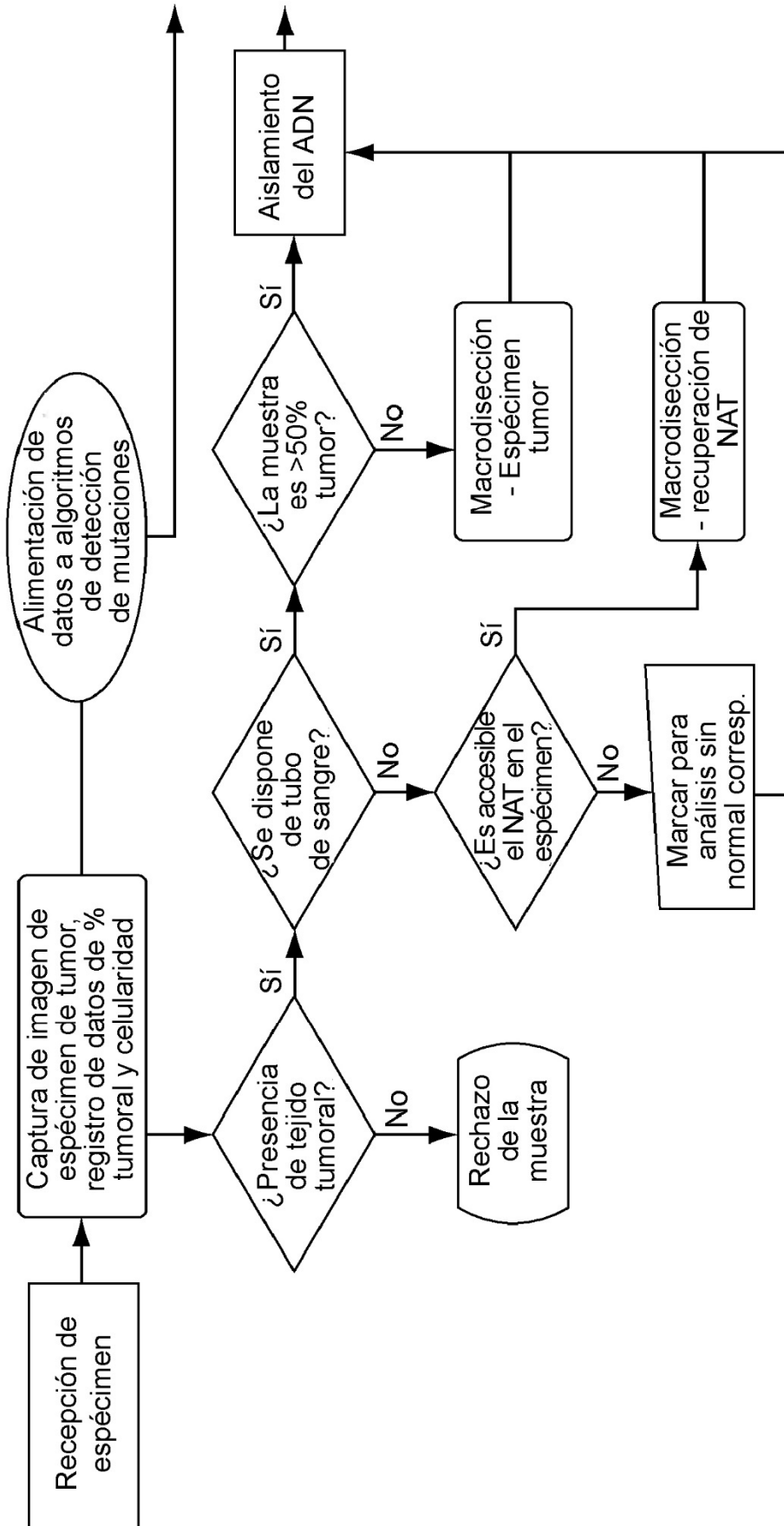


FIG. 3A

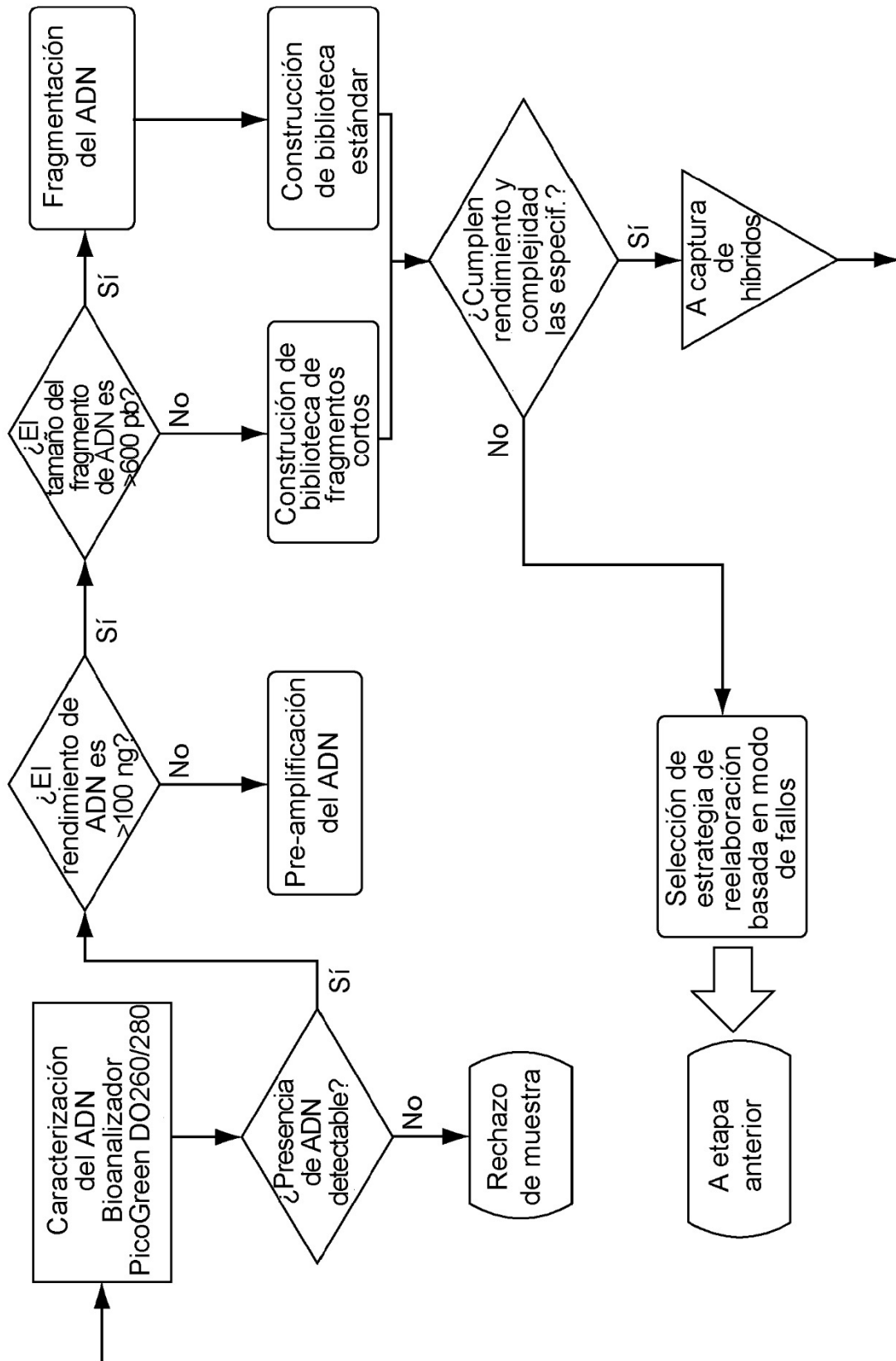
**Control de calidad del ADN y generación de bibliotecas**

FIG. 3B

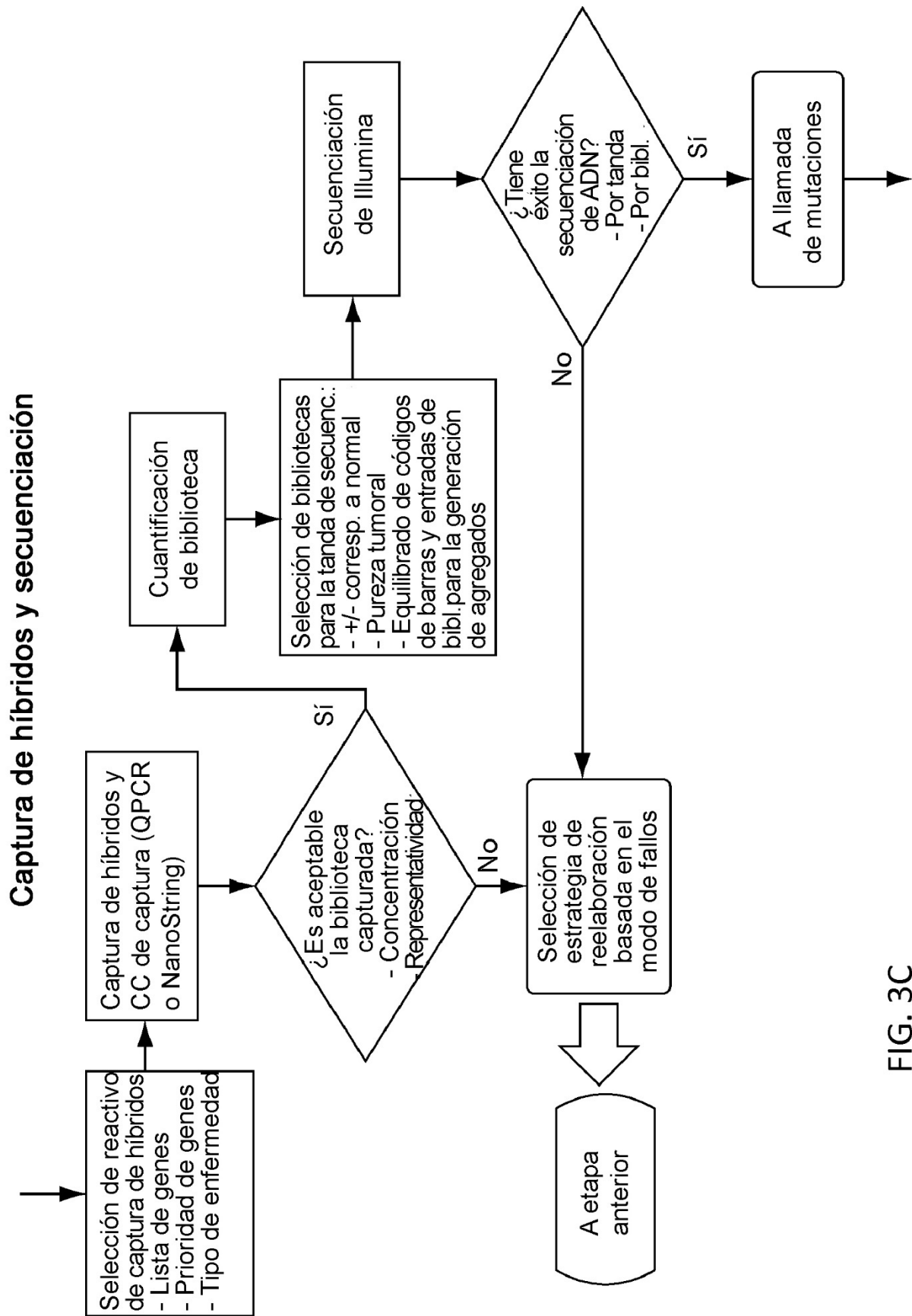


FIG. 3C

# Control de calidad de datos de secuencias y llamada de mutaciones

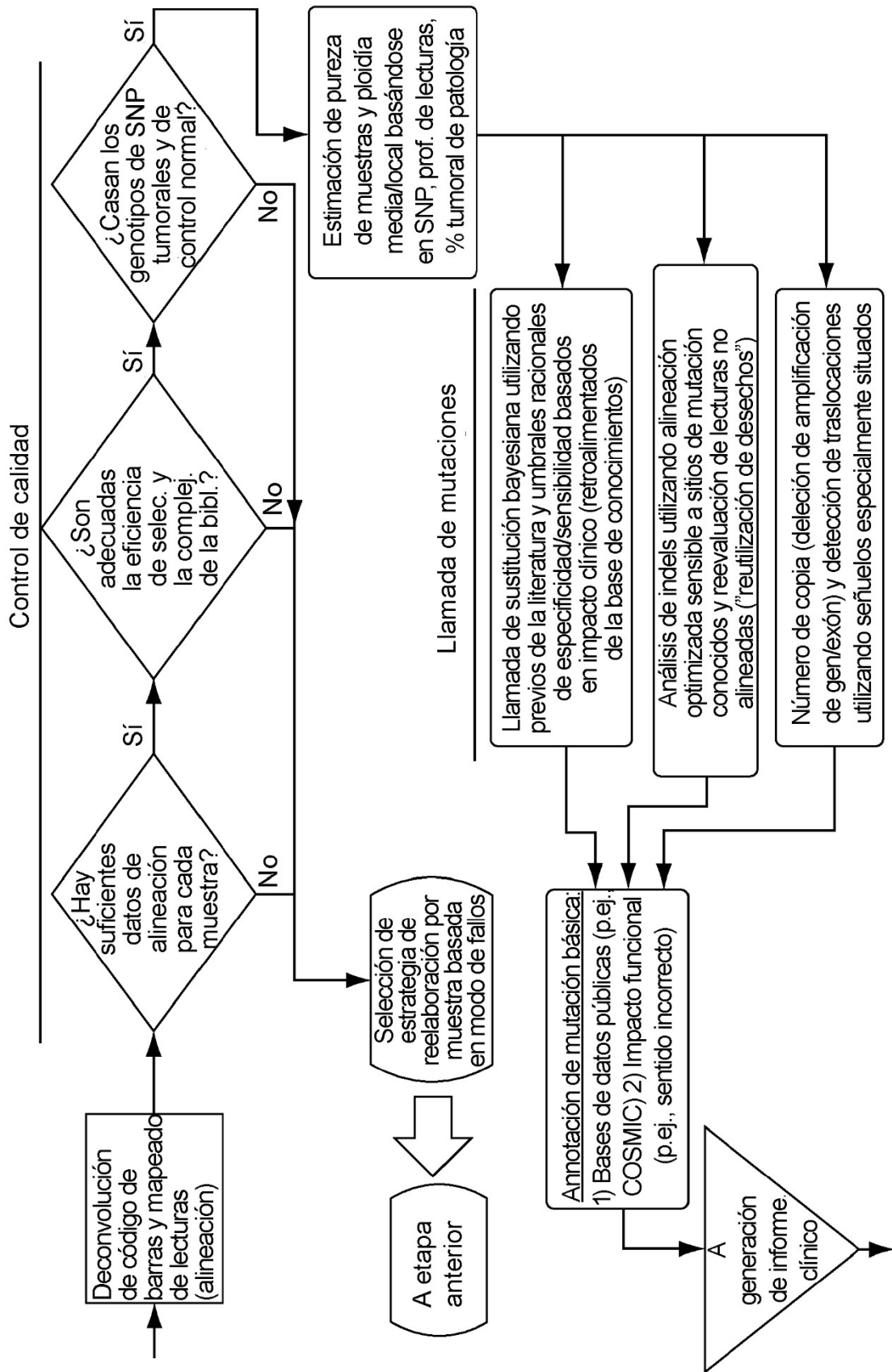


FIG. 3D

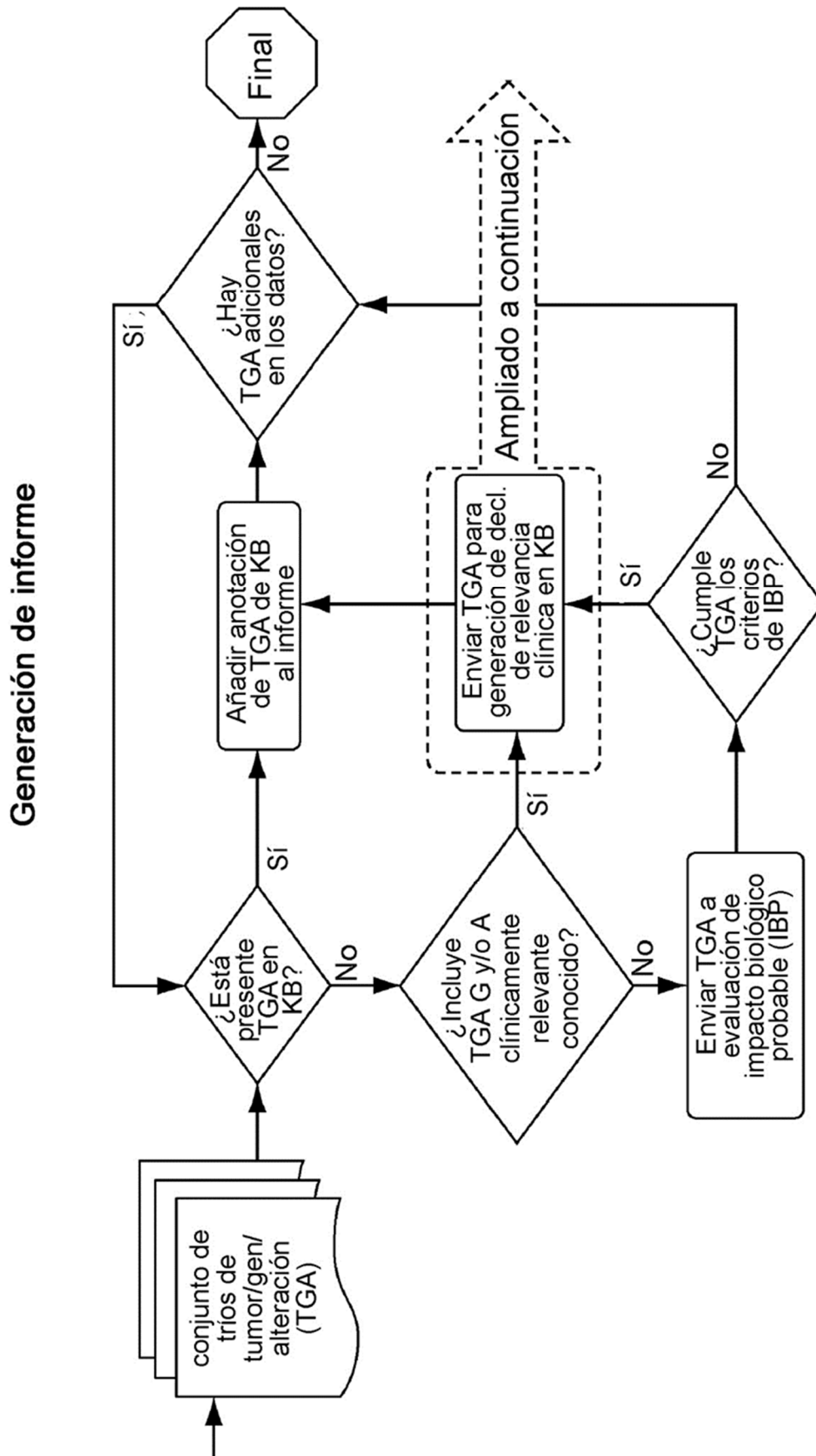


FIG. 3E

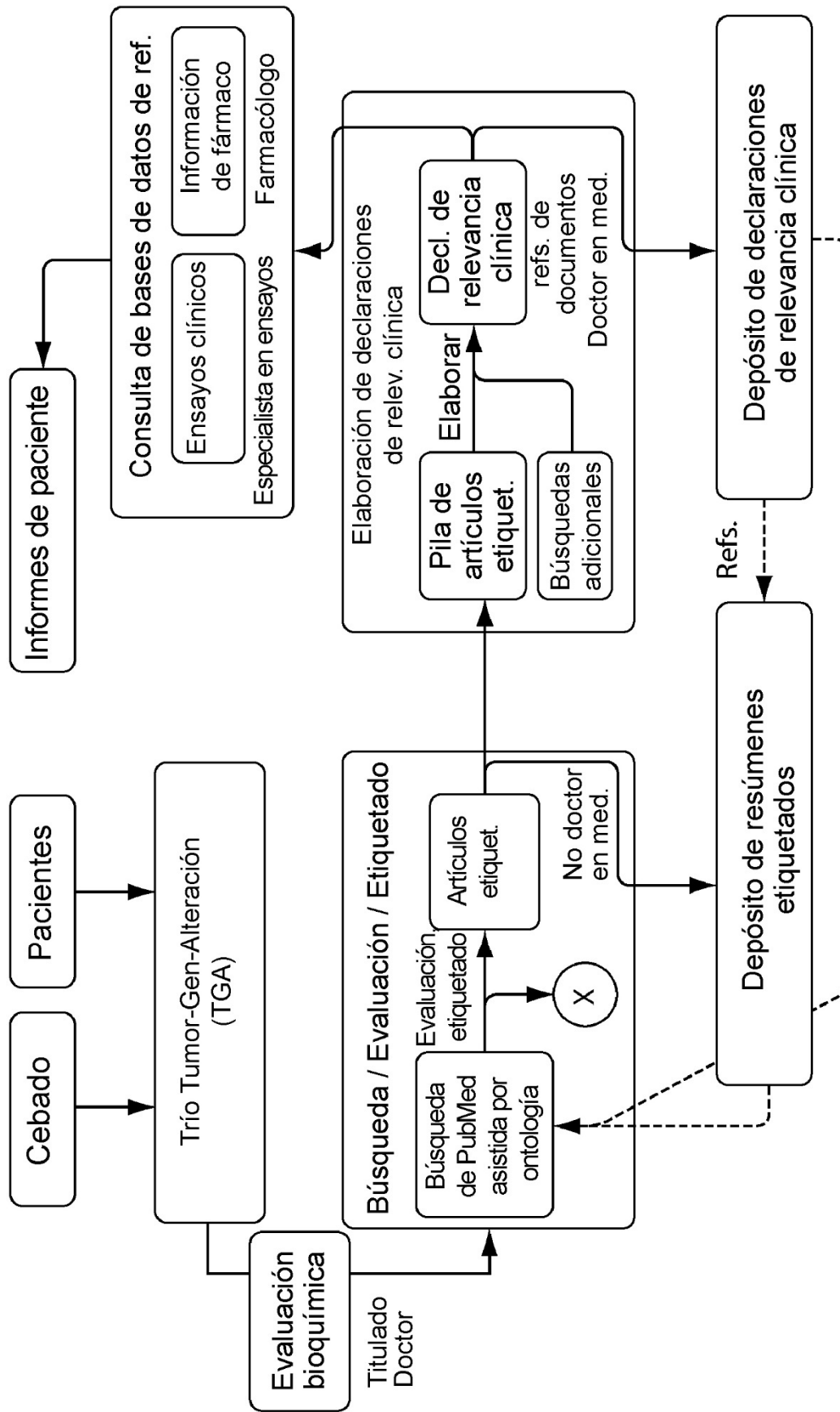


FIG. 3F

# Impacto de expectativas previas y profundidad de lectura sobre la detección de mutaciones

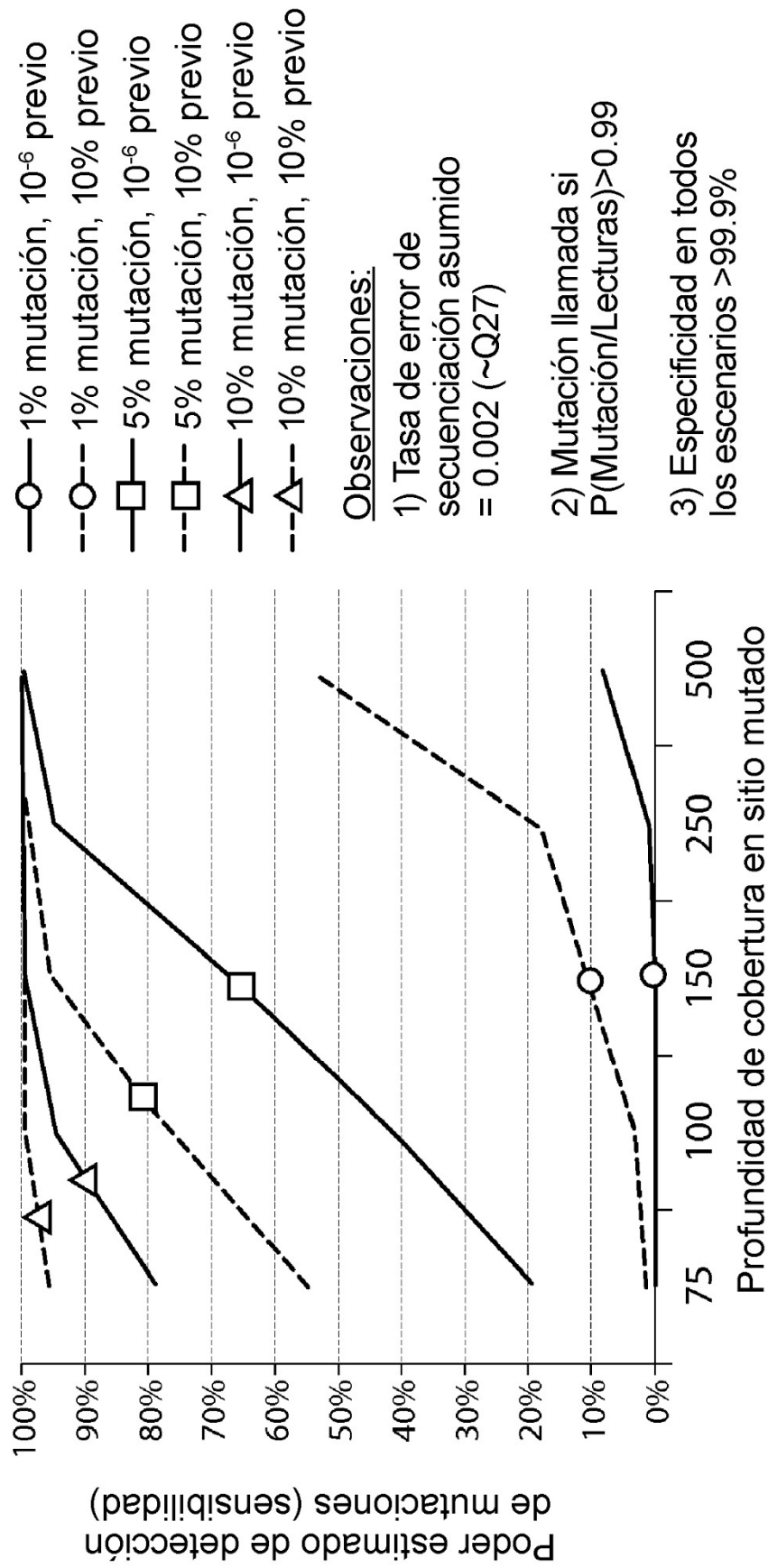


FIG. 4

# Frecuencias de mutación en > 100 muestras clínicas de cáncer

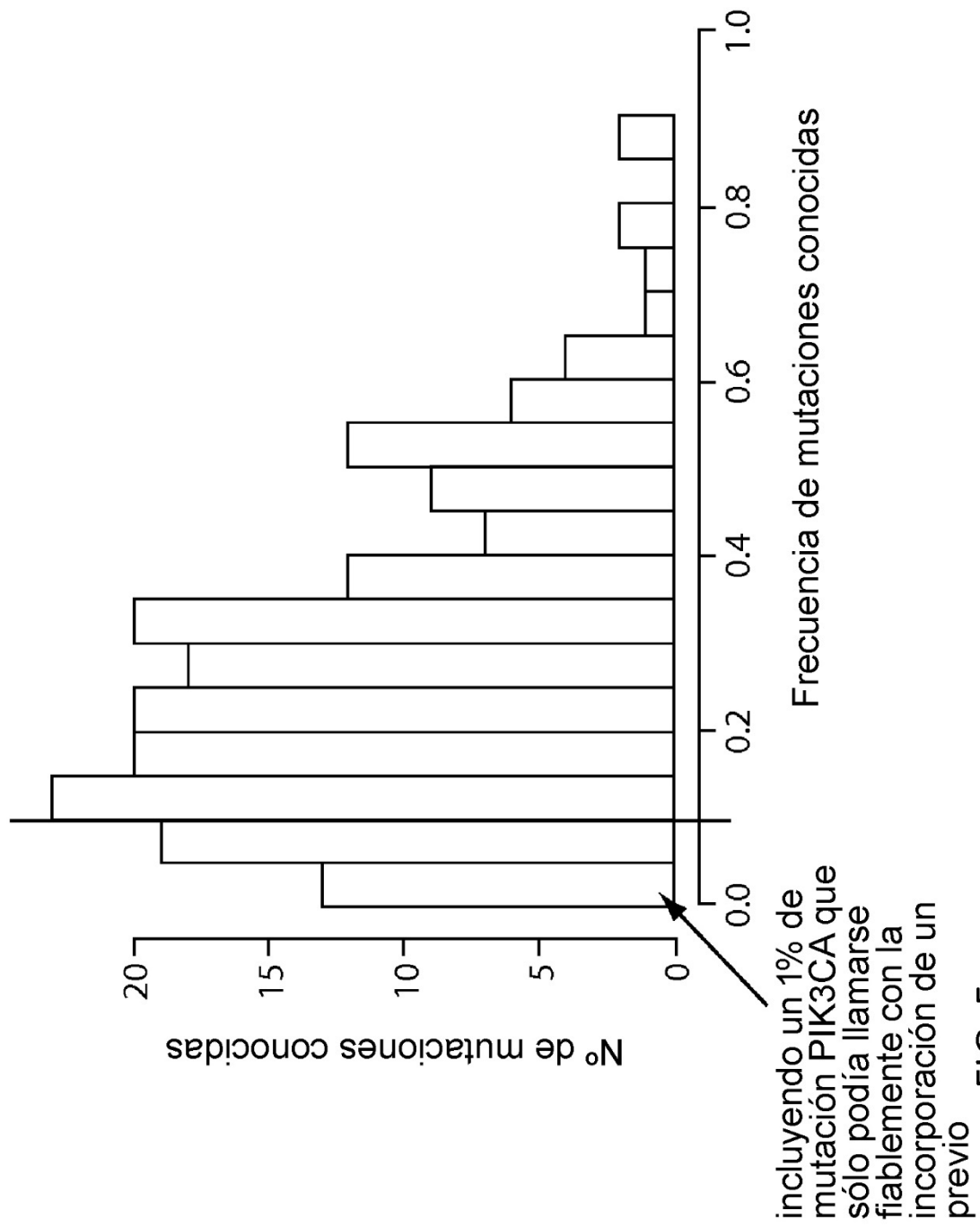


FIG. 5

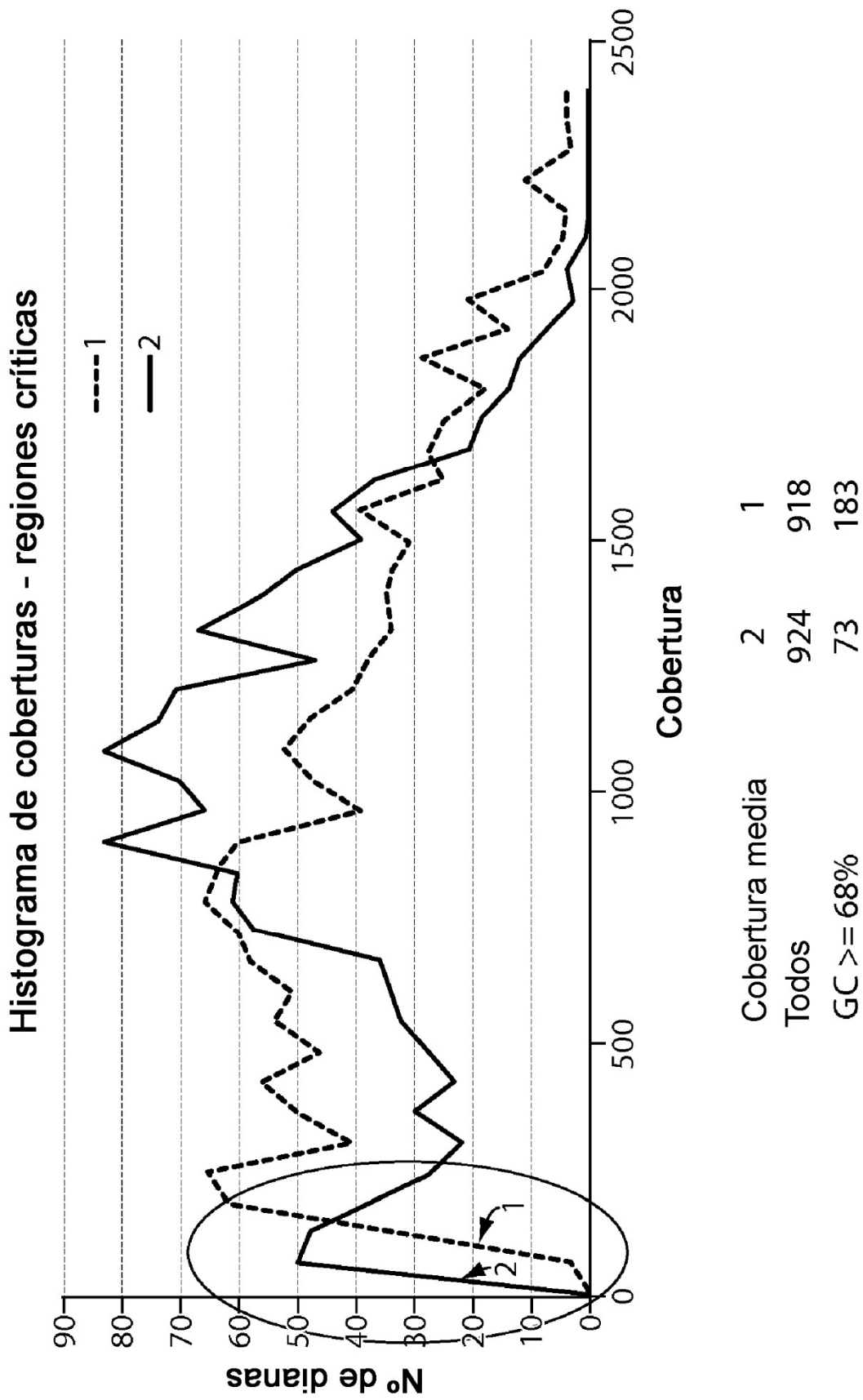


FIG. 6

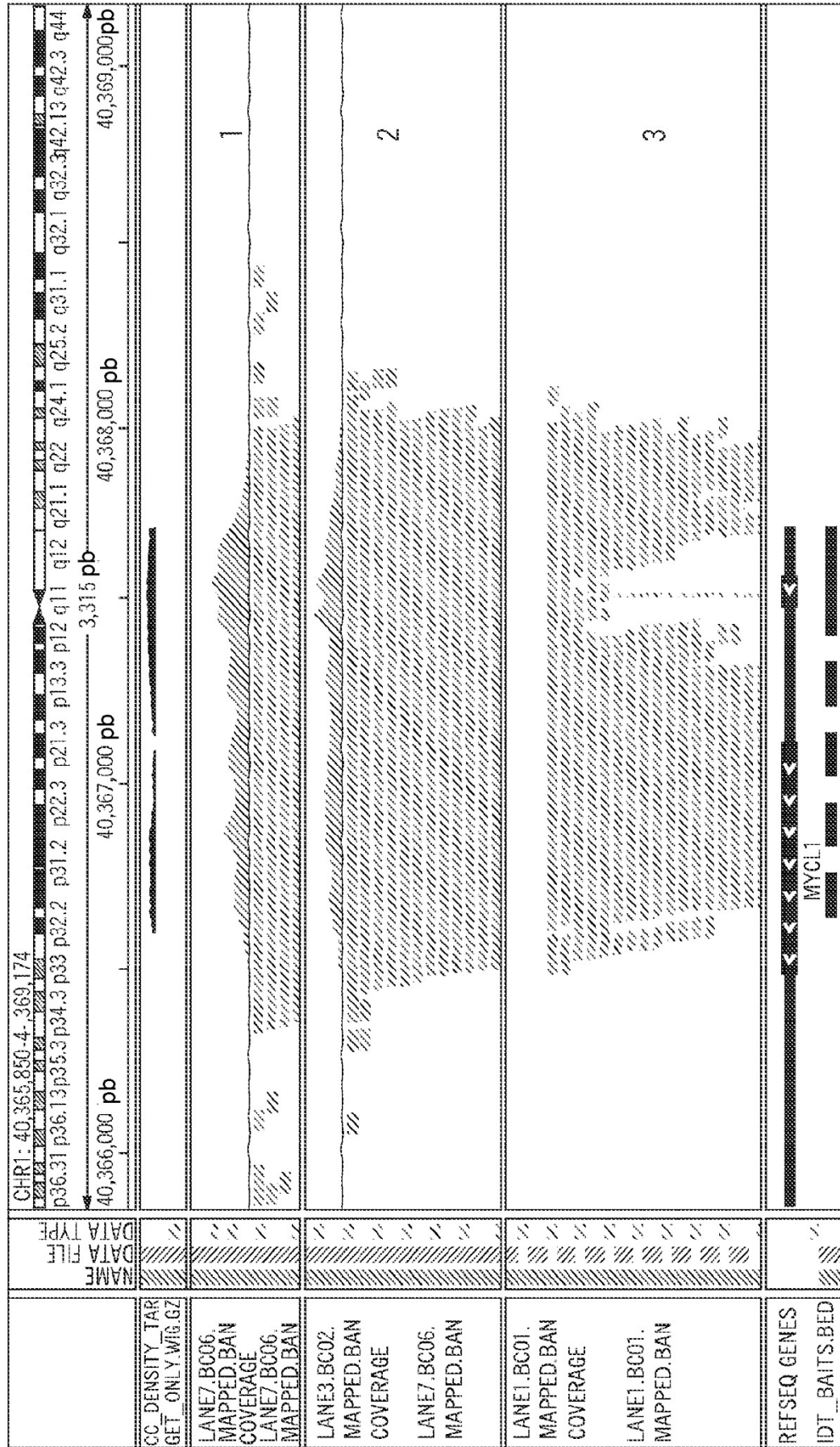


FIG. 7

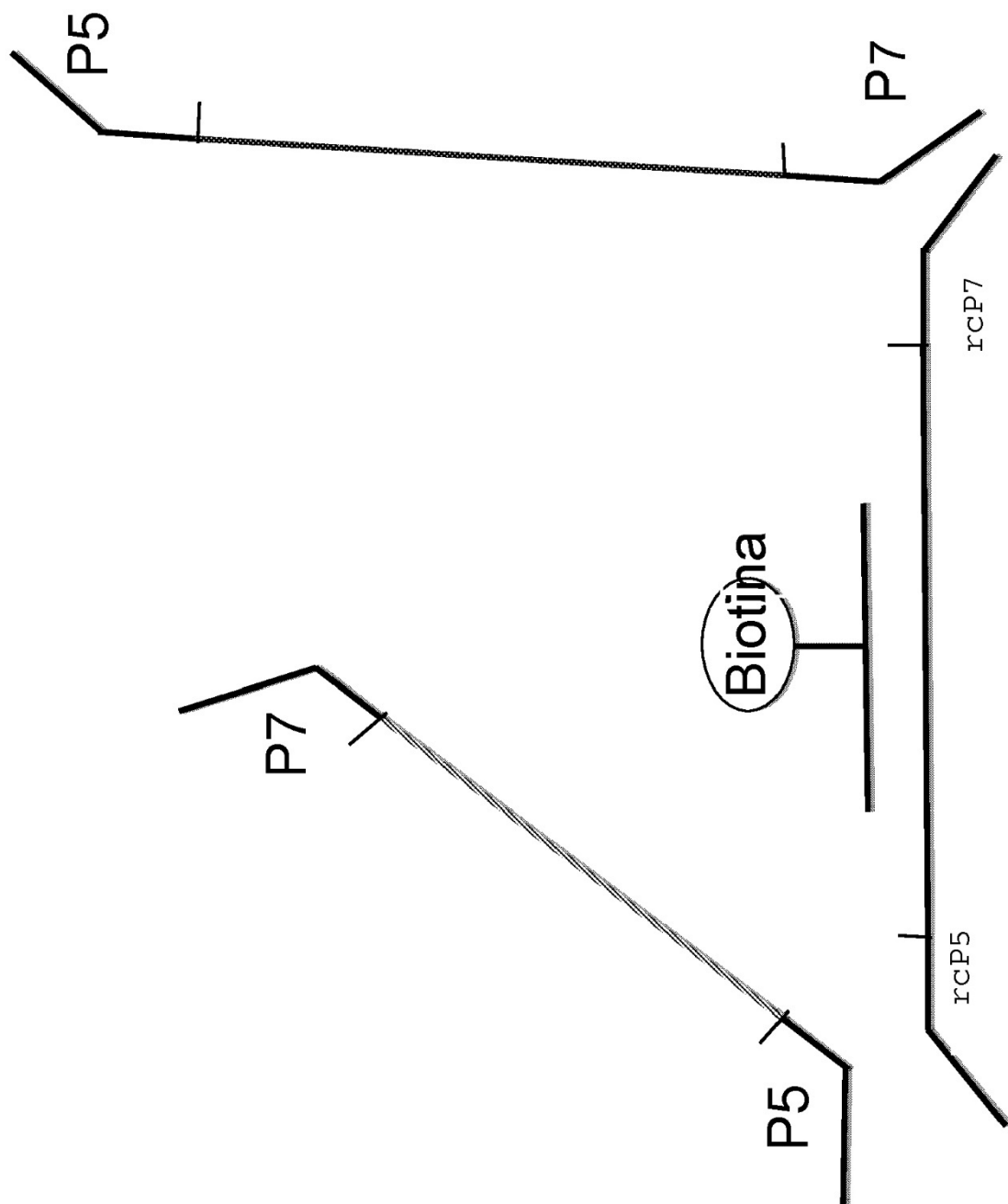


FIG. 8

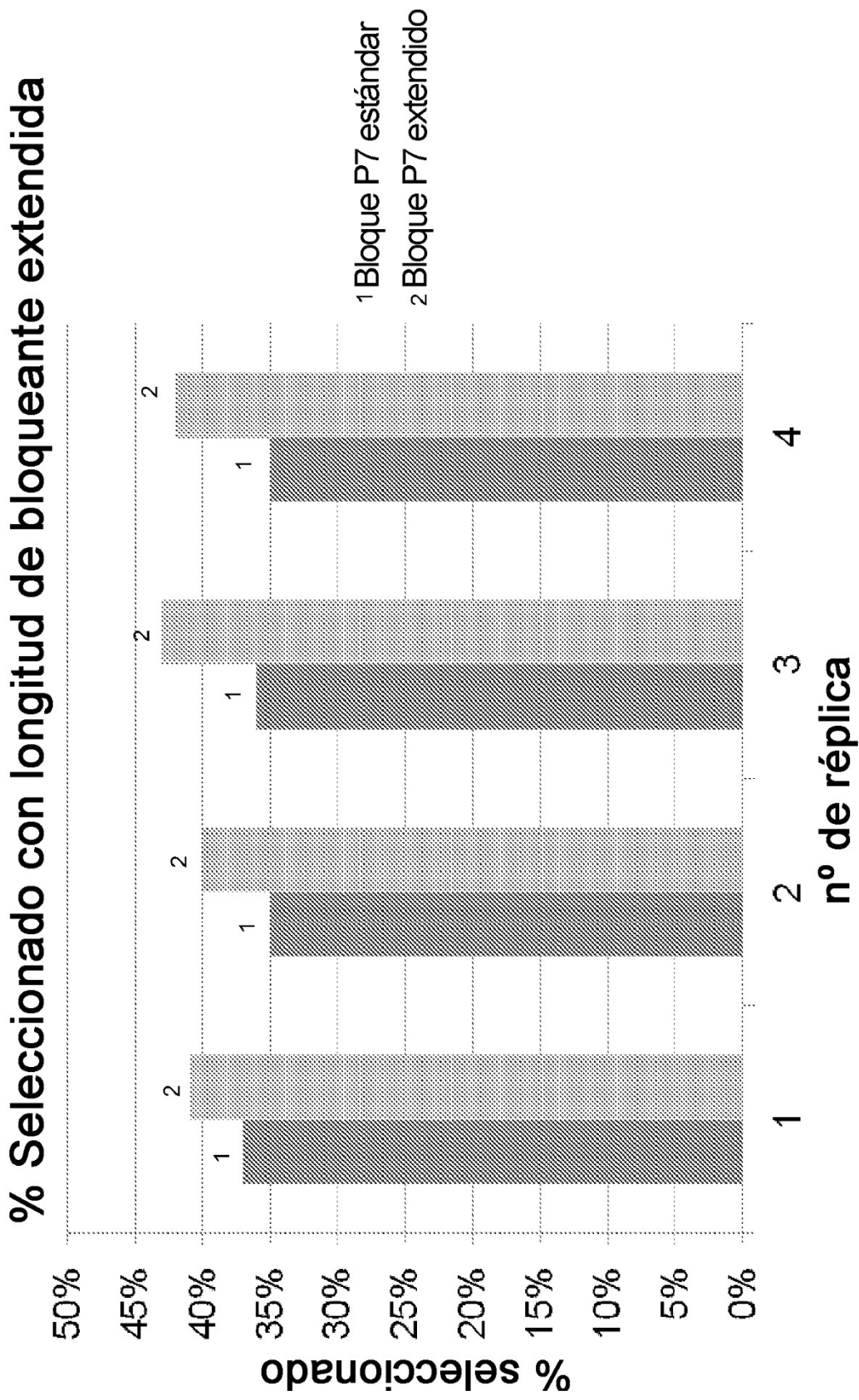


FIG. 9

# Histogramas de cobert. de exones con bloqueante estándar o extendido

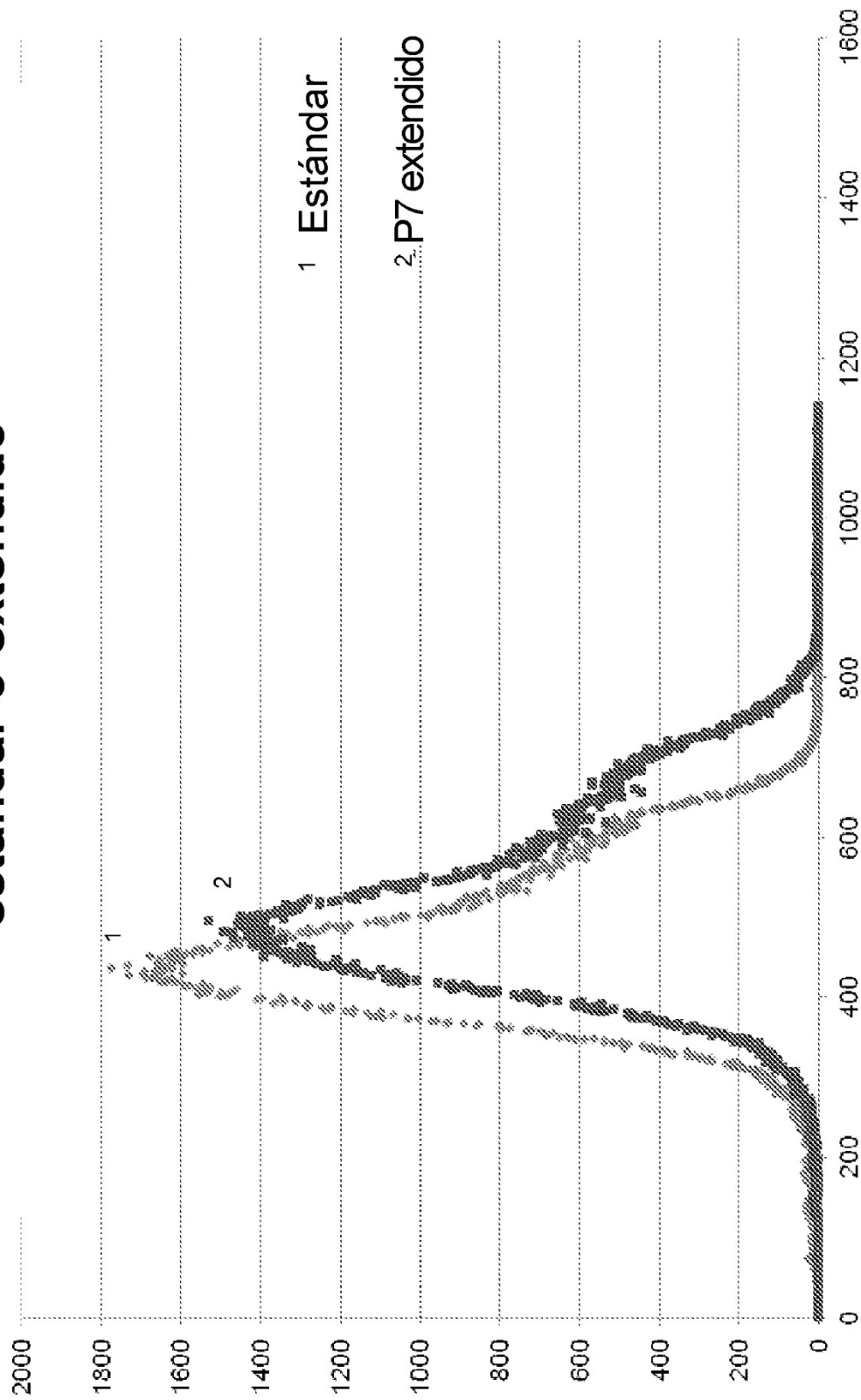


FIG. 10

## Número de lecturas de secuenciación por muestra

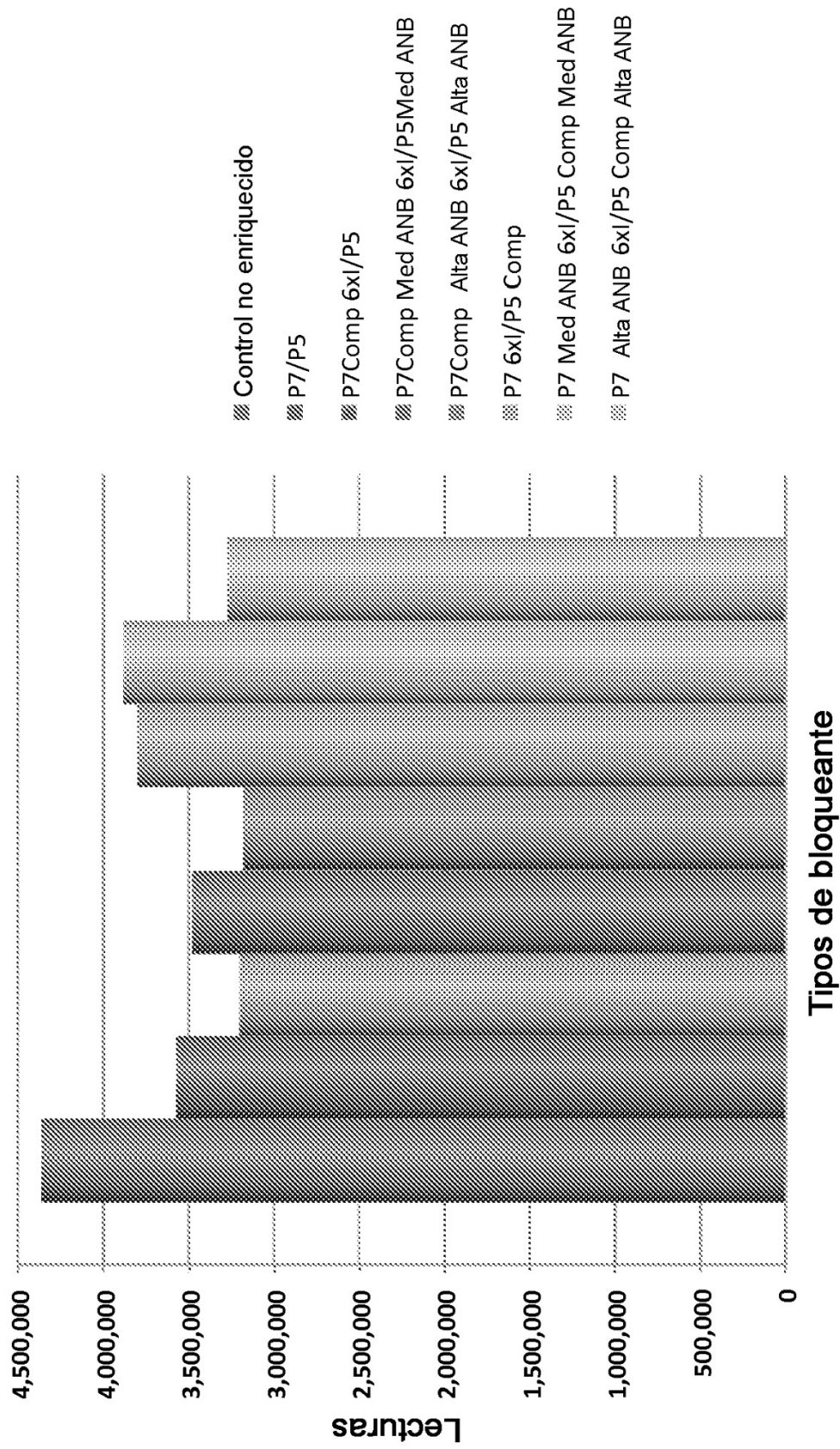
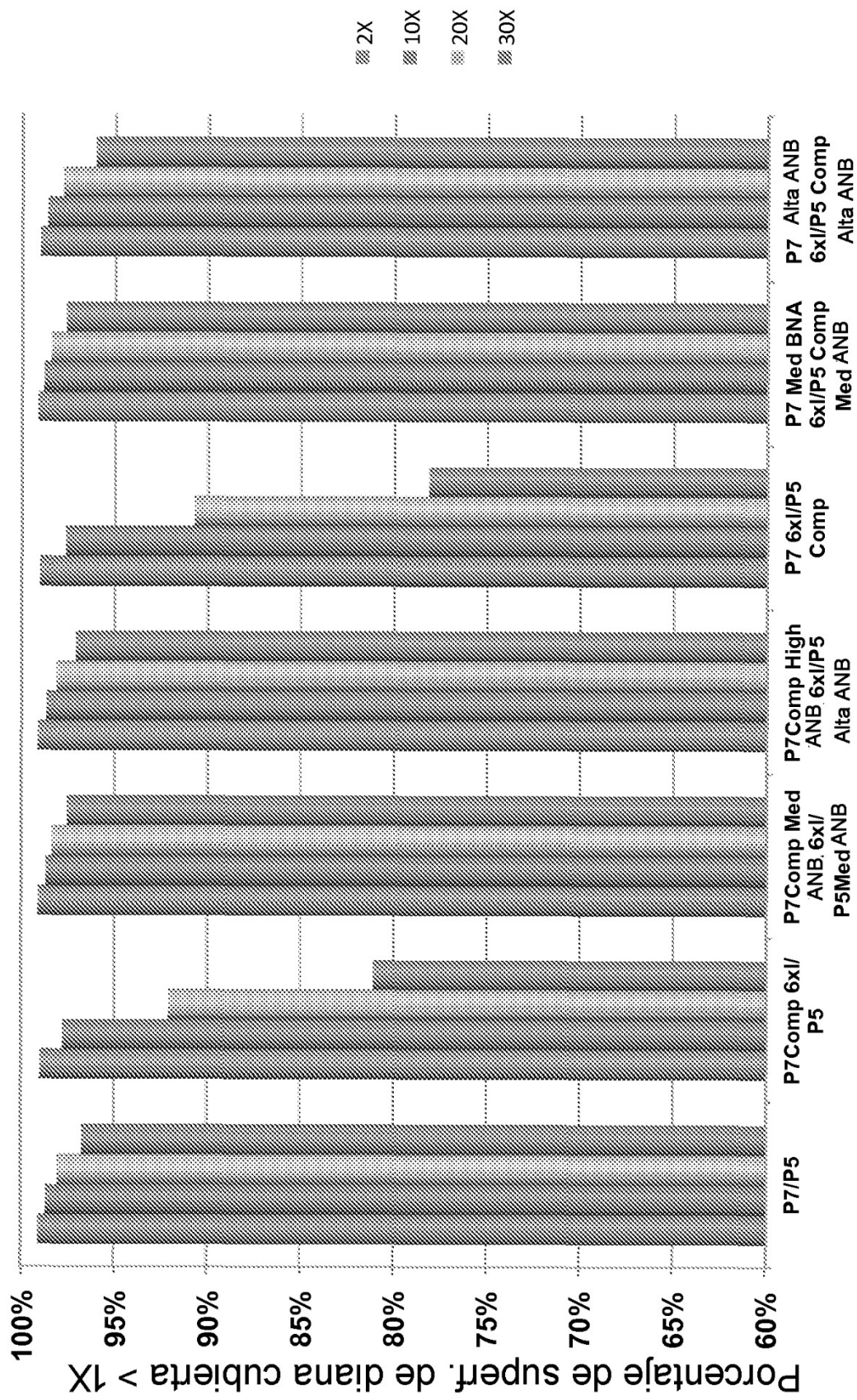


FIG. 11



Tipos de bloqueante

FIG. 12

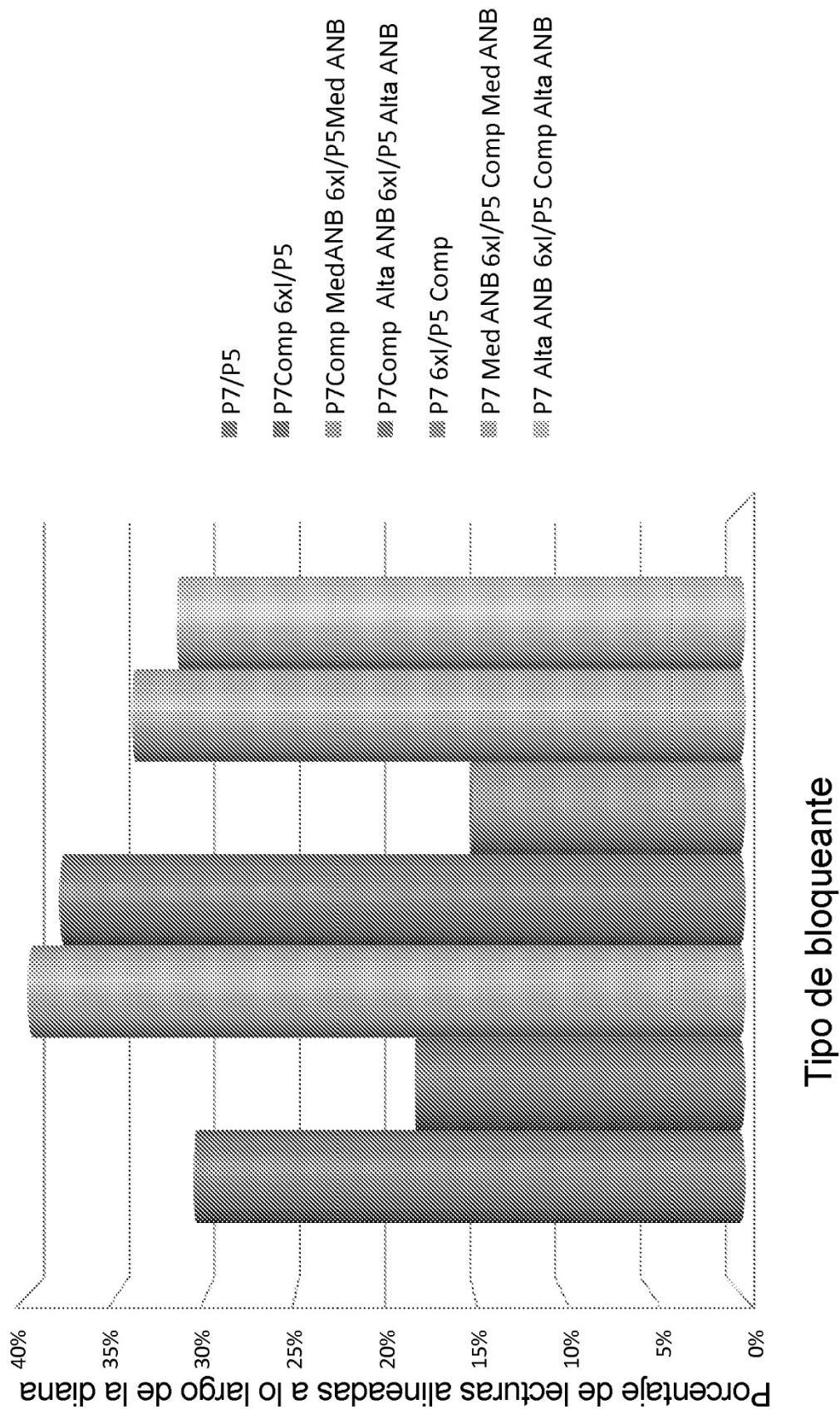


FIG. 13

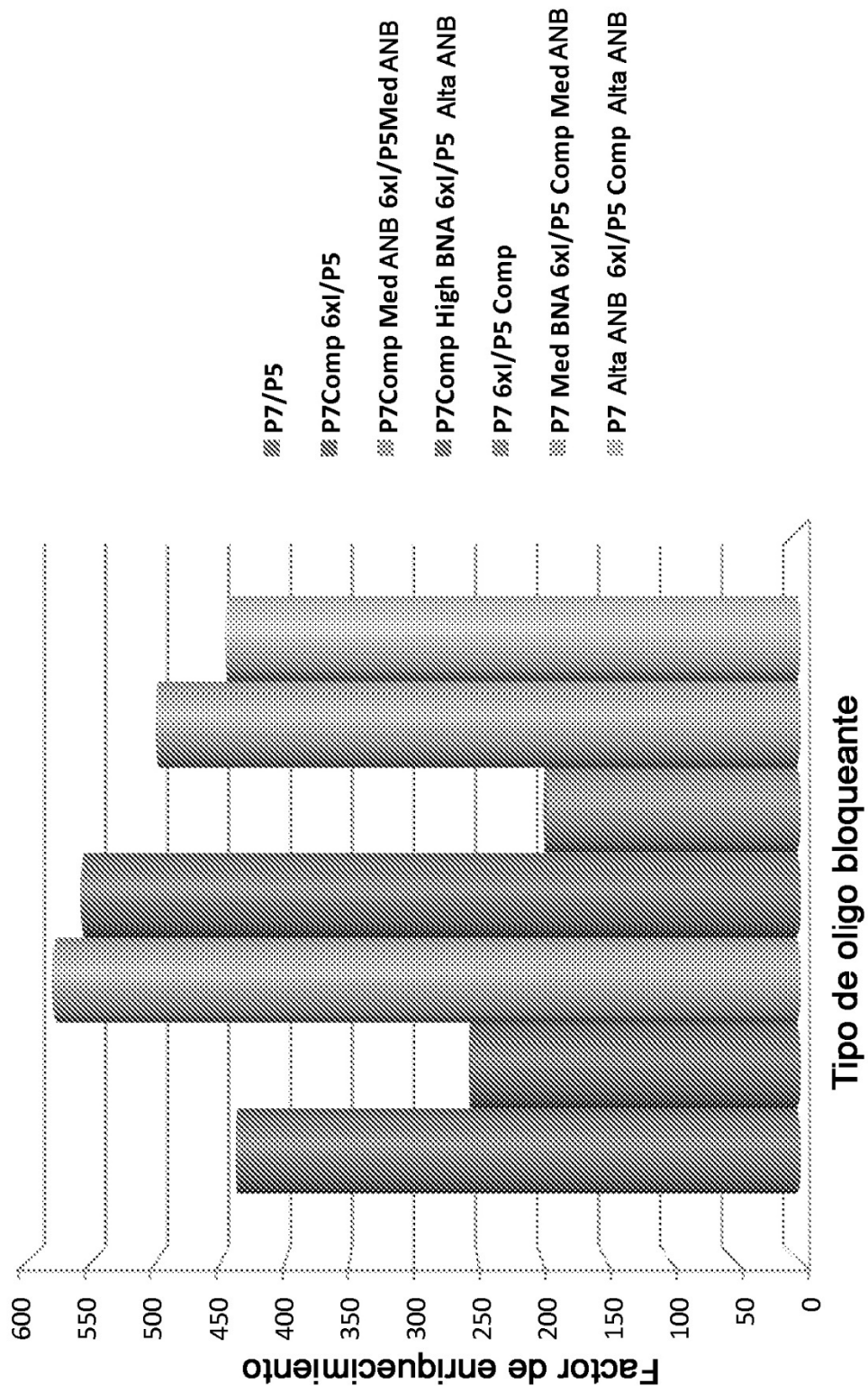
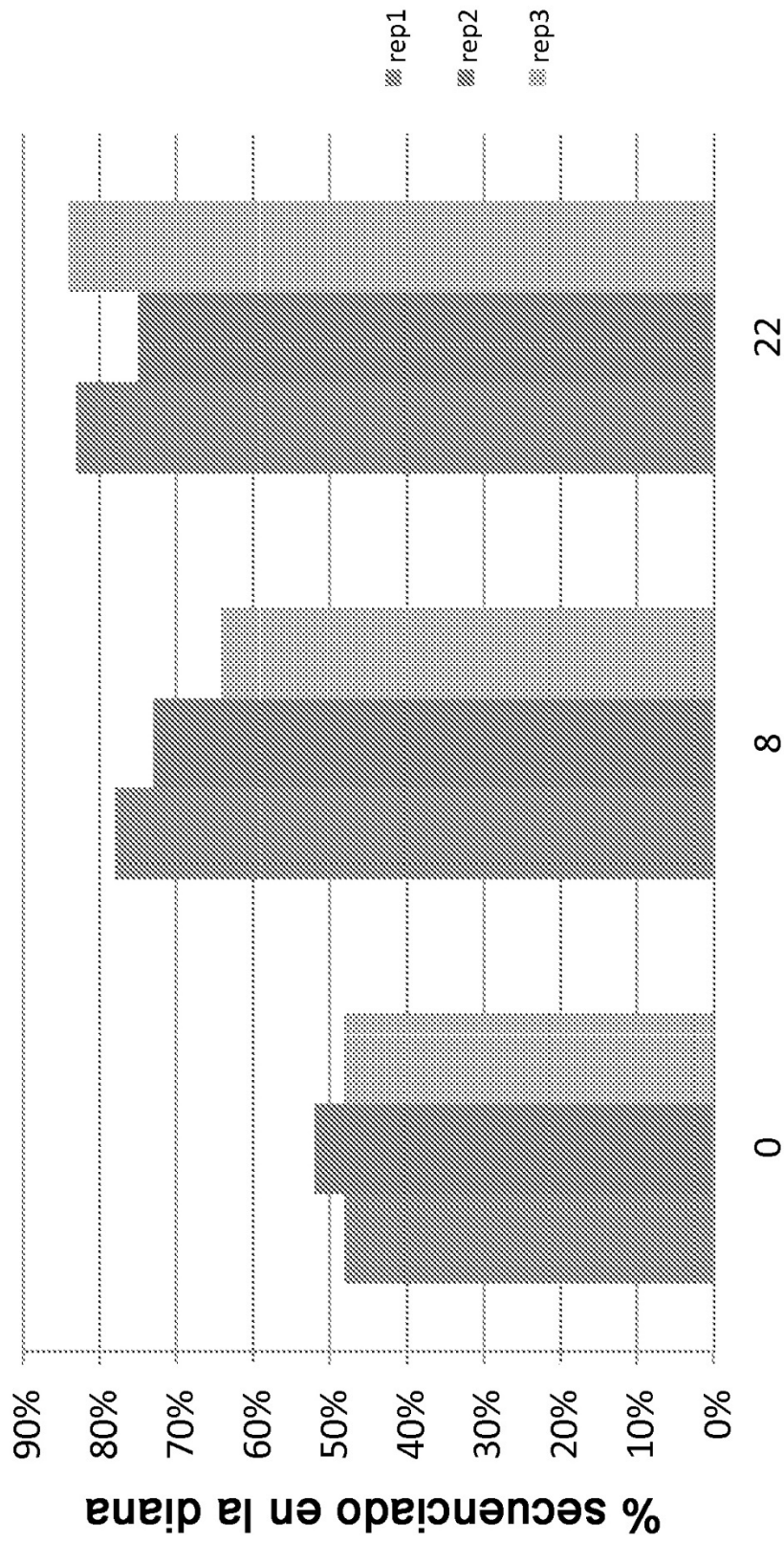


FIG. 14

## Variación del número de ANB sustituidos en bloqueantes Pe e In



Número de ANB sustituidos en los bloqueantes Pe/In

FIG. 15